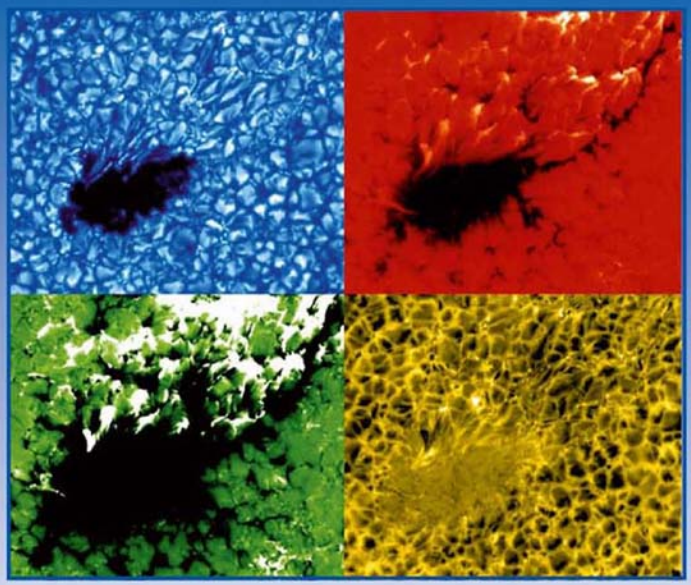
SPACE SCIENCES SERIES OF ISSI

The Origin and Dynamics of Solar Magnetism

M.J. Thompson, A. Balogh,
J.L. Culhane, Å. Nordlund,
S. K. Solanki and J.-P. Zahn (Eds.)



Springer

INTERNATIONAL SPACE SCIENCE INSTITUTE

M.J. Thompson • A. Balogh • J.L. Culhane •
Å. Nordlund • S.K. Solanki • J.-P. Zahn
Editors

The Origin and Dynamics of Solar Magnetism

Previously published in *Space Science Reviews* Volume 144, Issues 1–4, 2009

Springer

M.J. Thompson
School of Mathematics & Statistics
University of Sheffield
Sheffield, UK

A. Balogh
International Space Science Institute
Bern, Switzerland

J.L. Culhane
University College London
Mullard Space Science Laboratory
Dorking, UK

Å. Nordlund
Niels Bohr Institute
University of Copenhagen
Copenhagen, Denmark

S.K. Solanki
Max-Planck-Institut
für Sonnensystemforschung Katlenburg-Lindau
Germany

J.-P. Zahn
LUTH, Observatoire de Paris
Meudon, France

Cover illustration: Continuum image (blue) and line-of-sight components of the velocity (yellow) and magnetic field (red and enhanced as green) obtained from Milne-Eddington inversions of Stokes data observed with the CRisp Imaging SPectropolarimeter (CRISP) on the Swedish 1-m Solar Telescope (SST). Shown is a shortlived active region observed on 22 April 2008 in the Fe I 6302 line. The spatial resolution is close to the diffraction limit of 0.16".
Courtesy of Tomas Hillberg, Gautam Narayan and Göran Scharmer.

Library of Congress Control Number: 2009926695

DOI: 10.1007/978-1-4419-0239-9

ISBN-978-1-4419-0238-2 e-ISBN-978-1-4419-0239-9

Printed on acid-free paper.

1

springer.com

Contents

Introduction to Solar Magnetism: The Early Years

A. Balogh · M.J. Thompson

Originally published in the journal Space Science Reviews, Volume 144, Nos 1–4, 1–14.
DOI: 10.1007/s11214-009-9493-x

Abstract The year 2008 marked the one hundredth anniversary of the observational discovery by George Ellery Hale of magnetic field in sunspots (Hale in Astrophys. J. 28:315–343, 1908). This observation, the first to suggest a direct link between the best-known variable features on the Sun and magnetism, started a line of research that has widened considerably over the last 100 years and is continuing today. Knowledge about all aspects of the Sun has increased in a remarkable way over the past few decades. Variations in the appearance of the Sun and its corona, as well as deeper sources of quasi-regular and chaotic changes that make up solar variability have been extensively documented by both ground-based and space-based solar observatories. It has been recognized that solar magnetism is the key phenomenon that drives solar variability. The workshop devoted to the origin and dynamics of solar magnetism held in the International Space Science Institute in Bern, Switzerland, from 21 to 25 January 2008 reviewed the status of the field and has led to this volume that brings together the best available knowledge and understanding of solar magnetism 100 years after Hale's pioneering paper. This introductory paper gives an outline of the history of research into solar variability up to the work of Hale and his colleagues. The achievements of the past decades are discussed extensively in the other contributions to this volume.

Keywords Sun · Solar magnetism · Sunspots · Solar cycle

1 Sunspots as Indicators of Solar Variability

Sunspots are probably the most obvious and longest recognized manifestations of the variable Sun, and now of solar magnetic activity. These regions of relatively cool gas/plasma

A. Balogh (✉)
International Space Science Institute, Bern, Switzerland
e-mail: balogh@issibern.ch

A. Balogh
The Blackett Laboratory, Imperial College, London, UK

M.J. Thompson
School of Mathematics and Statistics, University of Sheffield, Sheffield, UK

M.J. Thompson et al. (eds.), *The Origin and Dynamics of Solar Magnetism*.
DOI: 10.1007/978-1-4419-0239-9_1

at the Sun's surface are caused by the suppression of convective heat transport by intense magnetic field generated in the solar convection zone. The number of sunspots visible on Sun and the area they cover on the photosphere exhibit an approximately 11-year cycle, and over that same time the solar latitude at which new spots appear migrates from mid-latitudes towards the solar equator. These are the two most recognizable features of solar variability, but the range of measures that are used to characterize temporal changes in the Sun is very large.

Sunspots have now been observed and counted since the early 17th century, and studied in ever increasing detail ever since then. (For a comprehensive account of the history of observations of sunspot and solar activity, see Chapt. 2.1 in Hoyt and Schatten 1997). They were first recognized as dark features, spots, on an otherwise idealized, unblemished Sun almost exactly 400 years ago, in and around 1610 or 1611, by Thomas Harriott, Johann Fabricius, Christophe Scheiner and Galileo, very soon after the introduction of the telescope as a tool into astronomy. (There is some evidence for earlier observations, but those who made them did not know what they saw and, for cultural reasons, certainly did not interpret their observations as blemishes or spots on the Sun.) A number of other observers also joined in documenting sunspots, although the controversy about their nature, whether really spots on the Sun or "clouds" or even "transiting planets" continued for up to two or three decades.

The discovery and sustained observation of sunspots in the first half of the 17th century immediately led to the discovery of solar rotation and its period, the determination of the rotation axis of the Sun, and the latitude dependence of the rotation period. Both Galileo and Scheiner claimed the merit for discoveries concerning the Sun, based on their observations of sunspots. These observations from this early phase of scientific solar research are very useful in providing a record of normality in solar activity prior to the start of a long, 70 year interval (from about 1645 to 1715) when the Sun had very few spots, now called the Maunder minimum.

We can now be grateful for the nearly systematic observing and counting of sunspots since Galileo's time because these records provide evidence of the longevity of what we now know to be the solar magnetic activity cycle, even if, historically, it can be seen to be chaotic rather than regular.

The close to 11 year periodicity in the number of sunspots was only noted in the 1840s by Heinrich Schwabe, originally an apothecary, but then turned full-time solar observer. The observations and data used in the Schwabe's conclusion that there was a ~ 10 year periodicity in sunspots are shown in Fig. 1 (Schwabe 1843). In fact, the original data set was complemented with seven extra years of observations by Schwabe included in Alexander von Humboldt's work, Cosmos, published in 1851. These extra years provided further evidence of the periodicity, by including one more maximum in the number of sunspots.

However, it was Rudolf Wolf (his portrait is shown in Fig. 2) who, following his own extensive observations in Switzerland, first in Bern, then in Zurich, established the standardized way to count sunspots, first called the Zurich or Wolf sunspot number. Wolf had been drawn to the study of sunspots following the work of Schwabe already in 1848. He then went on to put the quasi-cyclic variations in sunspot numbers on a firm, long-term basis by collecting, examining and standardizing past observations, as well as adding his own. The history of Wolf's work has been described by Izenman et al. (1983) who also has provided a critical statistical assessment of the "average" duration of the sunspot cycle derived by Wolf (11.11 years). Wolf's definition of relative sunspot numbers that he spent more than ten years to refine remained the accepted standard for measuring solar activity for over 100 years. Wolf combined counts of sunspot groups with those of individual spots. The historical data collected and refined by Wolf from the beginning of the 17th century

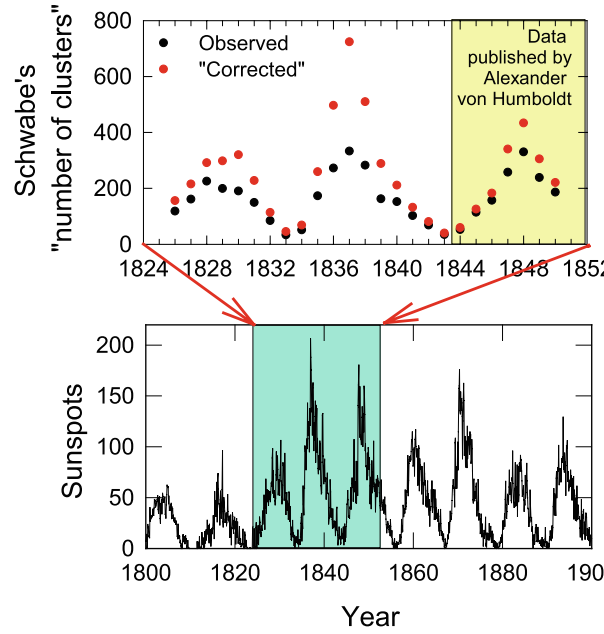


Fig. 1 The discovery of the ∼11 year solar activity cycle through sunspot observations. *Upper panel*: Schwabe's yearly sunspot group observations (from 1826 to 1844) and complemented to 1850 by Alexander von Humboldt. *Black symbols*: observed numbers, *red symbols*: corrected for the number of observing days in the year. *Lower panel*: Currently used (originally Wolf- or Zurich) monthly sunspot numbers for 1800 to 1900

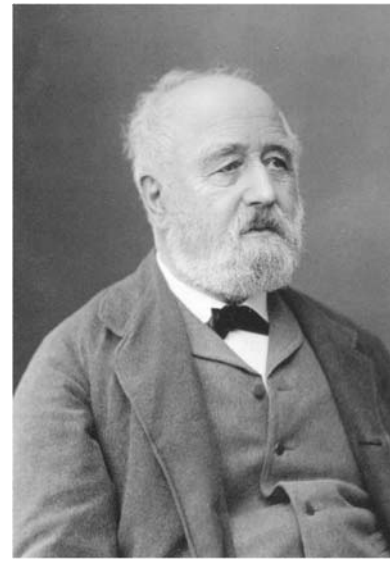


Fig. 2 Rudolf Wolf (1816–1893) who, as the Director first of the Observatory of the University of Bern then of the Observatory of the Eidgenössische Technische Hochschule (ETH) in Zurich made not only systematic sunspot observations over several decades, but also collected historic sunspot data and devised the still used measure of standard sunspot numbers

have become very important indicators to show that solar variability is very complex and not simply periodic. In Fig. 1, Schwabe's observations are put in the context of the record of the complete sunspot data in the 19th century, using the accepted Zurich monthly sunspot numbers.

During the Maunder minimum in the second half of the 17th century, the very small number of sunspots did not exhibit the cyclic behaviour that is now associated with what we know to be the 11-year periodicity (Eddy 1976). The lack of sunspots during the Maunder

minimum is well documented, observations covered about 95% of the total interval. The existence of other "grand minima" of long duration absence of sunspots in earlier epochs is inferred from proxy records, by such techniques as carbon dating and measuring isotopes in ice cores (see, for recent reviews, Usoskin et al. 2007; Usoskin 2008). We have no adequate theory to predict when the next grand minimum will occur. On these longer scales, solar activity and sunspots as its manifestation appear to behave in a chaotic/stochastic manner, and show no evidence of a cyclic behaviour (Spiegel 2009). The quasi-regularity of the 11-year cycle, when it persists for centuries, is therefore all the more remarkable. Some proxy records also support the continuation of the 11-year cycle through the Maunder minimum (Beer et al. 1998).

The linking of geomagnetic variations with the sunspot cycle is associated primarily with Edward Sabine in 1952 who was also leading an initiative for the establishment of geomagnetic observatories worldwide (see Cliver 1994 and references therein). In fact what Sabine noted was the coincidence of the 1843 minimum and 1848 maximum in sunspot numbers (see Fig. 1) with a minimum and maximum in geomagnetic storms at two widely separated geomagnetic observatories in Hobart (Tasmania) and Toronto (Canada). Others, including Rudolf Wolf, also noted the relationship between solar and geomagnetic variations (and other indicators such as the frequency of aurorae) that remained an unsolved puzzle for many decades after its discovery.

Another aspect of solar activity, introducing a concept of much shorter time-scale, explosive variability was discovered by Richard Carrington during his routine observations of sunspots (Carrington 1860). This observation, on 1 September 1859, was the first white-light, obviously very intense solar flare that was seen not only as a remarkable solar phenomenon, but also noted as being followed by a large geomagnetic storm (Cliver 2006). Flaring on the Sun is related to active regions and sunspot complexes, but while the connection is well established, the occurrence of very large flares, the kind that Carrington observed, is related to sunspots much less predictably. The terrestrial effects of large solar flares were noted, many decades after Carrington, by Hale (1931).

An observable effect of the Sun on terrestrial phenomena is related to the Sun's rotation. While this correlation, resulting in an apparent 27-day periodicity, was noted before, the key work on the details of the association were first published by Walter Maunder (Maunder 1904a), following an earlier compilation of observations by William Ellis. In particular, Maunder noted the association of the largest geomagnetic storms, with the shortest delay after the flares occurred when the sunspot group was within a privileged range of solar longitudes, between 19° East and 47° West of the central meridian, with a mean of 14° West. This was another association which remained a puzzle until the second half of the last century.

Another important indicator of solar variability that is associated with Edward Maunder is the evolution, in heliolatitude, of the location of sunspots as a function of heliolatitude as the solar cycle progresses (Maunder 1904b). At the time of minimum activity, there are a very few sunspots from the previous cycle close to the equator, and also a few spots associated with the new cycle at higher latitudes, at or poleward of 30°. As the number of sunspots increases towards solar maximum, they are seen to emerge progressively closer to the solar equator. Following maximum, the number of sunspots diminishes but the trend of approaching the equator continues. When the location of sunspots is plotted, the well-known butterfly diagram emerges. This is shown in Fig. 3 in which Maunder's original butterfly diagram is given, illustrating the location of sunspots from 1874 to 1902, as well as the equivalent 28 year interval including the sunspots (coded with their areas) up to the present.

Although the indications were recognized, the link between manifestations of solar variability and terrestrial, primarily geomagnetic, effects was missing, because solar magnetism

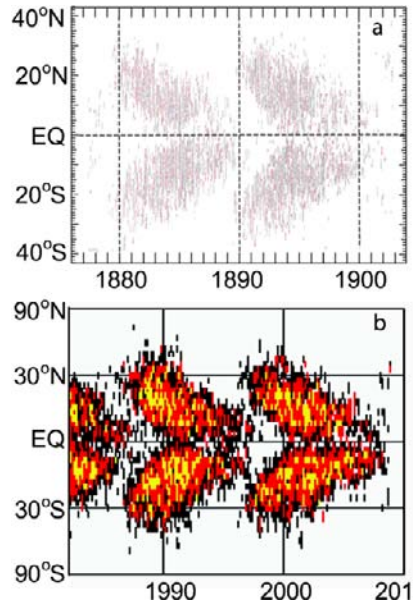


Fig. 3 The butterfly diagram of the location of sunspots in heliolatitude, shown as a function of time. The pattern has the clear periodicity of the solar cycle. *Upper panel*: The butterfly diagram as first published by Maunder (1904b). *Lower panel*: The modern butterfly panel over an interval of the same duration, colour coded with the area of the sunspots (courtesy: David Hathaway, NASA/MSFC)

had not been discovered. In fact, no less an authority than Lord Kelvin discarded the possibility of a remote effect of even a hypothetical magnetic field of the Sun on the Earth; the causal agents (interacting solar wind stream of different velocities, X-rays and energetic particles associated with flares and Coronal Mass Ejections) remained unknown until much later. And, in any case, there was no proof that the Sun was in any sense magnetic.

2 George Ellery Hale and the Discovery of Solar Magnetism

It is very satisfying to note that the discovery of solar magnetism was made by the new instrument, the spectroheliograph, invented and developed by George Ellery Hale in about 1891 (Hale 1891, 1929). Together with the very painstaking observations that helped solar physics make significant advances at the turn of the century, Hale started a new phase in solar research. He observed a very large flare or "eruption" as he called it, on 15 June 1892; this solar event was also followed within a day by a very large geomagnetic storm. However, Hale did not pursue the association: this was done, as already recounted, by Ellis, Maunder and others. The discovery of solar magnetism is now dated from Hale's paper on sunspots (Hale 1908), who used the spectral and imaging resolution of his instrument at Mount Wilson to detect Zeeman splitting in a number of spectral lines of sunspots that could only be due to the presence of strong magnetic fields.

In Hale et al. (1919) the findings of the earlier paper about the discovery of the magnetic fields were summarized as follows:

> photographs of the hydrogen flocculi made with the $H\alpha$ line showed clearly marked vortical structure in regions centering in sun-spots. This structure was found to be repeated in hundreds of spots, leaving no doubt as to the generality. … These photographs suggested the hypothesis that a sun-spot is a vortex, in which electrified particles, produced by ionization in the solar atmosphere, are whirled at high velocity. This might give rise to magnetic fields in sun-spots, regarded as electric vortices.

> A search for the Zeeman effect led to its immediate detection, and abundant proofs were soon found of the existence of a magnetic field in every sun-spot observed.

Although Hale attempted to explain his observations, in terms of vortices, in greater detail than was possible (with hindsight) from his observations, the key finding of magnetic fields, through the use of high resolution spectroscopic imaging for the detection of the Zeeman effect remains a lasting legacy of his and his colleagues' observations. In the same paper, the existence of different polarity sunspots was also noted.

Some ten years after his discovery that sunspots have magnetic fields, Hale published with Ellerman, Nicholson and Joy a paper on "The Magnetic Polarity of Sunspots" (Hale et al. 1919). This paper reported the reversal of polarity of sunspots over the 11-year cycle across the solar equator and the existence of opposite polarities in bipolar sunspot groups. This was confirmed by a more extensive study by Hale and Nicholson (1925). Thus, accounting for the reversal of polarity, the solar magnetic cycle is approximately 22 years.

The work that has led to what are now known simply as Hale's and Joy's law is very painstakingly reported in Hale et al. (1919). The study was based on 970 sunspots observed during the years 1915–1917. A very detailed magnetic classification of sunspots was established, a taxonomy that is primarily descriptive and spreads over a very large range of possible magnetic polarity arrangements in sunspot groups.

The details provided much raw material for further study of sunspots, although the key results, known now as Hale's and Joy's laws of sunspots originated from this work. These are illustrated in Fig. 4, modified from Fig. 6 in Hale et al. (1919), to make explicit both laws. Hale law states that the polarities in bipolar sunspots are always ordered so that the preceding spot has one polarity in the southern hemisphere, and has the opposite polarity in the northern hemisphere; similarly, following spots (that are of opposite polarity to the preceding spots) also have opposite polarities in the two hemispheres. However, the hemispheric ordering of polarities in bipolar groups reverses between solar cycles. Taken together with the Maunder butterfly diagram, this means that higher latitudes bipolar groups emerging as the new cycle begins, have the opposite ordering of polarities to the last bipolar groups (close to the equator, but still in the same hemisphere) of the previous cycle. Joy's law states that the preceding spot in a bipolar group is closer to the solar equator in both hemispheres.

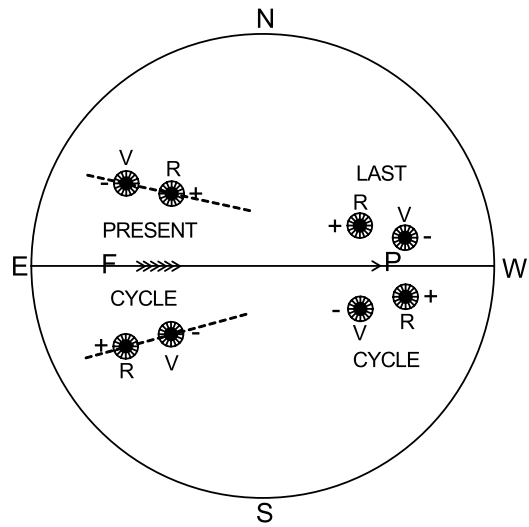


Fig. 4 The schematic illustration of Hale's and Joy's law of sunspots, as described in the text. This figure is a modified version of Fig. 6 published in Hale et al. (1919). The original caption reads: "*Diagram summarizing the results of polarity observations of sun-spots during the present and last cycles. The arrow indicates the Sun's rotation; the letters R* [red] *and V* [violet]*, the components of a normal triplet transmitted by the marked strip of the compound quarter wave plate; and the algebraic signs, the distribution of polarities between the preceding and following members of a bipolar group*"

Fig. 5 George Ellery Hale in 1918. Photograph taken in the Western Galleries of the Science Museum, London, showing Hale standing next to the 6 foot mirror of the Great Rosse Telescope

These laws are and have remained important observational building blocks of theories of the solar cycle.

Hale was clearly the towering figure of solar observations and discoveries in the first few decades of the 20th century. Quite apart from solar physics, Hale left an abundant legacy: the Yerkes Astronomical Observatory of the University of Chicago, the Mount Wilson Observatory, the 200-inch Mount Palomar telescope, the Astrophysical Journal that he founded (Hale 1895) remain the most memorable of his many achievements (Adams 1938, 1939). He was highly respected as a scientist, as a leader of research and research institutions and as a very effective fund-raiser for science. In Fig. 5, he is shown standing in London's Science Museum, next to the 6-foot mirror of the Great Rosse Telescope in 1918, shortly after this once-famous telescope was decommissioned, having been at the forefront of astronomical investigations in the mid-19th century through the study of nebulae and galaxies (Denning 1914).

The explanation, in terms of electric vortices and upwelling spirals of gases from sunspots that then spread out on the neighboring photospheric surface, is of course now dated both observationally and conceptually. Hale argues in his original paper (Hale 1908), not surprisingly, that higher optical and spectral resolutions are needed to make further progress. However, the discovery that strong magnetic fields existed in sunspots allowed a major step forward in solar research, as well as in starting to unravel the then still unknown links between solar and terrestrial phenomena. The systematic study of sunspot polarities and their ordered relationship to the solar cycle also provided straightaway the link between magnetic fields and the solar cycle.

3 Toward Today's Research in Solar Magnetism: Eugene Parker

It is symbolic that the first George Ellery Hale Prize of the Solar Physics Division of the American Astronomical Society was awarded to Eugene Parker in 1978, for his "imaginative and stimulating contributions in which plasma and magnetohydrodynamical physics have been applied to astronomy". When accepting the Prize, Parker delivered a lecture on "George Ellery Hale and active magnetic fields" (Parker 1979a). In the same year, following earlier work (Parker 1955a) on the topic, Parker published a highly influential series of nine articles under the general title of "Sunspots and the physics of magnetic flux tubes" (the first of which was Parker 1979b). Eugene Parker became the dominating figure in solar magnetism in the second half of the 20th century; Figure 6 shows Parker at the time of receiving one of his other awards, the Kyoto medal in 2003. The work of Parker on solar and stellar magnetism, including the generation and transport mechanisms spreads over more than fifty years and has been highly influential to the present (Parker 1955b, 1957, 2009).

In the summary of his acceptance lecture, Parker (1979a) identified many of the concepts concerning solar magnetism that would have been quite unknown to Hale and his contemporaries, such as turbulent diffusion in the convective envelope of the Sun; buoyancy that brings magnetic flux tubes to the surface; and the intrinsic non-equilibrium of the magnetic fields as they arise through the photosphere and lead to the many surface and coronal manifestations of solar activity. The underlying large-scale, longer-term mechanisms for magnetic-field generation that lead to the 11-year sunspot cycle, the 22-year magnetic cycle and other, longer quasi-periodic or chaotic intervals of variability were well beyond the conceptual framework in which Hale discussed his findings of magnetic fields in sunspots. Parker (1979a) suggested that "*the sunspot, and its magnetic field discovered by Hale, is a particularly vexing phenomenon, having resisted any overall self-consistent explanation up to the present time*". He went on to suggest further that "*the sunspot is merely the magnetic*

Fig. 6 Eugene Parker in 2003 when he was awarded the Kyoto Medal. His contributions to understanding solar magnetism, the solar cycle and the solar wind have had a profound and lasting influence on these scientific disciplines

debris on the surface of the Sun, marking the position over an unseen subsurface downdraft".

Probably neither Hale, nor modern observers of sunspots would agree with this provocatively dismissive statement, however well-intentioned to bring out a hierarchy of phenomena in solar magnetism. Progress in imaging sunspots has yielded photographs of increasing resolution; one example by the Swedish Solar Telescope is shown in Fig. 7 (upper panel). At the same time, the equally high spectral resolution allows imaging in the wings of spectral lines and thus use the Zeeman effect, as had been done by Hale, and take the differences to determine the map of the magnetic fields. This is also illustrated in Fig. 7 (lower panel) for the same sunspot group shown in the upper panel. Today's ground-based and spacecraft instruments for imaging the details of the Sun's magnetic field provide observations with high resolution on the solar surface and chromosphere, not only in or around sunspots, but

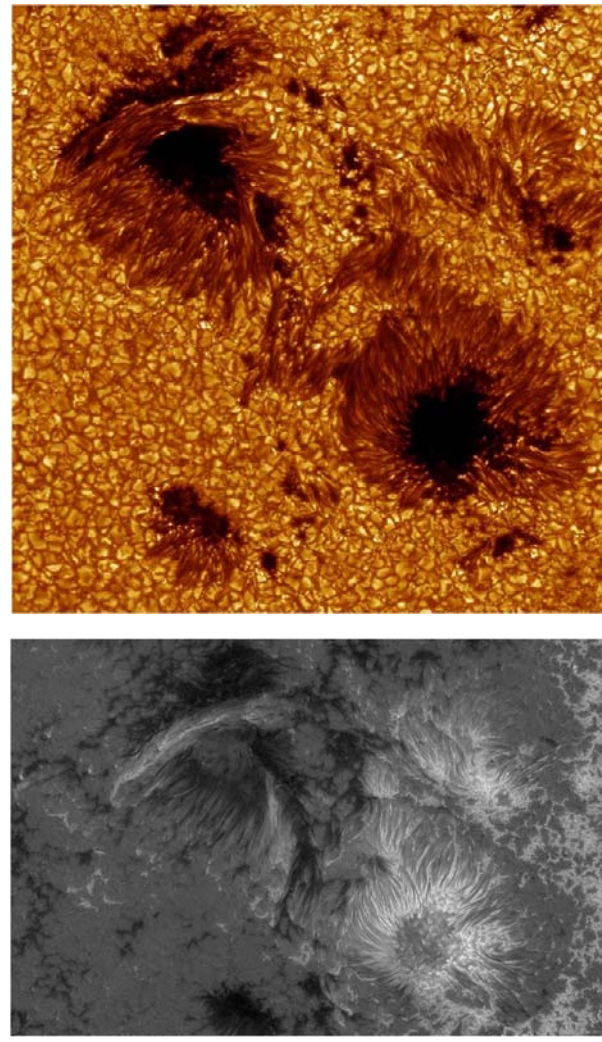


Fig. 7 *Upper panel*: a sunspot group imaged near the central meridian by Göran Scharmer and Kai Langhans of the Institute of Solar Physics of the Swedish Royal Academy of Sciences using the 1-m Swedish Solar Telescope. *Lower panel*: magnetogram associated with the same sunspot group at 630.2 nm (Fe I), in which white and dark areas represent the two opposite magnetic polarities

across the whole disk where the strength of the magnetic field is usually orders of magnitude weaker than in sunspots.

In addition to the direct observations that are similar to those of Hale, even though with a resolution that has increased by several orders of magnitude, the development of indirect probing of the Sun's interior using helioseismology has allowed giant steps in the understanding of solar magnetism (for a review, see Thompson 2006). This development was foreshadowed in Parker's (1979a) comments in which he refers to the then (and still now, in many cases) outstanding questions in terms of the processes in the Sun's convective interior.

Another development that needs to be mentioned concerns the link between the Sun and the terrestrial effects that were identified but which remained unresolved since the mid-19th century. The primary source for our concept of the solar wind is the paper, exactly half a century ago, by Eugene Parker (1958) in which he proposed the continuous supersonic outflow of plasma from the Sun. Parker's theory was in response, in part, to difficulties in explaining the existence of the very hot solar corona and its outer boundary conditions. In part, Parker's theory also addressed the observations of cometary tails by Biermann (1951) that showed what appeared to be a continuous outflow of plasma from the Sun in all directions. Parker's theory was controversial, but only for a short interval, as the first interplanetary spacecraft quickly proved the general correctness of Parker's theory. (For a detailed account of the context in which the solar wind concept was born and developed, see Parker 2001, as well as references therein.)

A natural component of Parker's solar wind theory is the heliospheric magnetic field, originating in the solar corona and dragged out in the highly conducting, radially flowing solar wind. Combined with the rotation of the Sun, the general three dimensional spiral pattern of the magnetic field has been fully confirmed and mapped in three dimensions by spacecraft observations (Forsyth et al. 2002; Smith 2007).

With the discovery of the solar wind and then with additional solar, interplanetary and terrestrial observations, many of the puzzles related to the Sun, its activity and its terrestrial effects could be resolved, about a century or more after Wolf, Sabine, Maunder and others. It is now understood that the link (unknown to Lord Kelvin and his contemporaries) is provided by the heliospheric magnetic field that acts as the propagation path to transport energetic particles from the Sun, from solar eruptions. At the same time, the explosive coronal mass ejections (also originating in the complex and frequently unstable magnetic configurations in the corona) propagate through the solar wind and, if directed towards the Earth, will produce geomagnetic storms and aurorae.

4 Solar Magnetism: The Current Status

Even though sunspots remain the longest-standing direct record, other aspects of solar activity also vary over the course of the solar cycle. These include the complexity of the coronal magnetic field, and the frequency of explosive events including flares and coronal mass ejections. In our technological society, we are increasingly susceptible to the consequences of the impacts of the particles and radiation from such explosive events. Power networks and communication satellites can be knocked out by them, and the hazard to humans in space is very real. For these practical considerations, therefore, as well as scientific interest, we need to understand solar magnetism and to develop our predictive capabilities in respect of solar activity.

Since the time of George Ellery Hale and his pioneering work, we have learned that magnetism controls the many manifestations of solar activity. Yet a complete understanding

Fig. 8 Participants in the ISSI Workshop on "The origin and dynamics of solar magnetism". (1) André Balogh, (2) Eugene Parker, (3) Roger-Maurice Bonnet, (4) Alan Title, (5) Mrs Title, (6) Bob Forsyth, (7) Ruedi von Steiger, (8) Ed Spiegel, (9) Sami Solanki, (10) Nigel Weiss, (11) Hannah Schunker, (12) Jan Stenflo, (13) Matthias Rempel, (14) Alfred de Wijn, (15) David Hathaway, (16) Axel Brandenburg, (17) Peter Gilman, (18) Laurent Gizon, (19) Mausumi Dikpati, (20) Lidia van Driel-Gesztelyi, (21) Mike Thompson, (22) Emiliya Yordanova, (23) Karel Schrijver, (24) Bruce Lites, (25) Len Culhane, (26) Nadège Meunier, (27) Yi-Ming Wang, (28) Saku Tsuneta, (29) Alan Hood, (30) Åke Nordlund, (31) Sasha Kosovichev, (32) Laura Bone, (33) Steve Tobias, (34) Jean-Paul Zahn, (35) Peter Cargill, (36) Sonia Danilovic, (37) Michal Svanda, (38) Junwei Zhao, (39) Rolf Schlichenmaier, (40) Andreas Lagg, (41) Laurène Jouve, (42) Sacha Brun

of solar magnetism is missing. At the threshold of the present millennium, solar magnetism was named in a survey by Physics World magazine as one of the ten great unsolved problems in physics (Physics World, Dec. 1999 edition). The centenary of Hale's discovery seemed an opportune time to take stock of the many advances that have been made in understanding the origins and behaviour of the Sun's magnetic field and to discuss intensively the problems that remain and how upcoming observations and future theoretical developments will address them.

The workshop on The Origin and Dynamics of Solar Magnetism held at the International Space Science Institute on January 21st–25th 2008 addressed all major aspects of the Sun's magnetism: the topics are reflected in the written papers presented in this volume. A group photo of the participants in the workshop is shown in Fig. 8. Observations and theory pertinent to understanding the solar dynamo which presumably generates the magnetic field are discussed in the papers by Weiss and Thompson (2009), Dikpati and Gilman (2009), Brandenburg (2009) and Tobias (2009). The large-scale dynamics represented by flows in the Sun's outer convective envelope are addressed in the papers by Meunier and Zhao (2009) and Brun and Rempel (2009). Papers by Kosovichev (2009) and Lites (2009) discuss how the magnetic flux emerges at the solar photosphere and how it is transported. Observations and theoretical interpretation of sunspots are discussed by Schlichenmaier (2009), Scharmer (2009) and, from a helioseismic perspective, by Gizon et al. (2009). Since the SoHO satellite observations, it is now appreciated that not only does the solar magnetic field manifest itself in sunspots, it is also present in fast-changing, small-scale magnetic flux known as the magnetic carpet: small-scale fields are discussed in the paper by de Wijn et al. (2009). Moving further out from the Sun, to understand the impacts on interplanetary space and the Earth environment one must understand the coupling of the magnetic field in the photosphere to the overlying chromosphere and corona, and the linkage to the heliosphere beyond: these aspects are discussed in the papers by Wedemeyer-Böhm et al. (2009), van Driel-Gesztelyi and Culhane (2009) and Wang (2009). Also key to mitigating effects of solar activity on human activity is the art of forecasting, which is discussed by Hathaway (2009). Perspectives on what has been understood and what still challenges us are provided by Parker (2009) and Cargill (2009). The paper by Christensen et al. (2009) gives an insight into other cosmic magnetic fields, namely those in planetary interiors, from a solar perspective. The workshop is summarized by the paper by Zahn (2009).

Acknowledgements The participants in the Workshop on "The origin and dynamics of solar magnetism" expressed their thanks to the staff of the International Space Science Institute for the very productive atmosphere in which the talks and the discussions were held. The Editors acknowledge the very willing cooperation of all the authors in this issue and their thanks to all those who took part in the review process.

References

W.S. Adams, George Ellery Hale 1868–1938. Astrophys. J. **87**, 369–388 (1938)
W.S. Adams, Biographical memoir of George Ellery Hale, 1868–1938. Biogr. Mem. Nat. Acad. Sci. **21**, 181–241 (1939)
J. Beer, S. Tobias, N. Weiss, An active Sun throughout the Maunder Minimum. Solar Phys. **181**, 237–249 (1998)
L. Biermann, Kometenschweife und solare Korpuskularstrahlung. Zeitschrift für Astrophysik **29**, 274–286 (1951)
A. Brandenburg, Advances in theory and simulations of large-scale dynamos. Space Sci. Rev. (2009, this issue)
A.S. Brun, M. Rempel, Large scale flows in the solar convection zone. Space Sci. Rev. (2009, this issue)
P.C. Cargill, Coronal magnetism: difficulties and prospects. Space Sci. Rev. (2009, this issue)
R.C. Carrington, Description of a singular appearance seen in the Sun on September 1, 1859. Mon. Nat. R. Astron. Soc. **20**, 13–15 (1860)
U.R. Christensen, D. Schmitt, M. Rempel, Planetary dynamos from a solar perspective, Space Sci. Rev. (2009, this issue)
E.W. Cliver, Solar activity and geomagnetic storms: the first 40 years. EOS Trans. AGU **75**, 574–575 (1994)
E.W. Cliver, The 1859 space weather event: Then and now. Adv. Space Res. **38**, 119–129 (2006)
A.G. de Wijn, J.O. Stenflo, S.K. Solanki, S. Tsuneta, Small-scale solar magnetic fields. Space Sci. Rev. (2009, this issue)
W.F. Denning, Lord Rosse's telescope. The Observatory **37**, 347–348 (1914)

M. Dikpati, P.A. Gilman, Flux-transport solar dynamos. Space Sci. Rev. (2009, this issue)
J.A. Eddy, The Maunder minimum. Science **192**, 1189–1202 (1976)
R.J. Forsyth, A. Balogh, E.J. Smith, The underlying direction of the heliospheric magnetic field through the Ulysses first orbit. J. Geophys. Res. **107**(A11), SSH19-1 (2002). doi:10.1029/2001JA005056
L. Gizon, H. Schunker, C.S. Baldner, S. Basu, A.C. Birch, R.S. Bogart, D.C. Braun, R. Cameron, T.L. Duvall, S.M. Hanasoge, J. Jackiewicz, M. Roth, T. Stahn, M.J. Thompson, S. Zharkov, Helioseismology of sunspots: a case study of NOAA region 9787. Space Sci. Rev. (2009, this issue)
G.E. Hale, Note on solar prominence photography. Astron. Nachricht. **126**, 81 (1891)
G.E. Hale, The Astrophysical Journal. Astrophys. J. **1**, 80–83 (1895)
G.E. Hale, On the probable existence of a magnetic field in sunspots. Astrophys. J. **28**, 315–343 (1908)
G.E. Hale, The spectrohelioscope and its work, Part 1. History, instruments, adjustments and methods of observation. Astrophys. J. **70**, 265–311 (1929)
G.E. Hale, The spectrohelioscope and its work, Part III. Solar eruptions and their apparent terrestrial effects. Astrophys. J. **73**, 379–412 (1931)
G.E. Hale, F. Ellerman, S.B. Nicholson, A.H. Joy, The magnetic polarity of sun-spots. Astrophys. J. **49**, 153–185 (1919)
G.E. Hale, S.B. Nicholson, The law of sun-spot polarity. Astrophys. J. **62**, 270–300 (1925)
D. Hathaway, Solar cycle forecasting. Space Sci. Rev. (2009, this issue)
D.V. Hoyt, K.H. Schatten, *The Role of the Sun in Climate Change* (Oxford University Press, Oxford, 1997)
A.J. Izenman, J.R. Wolf, H.A. Wolfer, An historical note on the Zurich sunspot relative numbers. J. R. Stat. Soc. A **146**, 311–318 (1983)
A. Kosovichev, Photospheric and subphotospheric dynamics of emerging magnetic flux. Space Sci. Rev. (2009, this issue)
B.W. Lites, The topology and behavior of magnetic fields emerging at the solar photosphere, Space Sci. Rev. (2009, this issue)
E.W. Maunder, The "great" magnetic storms, 1875 to 1903, and their association with sun-spots as recorded at the Royal Observatory, Greenwich. Mon. Nat. R. Astron. Soc. **64**, 205–222 (1904a)
E.W. Maunder, Note on the distribution of sunspots in heliographic latitude, 1874 to 1902. Mon. Nat. R. Astron. Soc. **64**, 747–761 (1904b)
N. Meunier, J. Zhao, Observations of photospheric and interior dynamics and magnetic fields: from large-scale to small-scale flows. Space Sci. Rev. (2009, this issue)
E.N. Parker, The formation of sunspots from the solar toroidal field. Astrophys. J. **121**, 491–507 (1955a)
E.N. Parker, Hydromagnetic dynamo models. Astrophys. J. **122**, 293–314 (1955b)
E.N. Parker, The solar hydromagnetic dynamo. Proc. Nat. Acad. Sci. USA **43**, 8–14 (1957)
E.N. Parker, Dynamics of the interplanetary gas and magnetic fields. Astrophys. J. **128**, 664–676 (1958)
E.N. Parker, George Ellery Hale and active magnetic fields. Bull. Am. Astron. Soc. **12**, 423 (1979a)
E.N. Parker, Sunspots and the physics of magnetic flux tubes, I—The general nature of the sunspot. II—Aerodynamic drag. Astrophys. J. **230**, 905–923 (1979b)
E.N. Parker, A history of the solar wind concept, in *Century of Space Science*, ed. by J.A.M. Bleeker, J. Geiss, M.C.E. Huber (Kluwer, Dordrecht, 2001), pp. 225–255
E.N. Parker, Solar magnetism: the state of our knowledge and ignorance. Space Sci. Rev. (2009, this issue)
G. Scharmer, Recent evidence for convection in sunspot penumbrae. Space Sci. Rev. (2009, this issue)
R. Schlichenmaier, Sunspots: from small-scale inhomogeneities towards a global theory. Space Sci. Rev. (2009, this issue)
H. Schwabe, Solar observations during 1843, Astron. Nachricht. 20(495) (1843)
E.J. Smith, The global heliospheric magnetic field, in *The Heliosphere Through the Solar Activity Cycle*, ed. by A. Balogh, L.J. Lanzerotti, S.T. Suess (Springer, Chichester, 2007), pp. 79–150
E.A. Spiegel, Chaos and intermittency in the solar cycle. Space Sci. Rev. (2009, this issue)
M.J. Thompson, Magnetohelioseismology. Phil. Trans. R. Soc. A **364**, 297–311 (2006)
S.M. Tobias, The solar dynamo: The role of penetration, rotation and shear on convective dynamos. Space Sci. Rev. (2009, this issue)
I.G. Usoskin, A history of solar activity over millennia. Living Rev. Sol. Phys. **5**, 3 (2008). http://www.livingreviews.org/lrsp-2008-3
I.G. Usoskin, S.K. Solanki, G.A. Kovaltsov, Grand minima and maxima of solar activity: new observational constraints. Astron. Astrophys. **471**, 301–309 (2007)
L. van Driel-Gesztelyi, J.L. Culhane, Magnetic flux emergence, activity, eruptions and magnetic clouds: Following magnetic field from the Sun to the heliosphere, Space Sci. Rev. (2009, this issue)

Y.-M. Wang, Coronal holes and open magnetic flux. Space Sci. Rev. (2009, this issue)
S. Wedemeyer-Böhm, A. Lagg, Å. Nordlund, Coupling from the photosphere to the chromosphere and the corona, Space Sci. Rev. (2009, this issue)
N.O. Weiss, M.J. Thompson, The solar dynamo. Space Sci. Rev. (2009, this issue)
J.-P. Zahn, Conclusions: Solar magnetism today. Space Sci. Rev. (2009, this issue)

Solar Magnetism: The State of Our Knowledge and Ignorance

E.N. Parker

Originally published in the journal Space Science Reviews, Volume 144, Nos 1–4, 15–24.
DOI: 10.1007/s11214-008-9445-x

Abstract We review some longstanding scientific mysteries related to solar magnetism, with final attention to the mystery of the "turbulent diffusion" essential for the theoretical $\alpha\omega$-dynamo that is believed to be the source of the magnetic fields of the Sun. Fundamental difficulties with the concept of turbulent diffusion of magnetic fields suggest that the solar dynamo problem needs to be reformulated. An alternative dynamo model is proposed, but it remains to be shown that the model can provide the quantitative aspects of the cyclic magnetic fields of the Sun.

Keywords Solar dynamo · Turbulent diffusion of magnetic fields

1 Introduction

Observational knowledge of solar magnetism began with Hale's (1908) detection and measurement of sunspot fields. The subject has expanded enormously over the last fifty years, beginning with the Babcock solar magnetograph (Babcock and Babcock 1955). Observations are currently probing the astonishing world of magnetic microactivity, with exciting new results from Hinode and the accumulating results from ground based observatories, as well as the SOHO, ACE, TRACE, etc. spacecraft. Knowledge is advancing rapidly as technical ingenuity provides increasing spatial resolution and a direct look into the microactivity. It is to be expected that the next ten years may provide observations down to the basic scales of 10–20 km, from which much of the activity is driven. An inevitable consequence of these observational advances, of course, is the discovery of many new puzzling phenomena, to be added to the scientific puzzles of long standing. Leighton (1969) remarked many years ago that "were it not for magnetic fields, the Sun would be as uninteresting as most astronomers seem to think it is"; this is the stuff that makes science so fascinating. So it is the purpose of this writing to review some of our outstanding ignorance of the physics of solar magnetism and to give an example of a fresh approach to the solar dynamo.

E.N. Parker (✉)
University of Chicago, Chicago, USA
e-mail: parker@oddjob.uchicago.edu

M.J. Thompson et al. (eds.), *The Origin and Dynamics of Solar Magnetism*.
DOI: 10.1007/978-1-4419-0239-9_2

Sunspots are the classic example of long standing ignorance, well known to observation for centuries, possessing fascinating internal fine structure, and still not understood from the basic laws of physics. Why is the Sun compelled by the basic laws to form sunspots at all? What causes the subsurface convection to sweep thousands of magnetic fibrils together and then to compress them into two or more kilogauss (Meyer et al. 1974; Meyer et al. 1977; Parker 1979)?

The sunspot dilemma has been extended by the observational discovery that there are red dwarf stars exhibiting cool patches covering half the visible disk. We presume these cool areas to be a magnetic phenomenon akin to sunspots. However, failing to understand the sunspot, we can say no more.

We should not fail to recognize that the remarkable internal rotation pattern of the Sun, established from helioseismology through the GONG Observatory, is still without explanation in terms of the hydrodynamics of the convection zone. Can magnetic forces be involved in some way? We do not know enough about the magnetic fields beneath the surface of the Sun to answer that question at the moment. Indeed the origin of those magnetic fields is one of the basic challenges to solar physics—and to astrophysics in general. We will have more to say on that subject. For the present, note that the magnetic fields of the Sun are in an inexplicable (Parker 1984) fibril state where they extend through the visible surface, causing us to wonder to what extent they are fibril throughout the interior.

These scientific puzzles, and many others, have been with us for years, and we are beginning to be haunted by Wigner's dictum: The important problems in physics are rarely solved; they are either forgotten or declared to be uninteresting. The hope is that sufficient attention to the problem may ultimately evade Wigner's dictum. Consider, then, the origin of the magnetic fields of the Sun, as a place to begin the discussion.

2 Solar Magnetic Fields

We possess a vast archive of observations on the magnetic field at the visible surface of the Sun, providing boundary conditions for the origin of the field but giving no clear picture of the magnetic fields below the surface, concerning which we have many questions. There has been substantial progress in modeling various forms of the presumed $\alpha\omega$-dynamo (combining the effects of the cyclonic convection and the nonuniform rotation) responsible for the 22-year magnetic cycle of the Sun, giving some idea of the available theoretical possibilities. Thanks to GONG, we know the internal profile of the angular velocity $\omega(r,\theta)$ of the Sun. However, the other contributing effects, including the basic cyclonic nature of the updrafts in the convective zone, the meridional circulation across the upper and lower convective zone (Choudhuri et al. 1995; Gilman and Miesch 2004; Dikpati et al. 2006), the nature of the turbulent diffusion of the azimuthal and the poloidal magnetic fields (Rüdiger and Kitchatinov 1997), and the proper boundary conditions at the surface of the Sun, are not adequately known from observation, nor are they available from theoretical considerations. So the usual procedure is to imagine a mathematical form that seems to describe each of the individual effects (Kitiashvili and Kosovichev 2008). Then we represent the strength of each effect with an appropriate parameter, i.e. coefficient, and proceed with the resulting $\alpha\omega$-dynamo equations to model our "solar dynamo" (Parker, 1955b, 1957).

The ensuing numerical or analytical simulations show how the several independent effects combine to provide the generation of magnetic field in forms resembling the observed solar magnetic fields at the visible surface. What is more, one finds that the values of the

individual parameters necessary to fit the observations do not appear to be unreasonable. It is gratifying to see that this exploratory approach over the years has provided a variety of circumstances that might provide the actual magnetic fields seen on the Sun. So there has been great progress, and some have been emboldened to apply the same $\alpha\omega$-dynamo concepts to others stars, to accretion disks, and to the Galaxy.

However, it must be recognized that these gratifying achievements are really only the first major step in establishing a scientific theory of the origin of the magnetic fields of the Sun, and by implication, of the other magnetic objects to be found in the astronomical universe. The physics implied by the appropriate values of the parameters must also be understood. That there exist values of the parameters such that a mathematical simulation can be made to conform to the observational facts is not sufficient for a scientific understanding of the solar dynamo. For there is the annoying revelation: Given four free parameters you can provide a gratifying fit to the New York, or Beijing, skyline.

To look into the problem of understanding the physics of the dynamo parameters, consider the magnetic diffusion coefficient η in the conventional $\alpha\omega$-dynamo equations. In order to construct a model of the solar dynamo with the observed latitudinal scales of the order of $l = 2 \times 10^{10}$ cm and time scales of a few years, say $t = 10^8$ sec, we must have l^2 and $4\eta t$ comparable in magnitude, from which it follows that $\eta = 10^{12}$ cm^2/sec, in order of magnitude. So large a diffusion coefficient can be understood only in terms of turbulent diffusion. The standard mixing length theory of turbulence gives a diffusion coefficient $\eta = \frac{1}{3}u\lambda$, where u is the characteristic velocity of the dominant eddies, which have a scale λ. The granules and the supergranules are the dominant eddies at the visible surface of the Sun, and they contribute about equally. For instance, the granule scale is of the order of 3×10^7 cm and the characteristic velocity is 10^5 cm/sec, providing the desired 10^{12} cm^2/sec. Low in the convective zone we expect λ to be comparable to the pressure scale height, of the order of 5×10^9 cm, where the convective velocities are of the order of 10^3 cm/sec (Spruit 1974), again yielding η close to 10^{12} cm^2/sec. So what is the problem?

The problem is that there is no physical concept for the turbulent diffusion of a vector field. The mixing length formula was constructed on the basis of the turbulent diffusion of a scalar field, e.g. smoke or ink, where the scalar field swirls around with the fluid, preserving the density of the scalar field as each initial element of fluid is drawn into a thin filament. In contrast, a vector field caught up in the turbulent fluid behaves quite differently, because the magnetic flux is conserved as the field is stretched out along each filament. Hence the strength of the field grows more or less exponentially with time, just as the length L of the filament grows exponentially with time while the thickness h decreases exponentially with time. We have

$$\frac{dL}{dt} = \frac{u}{\lambda}L,$$

so that

$$L(t) = L(0)\exp\frac{ut}{\lambda}.$$

Neglecting the slight increase in the width of a filament, an initial volume of dimension λ is stretched into a filament of length $\lambda\exp(+ut/\lambda)$ and thickness $\lambda\exp(-ut/\lambda)$, with the initial magnetic field $B(0)$ becoming something of the general order of $B(0)\exp(+ut/\lambda)$. This goes on until the thickness $h(t)$ becomes so small that the resistive diffusion time $h(t)^2/4\eta_R$ becomes comparable to the characteristic stretching time λ/u, where η_R is the usual resistive diffusion coefficient. At that point the magnetic field in each filament merges with the fields in the randomly oriented neighboring filaments, and the field is obliterated. In terms of the

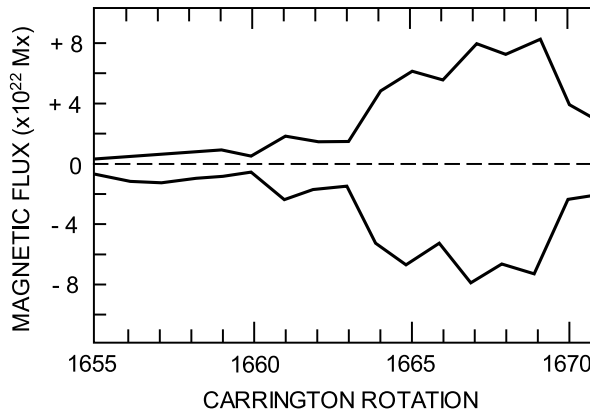


Fig. 1 The magnetic flux (separately for + and −polarities) in units of 10^{22} maxwells in a bipolar magnetic region as a function of Carrington rotation, from Gaizauskas et al. (1983)

magnetic Reynolds number $R_M \equiv u\lambda/\eta_R$ this happens when the field intensity has increased from its initial intensity by a factor of the general order of $R_M^{1/2}$. For conditions in the Sun R_M is a large number, of the order of 10^6 or more. The magnetic stresses increase by a factor of the order of R_M. So for anything but a very small initial field, the magnetic stresses overwhelm the turbulence long before the resistive diffusion has a chance to obliterate the field. In fact, in the convective zone of the Sun the estimated mean azimuthal magnetic field is nearly as large, if not substantially larger, than the equipartition field of the turbulent convection. So the convection would be constrained to be little more than Alfvén waves in the azimuthal field, and providing nothing like turbulent mixing of magnetic fields. The idea that the azimuthal magnetic field is subject to ordinary turbulent diffusion, with $\eta = \frac{1}{3}u\lambda$, seems unjustified.

Consider, then, a minimum estimate of the azimuthal magnetic field strengths to be encountered in the convective zone of the Sun. It is generally believed that the azimuthal magnetic field of the Sun is created in the tachocline—the layer of intense shear at the bottom of the convective zone. The magnetic field is buoyant and subject to upward eruption, forming the Ω-loops that emerge at the visible surface to form the active bipolar magnetic regions. Gaizauskas et al. (1983) monitored the magnetic flux in a long lived bipolar region, or activity complex, existing at a fixed solar longitude for over a year. Figure 1 is their plot of the visible magnetic flux over an 8 month period. The flux reached a maximum of about 8×10^{22} maxwells, suggesting that the azimuthal magnetic field that spawned the bipolar region contained a flux of at least that much.

Suppose that the band of azimuthal magnetic field at the bottom of the convective zone has a width comparable to the depth of the convective zone, 2×10^{10} cm. Assume that the band has a thickness of 2×10^9 cm, somewhat thicker than the tachocline. It follows that the average magnetic field strength is 2×10^3 gauss, to be compared with the equipartition field $(4\pi\rho u^2)^{1/2}$ that has a maximum of about 3×10^3 gauss low down in the convective zone, from where it drops to zero across the tachocline. It is evident that the convection may distort the mean field in substantial ways, but the turbulent mixing process is largely stymied, and it appears that the concept of turbulent diffusion of magnetic fields cannot be applied to the azimuthal magnetic field in the convective zone of the Sun. There is no way to account for the value $\eta \approx 10^{12}$ cm^2/sec, suggesting that it is necessary to rethink the $\alpha\omega$-dynamo for the Sun.

It should be noted that the same dilemma, with the necessary turbulent diffusion of 10^{25} cm^2/sec, arises for the Galaxy. The galactic magnetic field, of the order of 4×10^{-6}

gauss in the spiral arms, is equal to the equipartition field for the observed interstellar turbulence.

3 Thoughts on the Solar Dynamo

For some years I have looked for some theoretical scheme for which there might be "turbulent-like" diffusion of the azimuthal magnetic field of the Sun. The fibril structure of the magnetic field at the visible surface suggests the possibility of a fibril structure in the azimuthal field, although there is no theory for the formation of fibrils. However, I have not been able to show that rapid reconnection between intense fibrils is capable of providing the desired diffusion, because the magnetic Reynolds number at the distant tachocline, where the field is presumed to reside, is 10^{10}, or more. Other ideas have been explored, but are equally unpromissing. So it would appear that our theorizing has established a dilemma, bringing us to Willie Fowler's dictum: The abhorrent aspect of physics is the brutal murder of beautiful theories by ugly facts.

So let us go back to the beginning and start over, taking advantage of the knowledge gained since the $\alpha\omega$-dynamo was first conceived over fifty years ago. An essential point is the enormous intensity attributed to the azimuthal magnetic field at the base of the convective zone. Dynamical studies (Choudhuri and Gilman 1987; D'Silva 1993; D'Silva and Choudhuri 1993; Fan et al. 1993; Fan et al. 1994; Schüssler et al. 1994) show that the azimuthal magnetic field must be as strong as $0.5–1.0 \times 10^5$ gauss at the bottom of the convective zone if the buoyant rise of Ω-loops is to avoid domination by the Coriolis force.

D'Silva points out some special circumstances that may require no more than 10^4 gauss. Fixing the footpoints of the Ω-loop during the rise is one condition, and very small diameters ($<10^2$ km) for the individual flux bundles is another.

That is to say, the rise to the surface must occur in a fraction of the 25 day rotation period of the Sun in order (a) to limit the inclination, or tilt, of the bipolar regions (to the east–west direction) to the observed values of $\sim 10°$, and in order (b) that the apex of the Ω-loop rises more or less vertically rather than parallel to the spin axis of the Sun. With such intense azimuthal magnetic fields the cross section of the minimum azimuthal 8×10^{22} maxwells is only about 10^{18} cm^2 or 10^9 cm $\times$ 10^9 cm. This suggests that there may be intense flux bundles at the bottom of the convective zone occupying only a small fraction of the whole volume. The inference of 10^5 gauss flux bundles lies entirely outside the traditional view of the solar dynamo that created our dilemma. So we begin the rethinking process with the idea that there are one or more bundles of azimuthal field of 10^5 gauss lying at the tachocline, at the bottom of the convective zone.

Such a band of azimuthal field is unstable to the buoyancy of the field, bulging upward to form Ω-loops at various positions around the Sun (Parker 1955a; Schüssler et al. 1994) (sketched in Fig. 2). Once a rising Ω-loop starts to move up from the bottom of the convective zone it is strongly boosted by the convective instability of the super adiabatic temperature gradient. The upper portion the Ω-loop expands as it rises, of course, with the field intensity evidently falling below 10^3 gauss as it approaches the surface (Zwaan 1985). Matsumoto et al. (1993) suggest that the fibril structure of the field develops as the Ω-loop rises. Upon arriving at the visible surface the Ω-loop produces the familiar bipolar magnetic regions.

Sheeley (1966), Sheeley et al. (1989), Wilson et al. (1990) have studied the dispersal and subsequent polar transport of the emerging magnetic fields. They find that the flux bundles

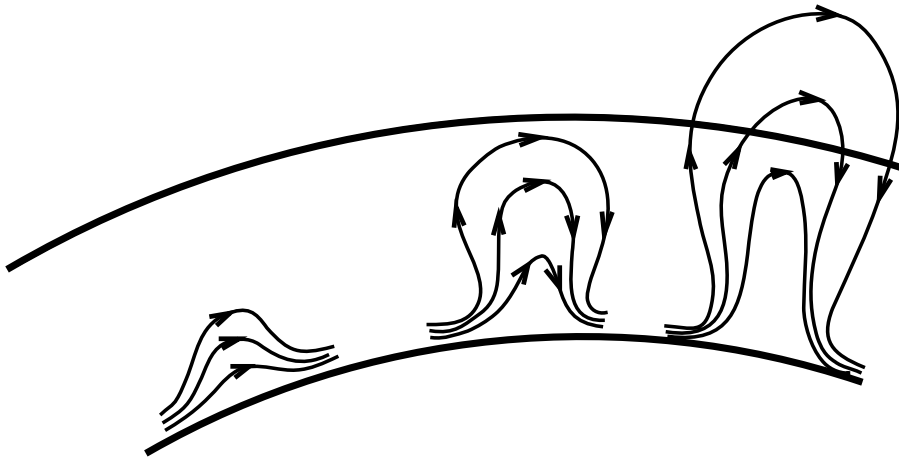


Fig. 2 Sketch of the formation of an Ω-loop from the azimuthal magnetic field presumed to lie in the tachocline, forming a bipolar magnetic region at the visible surface

at the surface random walk in the convection, exhibiting an effective turbulent diffusion coefficient of the general order of 6×10^{12} cm^2/sec at the same time that the magnetic flux is transported poleward by the meridional circulation of about 7 m/sec.

At this point we recall Babcock's (1961) suggestion that the observed inclination of the bipolar magnetic regions to the east-west direction provides the rotation of magnetic flux into the meridional plane (the alpha effect), thereby creating the poloidal field (Parker 1955b). If this idea is correct, then the upwelling Ω-loops provide the necessary generation of poloidal field, while Sheeley's surface studies find that the poloidal field is moved poleward by random walk and meridional circulation.

Now the principal shear (nonuniform rotation) lies in the tachocline at the bottom of the convection zone. We note that the upward displacement of magnetic field out of the tachocline to form an Ω-loop requires an equivalent downward displacement of gas back into the tachocline, carrying poloidal field down into the intense shear of the tachocline. This produces azimuthal field in the same location, either strengthening or weakening the existing azimuthal field at the same location. So the azimuthal field is not dispersed or dissipated by turbulent diffusion, but rather by the generating and superposition of new azimuthal field. Locally one would write

$$\frac{\partial B_\varphi}{\partial t} = B_r \frac{dv_\varphi}{dr}$$

for the generation of east-west field B_φ from the radial field component B_r by the shear dv_φ/dr. Integrating over r from immediately below to immediately above the thin tachocline, it follows that the time rate of generation of azimuthal flux Φ_φ per unit length in the north-south direction is $B_r \Delta v_\varphi$, where Δv_φ is the velocity difference across the tachocline. Across a length a in the north-south direction over a time t, the azimuthal flux generation is $at B_r \Delta v_\varphi$. So in a time of three years, 10^8 sec, a velocity difference of 10^4 cm/sec over a north-south band of width 10^{10} cm, with a radial field of 10 gauss, produces 10^{23} maxwells of azimuthal flux, comparable to the earlier estimate of 0.8×10^{23} maxwells, based on the observations of Gaizauskas et al. (1983).

Consider next the generation of poloidal field B_r from the observed inclination $\Delta\theta$ of the bipolar magnetic regions to the east-west direction. A modest Ω-loop with total flux Φ_Ω of the order of 10^{22} maxwells provides a magnetic flux $\Phi_\Omega \Delta\theta = 10^{21}$ maxwells circulating in the meridional plane. The polar fields of the Sun can be estimated at 10 gauss over a polar cap with a radius of, say, 4×10^{10} cm, representing a net flux of 5×10^{22} maxwells. On that basis the emergence of 50 Ω-loops over a few years of magnetic activity would provide the polar magnetic flux that accumulates at the poles.

These estimates suggest that the dynamo mechanism beginning with the intense azimuthal flux bundles and their Ω-loops may be potent enough to produce the observed solar magnetic fields. There are two obvious questions: (a) How is the intensity of the azimuthal magnetic field increased from the 10^2 gauss produced from B_r by the shear in the tachocline, to the inferred 10^5 gauss; (b) What is happening to the poloidal field in the convective zone where we cannot see it? In particular, how is it that the polar fields reverse relatively quickly and not necessarily simultaneously at about the time of maximum azimuthal field strength at low latitudes?

4 Discussion

We suggested some years ago (Parker 1994) that the convective instability initiated by the formation of an Ω-loop may be responsible for the inferred extreme concentration of azimuthal magnetic flux bundles to 10^5 gauss. The essential point is that the convective updraft set in motion by the buoyant rise of the magnetic Ω-loop drives an upward convective flow extending all the way across the convective zone, from the tachocline to the visible surface, sketched in Fig. 3. The convective forces are prodigious. The upward displacement of the plasma within the Ω-loop in the superadiabatic temperature gradient of the convective zone causes the plasma to push upward in the vertical legs of the Ω-loop, thereby pulling plasma out of the horizontal azimuthal flux in the tachocline at each foot of the Ω-loop. Reducing the plasma pressure within the azimuthal flux by 5×10^8 dyn/cm^2 is sufficient to increase

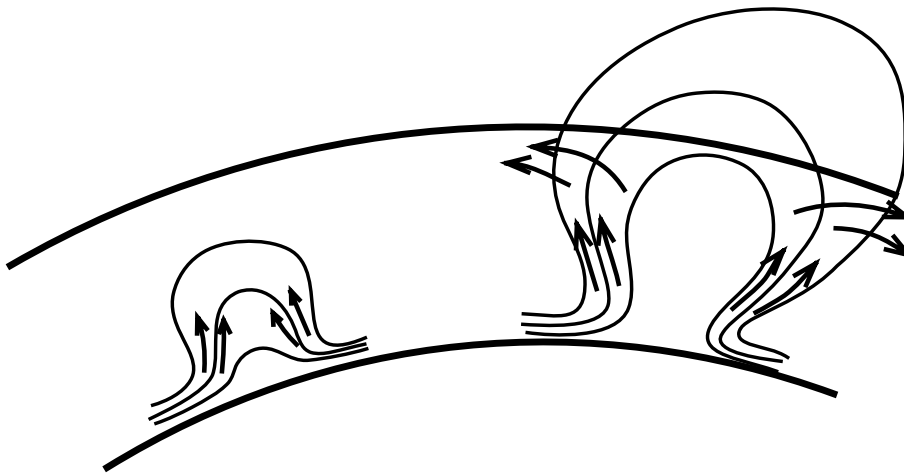


Fig. 3 Sketch of the convective upwelling within the legs of an Ω-loop during the rise of the loop from the tachocline to the visible surface

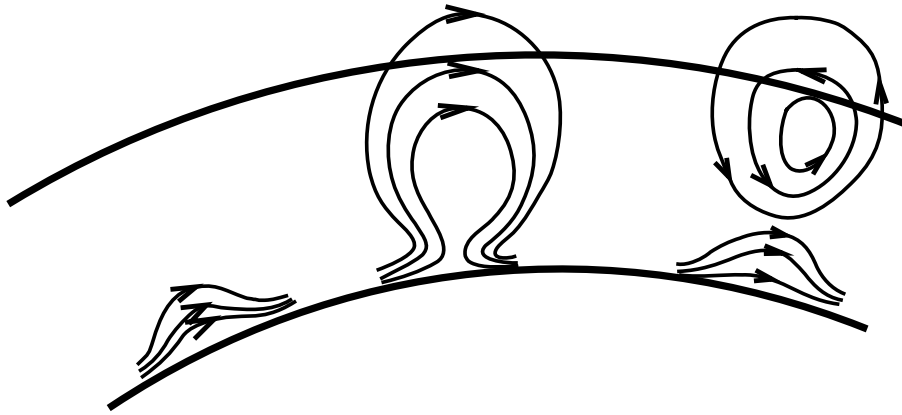


Fig. 4 Sketch of the conversion of an Ω-loop to an O-loop by rapid reconnection between the legs, restoring the azimuthal field to its original form and allowing formation of another Ω-loop at the same site

the magnetic field to 10^5 gauss, and the superadiabatic temperature gradient (Spruit 1974) across the convective zone appears to be more than adequate for the task.

Spruit et al. (1987) and Wilson et al. (1990) argue that the convective updraft around the rising Ω-loop pushes together the lower legs of the Ω-loop so that the field reconnects across the Ω-loop, cutting it off from the azimuthal magnetic field and converting it into an O-loop, as sketched in Fig. 4. Thus the azimuthal magnetic field is restored to its original form and is ready to repeat the Ω-loop process. The pumping of plasma out of the azimuthal field by the formation of an Ω-loop is repeated over and over. Gaizauskas et al. (1983) note the repeated emergence of new Ω-loops at intervals of 5–8 days, indicating ongoing convective pumping of plasma.

This brings us back to the 10^5 gauss azimuthal field with which we started.

Now in thinking over the individual steps in the dynamo process, it is evident that the poleward diffusion of the poloidal field at the visible surface, observed by Sheeley and others, raises some interesting and difficult questions about the subsurface poloidal field connections. In particular, the observed reversal of the polar fields is presumably a consequence of poloidal field of opposite sign diffusing into the polar cap. Recalling that the old polar field and the incoming new poloidal field exhibit a fibril state at the visible surface, we expect there will be a lot of rapid reconnection between fibrils of opposite sign. The photosphere and chromosphere are the favored levels for reconnection, where the plasma pressure falls below the magnetic pressure. Figure 5 sketches the new field connections that would follow from the reconnection. The new and old poloidal field lines are joined together, creating free loops of field above the reconnection level and captive loops in the convective zone. We can say little more because the initial connection of the old poloidal field between hemispheres is not known.

So in summary, we have outlined a sequence of effects that would seem to provide an $\alpha\omega$-dynamo for the Sun. Our treatment is not quantitative, and the individual effects that make up the sequence have yet to be studied properly. The outcome of each effect certainly depends upon the degree to which the solar magnetic fields are in a fibril form in the convective zone, about which we know nothing. So it is not unfair to ask whether the proposal contributes more knowledge than ignorance, or more ignorance than knowledge. And, of course, the skeptical reader is invited to propose better alternatives. A particularly

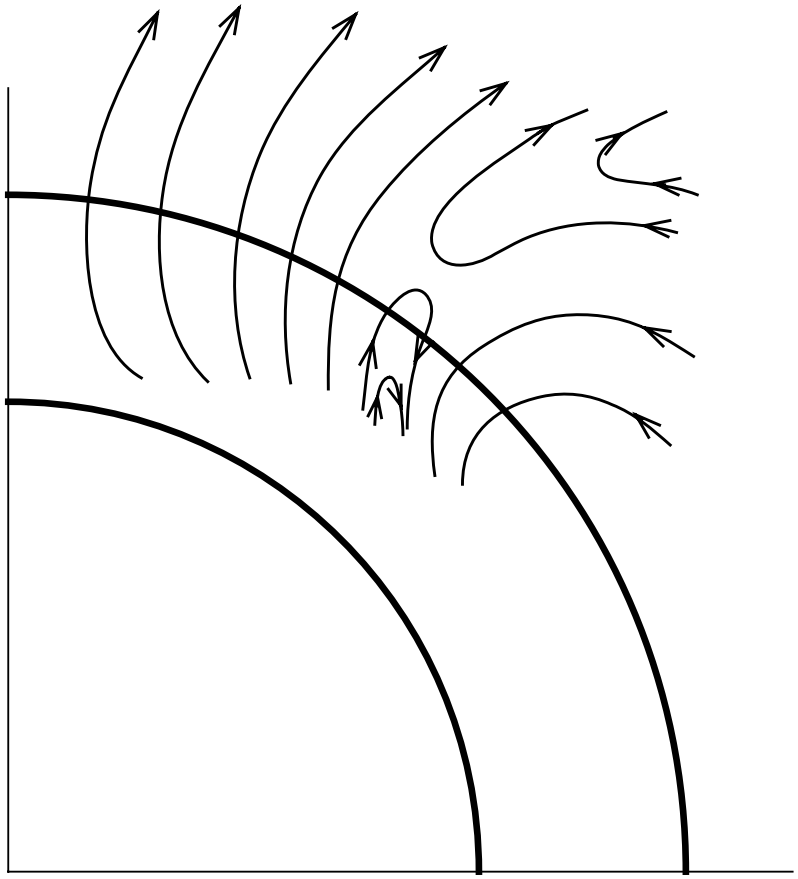


Fig. 5 Sketch of one quadrant of magnetic reconnection between the old polar field and the new intruding opposite poloidal field in the process of reversing the polar field

interesting twist would be success in showing how convectively driven turbulence might cause effective diffusion of a strong large-scale field, thereby returning us to more familiar dynamo theory. My own effort in that direction has had no success. Finally, we must keep in mind that until the theory of the solar dynamo can be raised above the level of conjecture, we should restrain our enthusiasm for extrapolating the $\alpha\omega$-dynamo concept to distant unresolved objects.

References

H.W. Babcock, The topology of the Sun's magnetic field and the 22-year cycle. Astrophys. J. **133**, 572–587 (1961)

H.W. Babcock, H.D. Babcock, The Sun's magnetic field. Astrophys. J. **121**, 349–366 (1955)

R.A. Choudhuri, P. Gilman, Astrophys. J. **316**, 788 (1987)

R.A. Choudhuri, M. Schüssler, M. Dikpati, The solar dynamo with meridional circulation. Astron. Astrophys. **303**, L29–32 (1995)

M. Dikpati, P. Gilman, K.B. MacGregor, Penetration of dynamo generated magnetic fields into the Sun's radiative interior. Astrophys. J. **638**, 564–575 (2006)

S. D'Silva, Can equipartition fields produce the tilts of the bipolar magnetic regions? Astrophys. J. **407**, 385–397 (1993)
S. D'Silva, A.R. Choudhuri, A theoretical model for tilts of bipolar magnetic regions. Astron. Astrophys. **272**, 621–633 (1993)
Y. Fan, G.H. Fisher, E.E. DeLuca, The origin of morphological asymmetries in bipolar active regions. Astrophys. J. **403**, 390–401 (1993)
Y. Fan, G.H. Fisher, A.N. McClymont, Dynamics of emerging active regions flux loops. Astrophys. J. **436**, 907–928 (1994)
V. Gaizauskas, K.L. Harvey, J.W. Harvey, C. Zwaan, Large-scale patterns formed by solar active regions during the ascending phase of cycle 21. Astrophys. J. **265**, 1056–1065 (1983)
P.A. Gilman, M. Miesch, Limits to penetration of meridional circulation below the solar convective zone. Astrophys. J. **611**, 568–574 (2004)
G.E. Hale, On the probable existence of a magnetic fields in sunspots. Astrophys. J. **28**, 315–343 (1908)
I.N. Kitiashvili, A.G. Kosovichev, Numerical dynamical modeling of solar cycles using dynamo formulation with turbulent magnetic helicity. Geophys. Astrophys. Fluid Dyn. (2008, in press)
R.B. Leighton, A magneto-kinematic model of the solar cycle. Astrophys. J. **156**, 1–26 (1969)
R. Matsumoto, T. Tajima, K. Shibata, M. Kaisig, Three-dimensional magnetohydrodynamics of the emerging magnetic flux in the solar atmosphere. Astrophys. J. **414**, 357–371 (1993)
F. Meyer, H.U. Schmidt, N.O. Weiss, P.R. Wilson, Mon. Not. R. Astron. Soc. **169**, 35–42 (1974)
F. Meyer, H.U. Schmidt, N.O. Weiss, Mon. Not. R. Astron. Soc. **179**, 749 (1977)
E.N. Parker, The formation of sunspots from the solar toroidal, field. Astrophys. J. **121**, 491–507 (1955a)
E.N. Parker, Hydromagnetic dynamo models. Astrophys. J. **122**, 293–314 (1955b)
E.N. Parker, The solar hydromagnetic dynamo. Proc. Nat. Acad. Sci. USA **43**, 8–14 (1957)
E.N. Parker, Sunspots and the physics of magnetic flux tubes. I. The general nature of the sunspot. Astrophys. J. **230**, 905–913 (1979)
E.N. Parker, Stellar fibril magnetic systems. I. Reduced energy state. Astrophys. J. **283**, 343–348 (1984)
E.N. Parker, Theoretical properties of Ω-loops in the convective zone of the Sun. I. Emerging bipolar magnetic regions. Astrophys. J. **433**, 867–874 (1994)
G. Rüdiger, L.L. Kitchatinov, The slender solar tachocline. Astron. Nachr. **318**, 273–279 (1997)
M. Schüssler, P. Caligar, A. Feriz-Mas, F. Moreno-Insertis, Instability and eruption of magnetic flux tubes in the solar convective zone. Astron. Astrophys. **281**, L69–L72 (1994)
N.R. Sheeley, Measurements of solar magnetic fields. Astrophys. J. **144**, 723–732 (1966)
N.R. Sheeley, Y.M. Wang, C.R. DeVore, Implications of a strongly peaked polar magnetic field. Sol. Phys. **124**, 1–13 (1989)
H.C. Spruit, A model of the solar convective zone. Sol. Phys. **43**, 277–290 (1974)
H.C. Spruit, A. M Title, A.A. Ballegooijen, Is there a weak mixed polarity background field? Theoretical arguments. Sol. Phys. **110**, 115–128 (1987)
P.R. Wilson, N.R. Sheeley, N.B. Snodgrass, The reversal of the solar polar magnetic fields. Sol. Phys. **127**, 1–9 (1990)
C. Zwaan, The emergence of magnetic flux. Sol. Phys. **100**, 397–414 (1985)

Chaos and Intermittency in the Solar Cycle

Edward A. Spiegel

Originally published in the journal Space Science Reviews, Volume 144, Nos 1–4, 25–51.
DOI: 10.1007/s11214-008-9470-9

Abstract Where a magnetic flux tube of sufficient strength and cross section protrudes from the sun, convection is locally inhibited and a sunspot appears. The number of spots on the sun at any time varies in a cyclic, but aperiodic, manner. Models with chaos and intermittency can capture the main qualitative aspects of this temporal variability, especially if they display the mechanism of on-off intermittency. Capturing the spatio-temporal aspects of the sunspot cycle requires a more complicated model but a description in terms of waves of excitation seems promising. To clarify these possibilities, qualitative introductory remarks about chaos theory itself are included in this narrative.

Keywords Sunspot number · Tachocline · Chaos · On-off intermittency · Dimensional reduction · Amplitude equations

1 Sensitive Systems

Four centuries ago, Galileo noticed that the period of a pendulum is the same for all small amplitudes. Not long afterwards, Galileo and his contemporaries established that sunspots really were on the sun itself. So the same person was involved in discovering the paradigm of periodicity and establishing an examplar of irregularity. But just how irregularly do sunspots behave? This question has become one of practical interest as more and more people venture into space and want to have warnings of disagreeable solar weather. And, if the word weather is truly appropriate, we should take note of this quotation:

> *. . . who could enumerate the countless changes that the atmosphere undergoes every day, and from that predict today what the weather will be a month from or even a year from now*?

To some, this may sound quite modern, but there is nothing new in it. When things get sufficiently complicated, it is usually not possible to predict what will happen after a

E.A. Spiegel (✉)
Astronomy Department, Columbia University, New York 10027, USA
e-mail: eas5@columbia.edu

M.J. Thompson et al. (eds.), *The Origin and Dynamics of Solar Magnetism*.
DOI: 10.1007/978-1-4419-0239-9_3

(certain) time. Decades ago, people studying fluid dynamical turbulence used to express this by saying that small causes (in turbulence) have large effects. For the most part, this behavior on the part of the weather and of turbulent processes in general has been accepted as a fact of life for which detailed explanations were rarely sought. Erratic behavior was, whether explicitly or tacitly, ascribed to some *deus ex machina*. So you may not be surprised when I tell you that quote is not at all recent; it is from Jakob Bernoulli's (1713) *Ars Conjectandi*. It is largely because of recent developments in the branch of dynamical systems theory known as *chaos* that the quote has a modern ring.

Since the sixties, much has been written about chaos, which is concerned with systems that, like turbulent systems, exhibit an unpredictable behavior despite being deterministic. The phrase that is often used to describe such systems is that they exhibit sensitive dependence on initial conditions. The excitement engendered by the study of sensitive systems (as I call them) was largely a result of the increased realization in recent decades that sensitivity can be observed in systems with very few degrees of freedom. In fact this awareness goes back to Poincaré or perhaps earlier, but computers were not making extensive studies of such behavior so easy a century ago.

Some people use the word revolutionary in speaking of the development of chaos theory and the description that it affords of aperiodic behavior. Yet what has been uncovered in this domain is not a new science, but a new understanding of an old science. That understanding has spread through many disciplines in the way that thermodynamics has made itself felt in many subjects. The well from which chaos springs is mechanics (with input from pure mathematics) following on Newton's laws and his derivation of Kepler's laws. A century after Newton had finally published the *Principia*, it was still true that the only problems of celestial mechanics that had been solved were those that were in some sense close to the two-body problem. Nevertheless, people soldiered on, motivated by the belief that,

> *An intellect which at a given instant knew all the forces acting in nature and the positions of all things of which the world consists—supposing the said intellect were vast enough to subject these data to analysis—would embrace in the same formula the motions of the greatest bodies in the universe and those of the slightest atoms; nothing would be uncertain for it, and the future, like the past, would be present to its eyes.*

Well, we have some good computers now, but they are not that good yet. The goals set by Laplace's dogma of determinism are well beyond us. Yet, just twenty years ago, Bondi could write that "most of science is not like the Solar System but much more like weather-forecasting." We may sympathize with the distinction he made but, by then, chaos theory had brought the thinking of modern mechanics into meteorology and other complex subjects. The question I address here is how we may adapt the new outlook to clarify our thinking about the workings of the solar cycle. But this discussion is not meant as a research paper. And it is definitely not a review. I shall simply try to provide a qualitative prelude to the study of the role of chaos theory in modeling solar magnetic variability.

2 Examples of Sensitivity

2.1 An Illustration from Bifurcation Theory

To illustrate the nature of sensitive systems let us look at a classic example: a bead sliding on a circular hoop. If the hoop is in a vertical plane, this system is equivalent to a simple pendulum. But let us suppose that the hoop is rotating at constant angular velocity, Ω, about

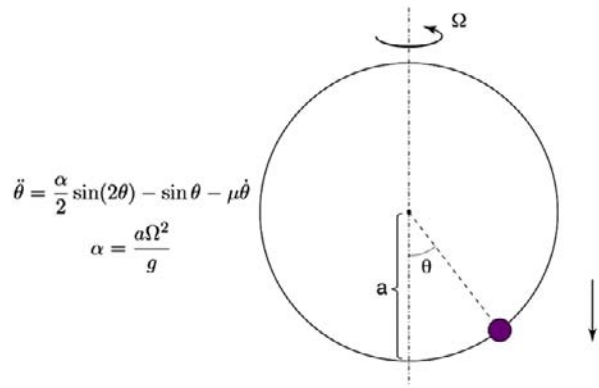


Fig. 1 A bead moving on a hoop rotating around a vertical axis

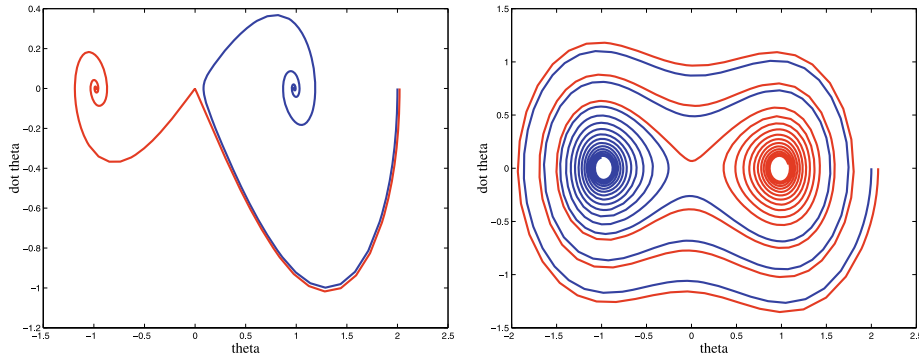


Fig. 2 Pairs of phase trajectories for the bead of Fig. 1 starting from adjacent initial conditions. *Left*: moderate friction; *Right*: small friction

a vertical axis, as shown in Fig. 1. The figure also displays the equation of motion of the pendulum in nondimensional form. From this we see that the bead has two positions of equilibrium, one is at $\theta = 0$ and the other, an unstable one, is at $\theta = \pi$. When we rotate the hoop, the lower equilibrium is stable so long as $\Omega^2 < g/a$, where a is the radius of the hoop. Once that value of Ω is exceeded, the lower equilibrium becomes unstable and two new, stable equilibria appear, one on each side of $\theta = 0$. This is an example of a stationary bifurcation. A plot of the angles at which the equilibria are found vs. Ω^2 resembles a pitchfork and so the formation of the new equilibria is also called a pitchfork bifurcation.

The equation of motion given in the figure is readily solved on a laptop computer. Figure 2 (after Spiegel 1985) shows pairs of trajectories of the bead in its state space (or phase space) whose coordinates are θ and $\dot{\theta}$ where the overdot denotes time derivative. In the figure, the two members of each pair move on trajectories that begin at $\dot{\theta} = 0$ with initial values of θ that are only slightly different from one another. The two panels of the figure show results for moderate (left) and small (right) friction. With lower friction, the orbits stay close together longer and, when μ is made small enough, my laptop cannot reliably predict on which side the bead will end up. If you have a large computer, tell me how many significant figures you can carry and I can make μ so small that your computer cannot make that prediction either. If you want to do this experimentally, let me know the frequency of your light source and similar limitations can be imposed. So here we see sensitivity in a

simple system. It is very like the problem of tossing a coin *in vacuo*, as Keller (1986) has shown, and yet many people think of coin tossing as a stochastic rather than a deterministic process. (I'll not get technical about these terms here but will hope that you see what I am driving at.) Few people would call the behavior of this system chaotic because the system runs down after a while. An interesting aspect of this system is that its sensitivity is not the same everywhere in the state space. A more marked nonuniformity of sensitive behavior in the state space is found in the chaotic systems that we discuss next (Smith et al. 1999).

2.2 A Chaotic System

The example we have just looked at is sensitive, but only until it enters the domain of attraction of one or the other of the stable equilibria. To make the system chaotic in the usually understood sense, we need to keep it going. To do this simply, let us restrict the model to small angles and rename θ in that case as x. For small angles, the trigonometric functions may be well approximated by multinomials in x rather than the transcendentals of the full pendulum problem. We may then write the equation of motion in the form

$$\ddot{x} = -\frac{\partial V}{\partial x} - \mu\dot{x} \tag{1}$$

where $V = V(x, \alpha)$ and α is again a parameter that, like the α of Fig. 1, quantifies the ambient conditions. Let the potential, V, be a quartic in x. Then we can get behavior qualitatively like that of the previous problem, for some values of α. Again, this oscillator always comes to rest eventually. But when α varies slowly in time, we find new regimes of behavior including what is called continuous chaos.

Variation of α may arise in two ways. We may simply impose variation of α (periodic, say) or we may let α vary in response to the changes in x. In the latter case, we may impose a variation of α by specifying a differential equation for it such as

$$\dot{\alpha} = -\epsilon[\alpha + g(x)], \tag{2}$$

where ϵ is positive and small and g is chosen according to the problem at hand. As an example, consider the case

$$V = \frac{1}{4}x^4 - \frac{1}{2}x^2 - \alpha x \quad \text{and} \quad g = ax(x^2 - 1) \tag{3}$$

where a is a constant chosen to give interesting behavior, which is readily found numerically. We may imagine the trajectories in the four-dimensional spacetime (appropriate for a meeting in Bern) with coordinates (x, y, α, t). Though this may seem hard to visualize, we may look at an extended trajectory in spacetime from different directions to get some feel for the structure that is formed. Neil Balmforth has provided a numerical solution that is very helpful in this. He has plotted a single, long run in projections onto the $x - \dot{x}$ and $x - t$ planes and has overlaid them in the manner of the cubists (or tesserists) as shown in Fig. 3. We see there what is called a strange attractor in the language of chaoticists. Extensive discussions of this system are given in Balmforth and Craster (1997) and Smith (2007).

Here I should inject a few words about terminology. The word chaos derives from Greek mythology (Lloyd 1975): "Hesiod's *Theogony* provides an account of the origins of [the cosmos]. This begins with Chaos in the Greek sense of gap." Whatever the nature of this being, its association with vanishing density must have seemed to suggest a good name for

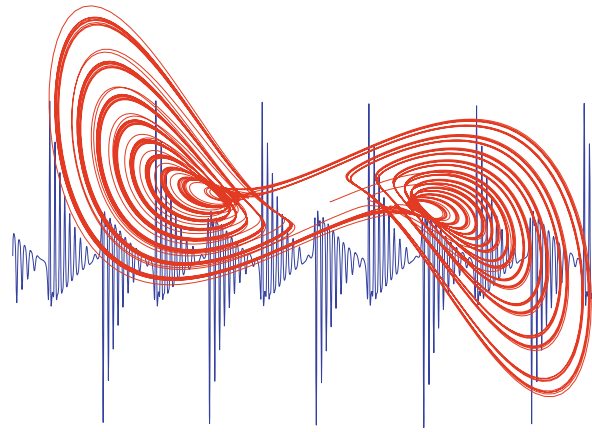


Fig. 3 Two projections of a trajectory of system (1)–(3)

a low density medium—a gas—to Jan Baptista van Helmont (Hoffman and Torrence 1993). The difficulty of predicting the behavior of gases perhaps made it reasonable to associate the word chaos with unpredictability and disorder, though here I am on a slippery slope. Perhaps all this was in the mind of Jim Yorke when he proposed the term chaos to name unpredictable systems more generally. Once the term was proposed and came into current use, there was a long period when people could not agree on a suitable precise definition of chaos. As this discussion is meant to be qualitative, the general descriptions offered throughout should suffice.

In any case, the word I have used for someone who studies chaos is found in the writings of Thurber (1947): "I had broken away from an undulant discussion of kinetic dimensionalism and was having a relaxed moment with a slender woman I had not seen before, who described herself as a chaoticist, when my hostess, an avid disturber of natural balances and angles of repose, dragged me off ...." However, Michael Berry has proposed that we use the word "chaologist," whose early appearance in the OED gives it precedence. So we have a choice in this perhaps.

Another term used in this field may be appreciated on realizing that the velocity in the state space of the system just discussed, is $(\dot{x}, \dot{y}, \dot{\alpha})$, where $y = \dot{x}$; its divergence is $\partial\dot{x}/\partial x + \partial\dot{y}/\partial y + \partial\dot{\alpha}/\partial\alpha$. If we start off an ensemble of realizations of the system, the rate of change of the volume of the swarm of representative points, per unit volume, is then $-(\mu + \epsilon)$, which is negative under physically reasonable assumptions. So the volume of the swarm goes to zero (eventually) and the object so formed is called an attractor. (In a state space of dimension n an object embedded in a lower dimensional subspace has zero volume though, in the subspace, it may be seen to have finite volume.) In the example of Fig. 3, the attractor has a (suitably defined) dimension that is not an integer and it is known as a "fractal" in the language of Mandelbrot (1997). Such attractors are called "strange" (Ruelle and Takens 1971) and their dimensions may be calculated as described in many places such as Wolf et al. (1985).

2.3 Qualitative Considerations

There are various ways to think about chaos, most of them are inevitably mathematical. But here is a softened way to picture a chaotic process. (If you have mathematical friends, it might be best not to show them the following few paragraphs.)

Let us think back on the model oscillator described by (1)–(3). For α fixed at a value where the quartic has two wells and with $\mu = 0$, the oscillator, if given a suitable value of energy, will bounce back and forth periodically. Now if α is allowed to vary slowly, say in response to the variation of x, the behavior changes. The oscillator leaves a turning point, moves into the well, crosses it, and bounces back. But when it returns to its starting position, it may encounter a different value of V there. It may be bounced back into the well sooner than expected, or later, depending on how V is varying. If, on the second return, it hits the original value of V in the original place, it will have found itself in a periodic orbit with about twice the period of the first one. Indeed, there may be a few such orbits but the system will seek those that are stable. If the time scale of α's variation changes in response to a change in ϵ, it may take more traverses of the well before the oscillator again encounters the original value of V at the original place. But if it does have such an experience, it will again be in a periodic orbit but with a yet longer period.

There are parameter regimes where quite a few periodic orbits are found; in some cases, there are infinitely many. For systems like (1)–(3), conditions that lead to an infinite number of unstable periodic orbits may be given explicitly and this possibility is part of what is known as Shil'nikov's theorem. In fact such systems can be derived from the fluid dynamical equations for processes such as convection with competing stability effects (Arneodo et al. 1985). The phase trajectories of many of the periodic solutions look alike. They are so much of a muchness that a slight change in the state space coordinates of the system will move the system from the realm of one periodic orbit into that of another. Qualitatively speaking, when the periodic orbits are very similar, they all tend to be unstable, because none of them is preferred. However, there is a tendency for the system to hover briefly near one or the other of these orbits before moving on to another one. This behavior is recognizable in Fig. 3 where you see that the system is doing almost the same thing over and over again but never (or hardly ever) quite the same way twice. In these conditions, unless the system is on a periodic orbit and unperturbed, it will drift amongst them all. So one question for us here is whether this is a fruitful way to think about the solar cycle. As we shall see, the answer depends on which aspect of the data you are trying to model (see, for instance, Spiegel and Wolf 1987).

In the case of a third-order system such as the one we just have looked at, a useful way to build intuition is by constructing a surface of section or Poincaré map. In this approach, one considers a plane (or other smooth surface) in the three-dimensional state space chosen to cut through the region where the motion is taking place. Each time a given trajectory traverses this plane going from one particular side to the other, a point is placed where the crossing occurs. In the case of a simply periodic orbit, each traversal takes place at the same point once the transient approach to the periodic orbit has ended. That single point is called a fixed point of the map. When a parameter change produces a doubly periodic orbit, the map acquires two new recurrent points. (We again omit the transients.) This is illustrated by the sketch in Fig. 4 for a sequence of stationary bifurcations.

We see at the top of Fig. 4 the case of the simply periodic orbit. With a good description of the map, one can study the stability of such a fixed point and the result is a pair of eigenvectors and corresponding eigenvalues. In the case of a stable periodic orbit, we get a pair of negative real eigenvalues. The eigenvectors are the backbones of a flow in the surface of section that carries the successive crossing points toward the periodic orbit along with their close neighbors. I have also sketched in the nonlinear extensions of the linear eigenvectors in a qualitative manner. These extensions of the eigenvectors are called the stable manifolds of the periodic orbit.

The drawing just below the first one in Fig. 4 illustrates the case when some parameter in the system has been changed so as to render the simply periodic orbit unstable. At that

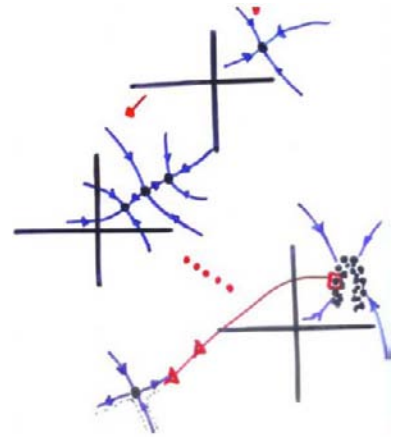


Fig. 4 Surfaces of section with their stable and unstable manifolds. Explanation in text

parameter value, in addition to the original fixed point of the simply periodic orbit, we find a pair of points that represent a periodic orbit of approximately twice the period of the previous one. Here the stability situation is richer. In many cases, the situation is as shown in the figure. The nonlinear extensions of the eigenvectors of linear theory around the original fixed point are indicated but only one is a stable manifold while the other is an unstable manifold. That unstable manifold of the fixed point leads to the two points that are found for the map of the newly formed doubly periodic orbit. In the case illustrated, that unstable manifold is in fact one of the stable manifolds of the doubly periodic orbit, which also has a second stable manifold. Both stable manifolds are shown with their invariant nonlinear extensions. Just beyond the *bifurcation* point where the simply periodic orbit loses stability, and the two new points in the map are born, the original simply periodic orbit, which is still a solution, has become an unstable one.

As we keep changing the control parameter along a suitable path in parameter space, we pass through a sequence of instabilities that represents the storied period-doubling bifurcation sequence of chaos theory. The result is an infinite number of points representing the multiply periodic orbits that have been formed. Also suggested by a crude sketch at the bottom of the figure representing all those periodic solutions is the notion that every point in the limiting map has one stable and one unstable invariant manifold. A representative point that gets near a stable manifold of a periodic orbit is attracted to that orbit and, when it gets close to the orbit, it moves mainly along the periodic orbit.[1] The successive crossings of the surface of section are then close to each other. But finally, the representative point of the system gets into the zone of influence of an unstable manifold, which happens also to be a stable manifold of another periodic point, and the system heads there before being diverted again.

From a great distance, we resolve only the form of the original pair of stable manifolds (somewhat distorted). So a representative point that starts out far from the action will be drawn into the globally attracting crowd of points and wander around in there forever. That cluster of points representing the surface of section represents all of the unstable periodic orbits. All those unstable periodic orbits in the phase space together constitute the strange attractor or, as here, its surface of section. For a thorough discussion of such chaotic processes

[1] Here, I am asking that you keep both the full motion in phase space *and* the sequence of points at which the orbit crosses the surface of section in mind.

in astrophysics see Regev (2006) and for a very brief sketch of how those considerations may apply to the solar cycle see Spiegel (1997).

3 Quantifying Chaotic Behavior

Some people distinguish between deterministic systems whose future behavior is predictable in the sense of the Laplace quote above and those systems whose futures are indeterminate and that are sometimes called *stochastic*. This distinction brings to mind the probabilistic interpretation of quantum mechanics, but I shall put quantum processes aside here and consider only classical systems. In that case, for me, all physical systems are deterministic. However, their futures may not be determinable in practice. That is what is meant by 'sensitivity.' To test for this property we may start the system off with two initial conditions very close to each other to see how the separation between the two (seen as representative points in state space) evolves. The mean over the orbit (or a circuit of the attractor) of the rate of increase of (the log of) the separation is called a Lyapounov exponent. If for short times, the separation of two close-lying representative points grows exponentially we may call the system sensitive (that is, chaotic). In fact, the rate of separation varies over the state space and it is not simple to determine the average rate (see Wolf et al. 1985). If the separation grows exponentially in only one direction at each location in state space, this is ordinary chaos, a fairly mild form. When there are many such directions, the chaos is in a regime sometimes called hyperchaos. As to real fluid turbulence, perhaps it should be called ultimate chaos. Here I concentrate on mild chaos.

If we suspect that a system is chaotic, how should we model it, particularly if we are uncertain about the underlying physics? If the number of degrees of freedom that seems required is large, you might prefer to use a stochastic model. So the interest in chaotic behavior has centered on systems with only a few degrees of freedom. Hence one goal of the tests used to study aperiodic behavior is to determine the number of degrees of freedom involved in that behavior. That is a rather technical subject (see Eckman and Ruelle 1985) and I will indicate only the road one takes without describing the mathematical landscape it passes through.

There has been much discussion about the efficacy of these methods but, as many specialists agree, the data on the sunspot number are not adequate for accurately deciding the number of degrees of freedom needed to model the variation of the sunspot number in time (see Fig. 5). Even beyond this question, there lies the deeper issue of whether a deterministic model such as the (potentially) chaotic system of equations (1) and (2) with its relatively few degrees of freedom, can be used to give us some insight into the workings of the solar cycle. At first glance one might think this is not a promising line of attack since the cycle is a spatio-temporal process while systems like (1)–(3) do not allow for spatial variation. Below, I shall describe some of the (perhaps overly optimistic) grounds for thinking that simple systems may be helpful in studying the solar cycle. But before going into the why of this problem, I want to turn to the how of it and to indicate a method of extracting some helpful information from data by what is called time series analysis.

3.1 Studying Empirical Attractors

Say that we have made a series of observations on a quantity $X(t)$ at times $t = (0, 1\Delta t, 2\Delta t, \ldots)$. (Making observations at uniform time intervals is often not possible. The failure of that as well as other idealizations I shall implicitly assume needs to be accounted

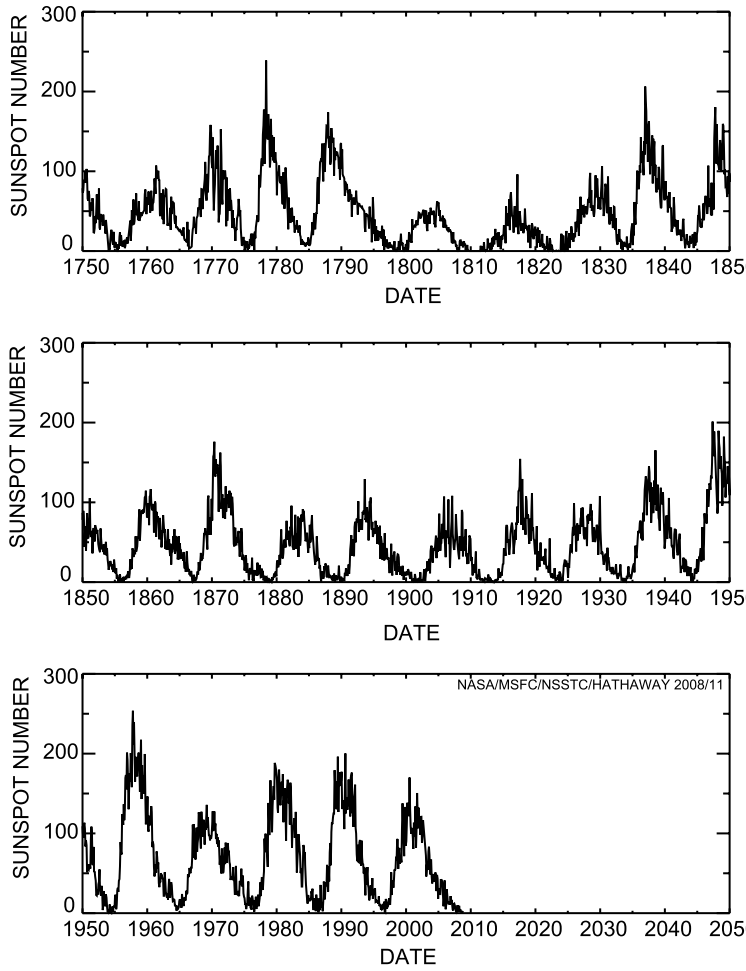
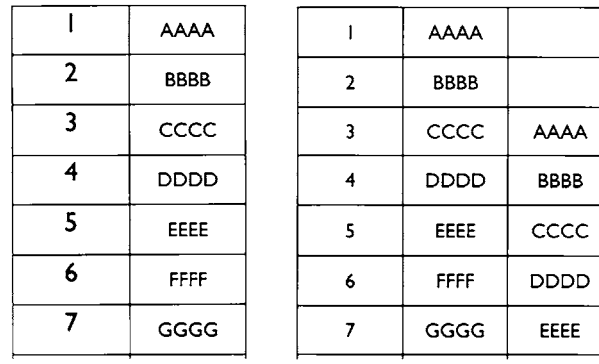


Fig. 5 The monthly sunspot number from The World Data Center for Solar-terrestrial Physics

for, but I will not go into such nagging problems here.) If, for instance, $X(t)$ is governed by a system of n first-order ordinary differential equations, we may expect that a suitable state space for the underlying theoretical system would have coordinates $X, \dot{X}, \ddot{X}, \ldots$ up to a derivative of order $n-1$. However, when we are dealing with measurements we have to accept that numerical differentiation of data is not a very accurate process, especially if high derivatives are needed. So, for studying measurements on time series, a device called the method of delay coordinates is often used to analyze the data.

Suppose the series of measurements (in time) that you have made is as represented schematically in Fig. 6a. Then, next to the original column of data, place a copy of the same column translated down by a suitable interval of time. That interval, as shown by experience and a certain amount of judgement (see Fraser and Swinney 1986), ought to be of the order of one quarter of the characteristic time scale of the process under study. Now you have two columns of data, as shown in Fig. 6b, that present you with a set of ordered pairs representing a sequence of points along a trajectory in a trial two-dimensional state space.



1	AAAA
2	BBBB
3	CCCC
4	DDDD
5	EEEE
6	FFFF
7	GGGG

1	AAAA	
2	BBBB	
3	CCCC	AAAA
4	DDDD	BBBB
5	EEEE	CCCC
6	FFFF	DDDD
7	GGGG	EEEE

Fig. 6 *Left*: **a** The measurements vs. time; *Right*: **b** Delayed ordered pairs

(We lose some data at the ends of the run in this way, but this loss is made up for by the convenience of this scheme.) Connect up these points identified by the coordinate pairs and see what that trajectory looks like. If, for example, it crosses itself, this would not be a correct description of what the system does if it is deterministic since the motion after the crossing would be indeterminate. If this does happen, make three columns or more instead of two until there are no more crossings. Then perhaps go one more to see if you have converged. (That is, see whether the topology of the phase trajectory that you are constructing seems to have converged.) If the system is chaotic, the trajectory will wander around in the (putative) strange attractor and you will have a reasonable sketch of it if your measured time series is long enough.

There is some mathematics that goes with this story. In particular, a theorem of Takens (1981) says that the attractor constructed in this way has the same topological structure as that of the system that produced the data. (It may be geometrically different and not immediately recognizable to the naked eye, even to that of an astronomer.) There is a lot of fine print that goes with this so it is best to have an expert near by when going on such an undertaking. But such cautioning is not so relevant to the present case since the data for the solar cycle are not sufficient to make theorems very meaningful, according to my gurus of this subject. Nonetheless, I feel one may well use the approach for qualitative examination of time series. If you wish to see how applications of the procedure do work for experiments where the data are abundant see Wolf et al. (1985); to read about the theory in a lengthy review see Eckman and Ruelle (1985); or to get an appraisal of what may happen in real life from a slim volume consult Smith (2007).

3.2 An Illustrative Example

Figure 7 shows a trajectory calculated by Lenny Smith for the system (1)–(3) seen from a different viewpoint than that of Fig. 3 and shown as a series of points along the orbit. When you have a simple system like that, you can generate all the data you might want and the x coordinates of those points can be used to reconstruct the trajectory. When you apply the method of delay coordinates to the data so obtained (to $x(t)$, say) you get very nice trajectories. I use the plural because you must make choices about the delay time and other details. Figure 8 shows a couple of such outcomes for low dimensional attempts done with three-column reconstructions (in a trial three-dimensional state space) for different delay times and projected onto two dimensions. Though it is not obvious to visual inspection, the results do seem consistent with the mathematical result that, once we get to high enough dimension for the embedding space, the topology of the attractor does not depend on the

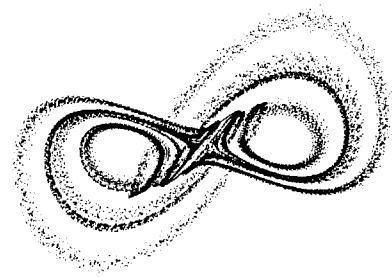


Fig. 7 Points along a trajectory of the system (1)–(3)

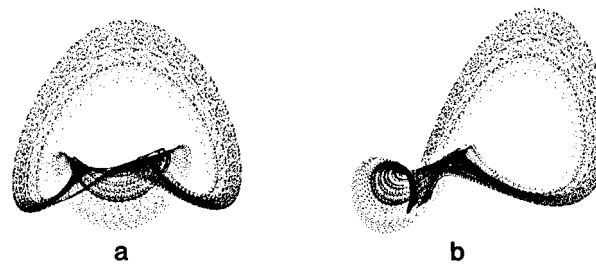


Fig. 8 Two possible delay reconstructions from a data series from Fig. 7

choices of such details as delay times (within reason). One may try to think of ways to change such an outcome but you must stay within the rules of procedure laid down in the theorems. When you look at the reconstructed objects of Fig. 8, would something like the one of Fig. 7 made from the (so to say) raw data suggest itself? Possibly not, and that is why people have worried more about the dimensions of the trajectories constructed from the data: the qualitative implications of the reconstructed trajectories are harder to characterize. On the other hand, if your data do not conform to the desiderata mentioned in the fine print that goes with the theorems about these methods it makes little sense to get very quantitative about all this. As far as I know, not a lot has been done on giving error estimates on things like the dimensions of the attractors suggested by the analyses. However, if you are not distressed by the disapprobation of experts, why not try to derive some insights into the workings of the solar cycle from the appearance of the reconstructed attractor, as we do next.

3.3 A Solar Attractor

Though the data for the solar cycle are too few to allow a fully justified application of the method described the previous section, I suggest that the application of the procedure can produce synthetic trajectories in model state spaces that may give a qualitative impression of what the attractor of the system may look like. From that, one can try to make a mathematical model that produces such an attractor and use it to infer what the possible underlying physical mechanisms may be. In the case of the solar cycle, this approach has been of some help. So I shall here describe the application of the method of delay coordinates to the solar cycle.

The sunspot number[2] has been recorded daily since the middle of the nineteenth century and reconstructed to over a century before that. But the sun turns once a month. Since we see a slightly different portion of the sun each day, it makes sense to look at monthly averages of some measure of the solar activity such as the sunspot number. (Even that is not perfect

[2]This is not simply the number of spots but a weighted average of the numbers of spot groups and individual spots and averaged again over the reports from various observatories.

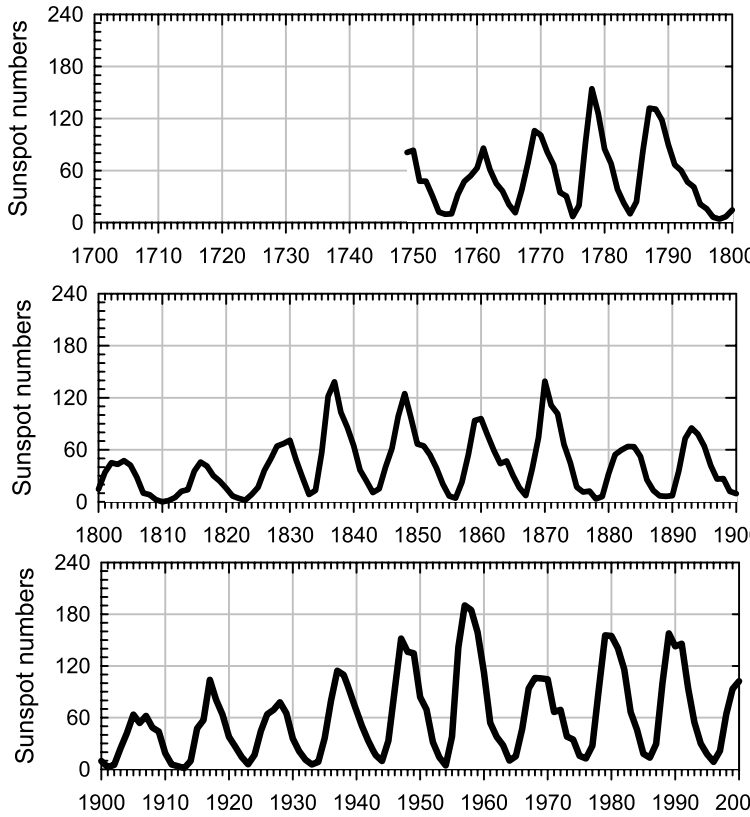


Fig. 9 The annual sunspot number from The World Data Center for Solar-terrestrial Physics

since the sun is not a rigid rotator, but such errors are negligible at present.) However, as we see in Fig. 5, the monthly variations are lively, so that modeling those fluctuations on that time scale is likely to involve us in more degrees of freedom than we can usefully cope with as yet. If you go directly to the delay coordinate representation of these data in an embedding space of three dimensions (for instance) you will see various small loops in the phase trajectory in addition to the overall structure. My qualitative impression is that there are more one-year loops than I would have expected and I do not know what they indicate. Others see signs of other time scales and these all deserve to be considered more carefully, but I shall restrict my attention to the time scales of the solar cycle and the rough recurrence times of the grand minima when very few spots are seen. For this, I shall deal with smoothed versions of the data in Fig. 5, with time scales of less than two years filtered out. One may do many things with such data—Wolf and I exhibited a variety of things with them in Spiegel and Wolf (1987) where some relevant references for that era may be found—and many of them give suggestive images. But I shall limit myself to a single example. Figure 9 shows the annual sunspot number for the past two and a half centuries obtained from the World Data Center for Solar-terrestrial Physics to help us keep in mind what we are trying to understand.

In the century or so since Hale first measured the solar magnetic field it has been found that the field reverses polarity on every cycle. This is so both in the global weak field and in the way that the polarities of the stronger fields in pairs of spots organize themselves in

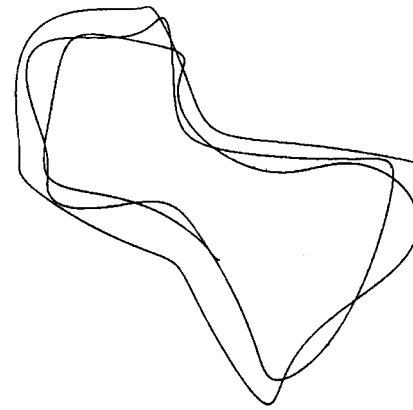


Fig. 10 A delay reconstruction of a solar attractor using the Bracewell artifice

the northern and southern solar hemispheres. Thus, the time scale of the solar magnetic cycle should perhaps be taken as twenty-two years not as eleven years as the sunspot number variation might imply. Bracewell (1985) suggested that one could make this explicit in the presentation of the data by simply calling the sunspot number negative on alternate cycles. There are other ways to put the twenty-two year variation in evidence such as taking the square root of the sunspot number with suitable signs or more elaborate approaches (Letellier et al. 2006). As I am here being qualitative, the Bracewell modification will do. For this example, Wolf and I filtered periods less than about two years from the data. Then we constructed triplets from the delay coordinates and projected the reconstructed state space trajectory onto two dimensions for plotting purposes as shown in Fig. 10. The orbit we found looks somewhat like a trajectory from system (1)–(2) with

$$V = \frac{1}{4}x^4 - \frac{\alpha}{2}x^2 \quad \text{and} \quad g = a(x^2 - 1) \tag{4}$$

where a is again a constant. *Mirabile dictu*, the system (1), (2), (4) is equivalent to the Lorenz equations (Sparrow 1982) after a simple transformation of the dependent variables (Robbins 1997; Marzec and Spiegel 1980). Moreover, the Lorenz equations govern the dynamics of the shunted disk dynamo described by Malkus (1972). Such coincidences may owe much to the commonality of many low-order irregular oscillators. Besides, the appearance of the trajectory segment in Fig. 10 is conditioned by the various choices made in its construction. Yet, it may be reasonable to proceed as if the solar cycle has behind it something like a simple dynamo process such as the disk dynamo. Indeed, a third-order system like that of Lorenz may produce a cyclic oscillation whose amplitude variations resemble those of the sunspot number when smoothed over a year or two. But it will need a more elaborate model to produce an analogue of the grand minima. I will discuss one such model after the next section, which is a technical digression aimed to provide some rationalization for seeking a relatively simple model of the behavior of the sunspot number.

4 Dimensional Reduction[3]

The neurons in our brains are as numerous as the stars in our galaxy, yet they work together well enough to allow me to write this story and you to read it (if you choose to). What leads to

[3]This chapter may be skipped without loss of comprehension of the sequel.

this coherent behavior of so many constituents? Such questions about the origins of ordered behavior arise throughout science as when now we ask how it is that the particles in the sun conspire to produce the macroscopic hydromagnetic processes that we are considering. And I am imagining just such a marvel in attempting to describe the variation of the sunspot number as the output of an oscillatory system with only a few degrees of freedom.

How do complicated systems manage to carry on their activities in subspaces that have many fewer dimensions than the full state spaces? This question lurks behind many scientific theories and I suppose that most of us have some intuitive notions about its answer. The idea that aspects of highly complex, spatially structured systems can be derived from relatively simple mathematical descriptions—ODEs, say—is referred to as dimensional reduction. In kinetic theory, when one goes from a kinetic description such as the Boltzmann equation in six (or more) dimensions to the Navier-Stokes equations in only three, the procedure is called contraction (Uhlenbeck and Ford 1963).

It is generally accepted that the hydromagnetic extension of the Navier-Stokes equation, when applied to the sun, contains an adequate description of the workings of the solar cycle. Since we cannot as yet properly solve those equations it may not be unreasonable to seek dimensional reductions that may guide us. Here, we are looking at an extreme example of this approach and I shall attempt to provide its motivation using a mathematical *sounding* language to give you a feeling for how this works. I would not however call it mathematics.

With that disclaimer, let me suppose that we have to deal with a system whose fundamental fields (or state variables) are denoted as $\mathbf{U} = U_K(\mathbf{x}, t)$ where $K = 0, 1, 2, \ldots, N$ and N may be any finite integer. (If we are dealing with fluids, the fields would be velocity, pressure, density and so on.) As already mentioned, all such systems are to be considered deterministic. (Even the Schrödinger equation is deterministic, though it is up to you what you make of its solutions.) So we shall presume that the physical system in question is governed by a system of equations schematically written as

$$\partial_t \mathbf{U} = \boldsymbol{\mathcal{F}}[\mathbf{U}, \boldsymbol{\nabla}] \tag{5}$$

where the right side stands for some (possibly complicated) expression involving the fields and their spatial derivatives. (A negative power of $\boldsymbol{\nabla}$ stands for an integral operator.) Neither time nor time derivatives appear explicitly on the right side of this equation. And let us also be kind to ourselves and suppose that there are no fractional powers or other such inconveniences.

The condition for a stationary solution is

$$\boldsymbol{\mathcal{F}}[\mathbf{U}^{(0)}, \boldsymbol{\nabla}] = \mathbf{0} \tag{6}$$

and it is satisfied by a $\mathbf{U}^{(0)}(\mathbf{x})$ which typically represents a continuum of solutions of (6) and whose global properties provide control parameters for the ensuing dynamics. Stationary solutions may often be found in straightforward calculations by people with the right skills. (Those of a certain age will remember how much effort was once needed to solve for stationary solutions of the stellar structure equations and even those were achieved by questionable treatments of the convection.)

Returning to the full problem presented by (5) (with boundary conditions to be specified), let us consider solutions of the form

$$\mathbf{U}(\mathbf{x}, t) = \mathbf{U}^{(0)}(\mathbf{x}) + \mathbf{u}(\mathbf{x}, t). \tag{7}$$

On inserting this into (5), we may turn it into an equation for the unknown $\mathbf{u}$. After some manipulations, we may normally separate the terms linear and nonlinear in $\mathbf{u}$ so that the

governing equation takes the form

$$\partial_t \mathbf{u} = \mathcal{L}[\mathbf{u}] + \mathcal{N}[\mathbf{u}], \tag{8}$$

where $\mathcal{L}$ and $\mathcal{N}$ are respectively linear and nonlinear operators that depend on the properties of $\mathbf{U}^{(0)}$ and on ∇.

4.1 Linear Theory

Much depends on the nature of the linear operator, so we first examine the associated linear problem, though the procedure does not end with that. In the linear theory, $\mathbf{u}$ represents a small perturbation to the chosen stationary state and so we consider

$$\partial_t \mathbf{u} = \mathcal{L}[\mathbf{u}]. \tag{9}$$

Since $\mathbf{U}^{(0)}$ is a stationary solution, there is again no explicit appearance of t on the right hand side. Of course, since the world is always changing, this is not ever strictly true, but it is often close enough for our purposes. Therefore we need not worry about the slow evolution of the sun as we look for separable solutions of the relevant linear problem (9). These have the form

$$\mathbf{u}(\mathbf{x}, t) = \exp(\lambda t)\boldsymbol{\Upsilon}(\mathbf{x}) \tag{10}$$

where λ is an eigenvalue of

$$\mathcal{L}\,\boldsymbol{\Upsilon} = \lambda \boldsymbol{\Upsilon}. \tag{11}$$

Let us suppose that the spectrum of eigenvalues is discrete. Then we have a countable set of eigenvectors that is either complete or that may be made complete by the inclusion of generalized eigenvectors (see Friedman 1956). The generalized eigenvectors have algebraic as well as exponential dependence on t and are related to the currently much-discussed non-normal modes of stability theory. We do not need to go into these fine points here except to stress that the assumption that the spectrum is discrete is significant. As we'll see later, a continuous spectrum enriches the story.

We express the eigenvalues of (11) as

$$\lambda = \kappa + i\omega, \tag{12}$$

where κ and ω are real. In the region of parameter space where an eigenmode (or eigenvector) has an eigenvalue with $\kappa > 0$, the stationary base state is unstable. That stationary state of the system is characterized by parameters, P_k, that may be regarded as coordinates of a space whose points correspond to the conditions imposed on the system. The linear theory shows that this parameter space is typically partitioned by neutral surfaces on which $\kappa = 0$. These separate the parameter regions in which a given mode is unstable from those in which it is stable. Parameter space with its neutral surfaces thus resembles a blue cheese.

In Fig. 11, we illustrate the situation with two neutral surfaces in a three-dimensional parameter space by showing the surfaces $\kappa = 0$ for the two neutral modes. Such pictures may become complicated in parameter regions where several modes can pass from stability to instability. In such cases it is better to simply indicate the spectrum of neutral modes as in Fig. 12. This is an example of another representation of a spectrum, this time with three neutral modes—two with complex conjugate pairs of eigenvalues and one with $\lambda = 0$. We are considering only the case where, at marginality, all the other modes are stable with

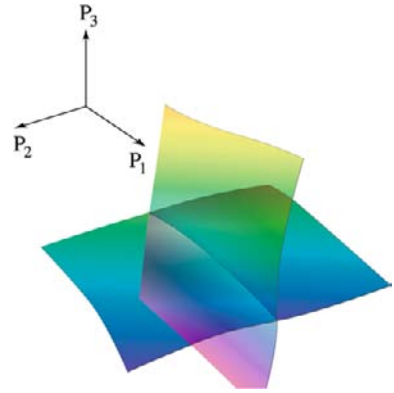


Fig. 11 The structure of parameter space with two neutral surfaces

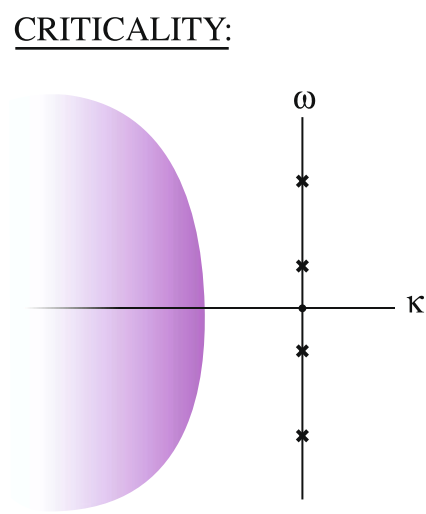


Fig. 12 Illustrating a spectrum for a case with three marginal modes

$\kappa < 0$; their eigenvalues all lie in the shaded region which extends out to $\kappa = -\infty$. Since the eigenvalue spectrum is discrete, there is always some gap between the κ of the least stable mode and $\kappa = 0$.

4.2 Weakly Nonlinear Theory

Now we may enter cautiously into the nonlinear regime by expanding about a point of marginality. For this exploration, we select a region of parameter space where there are no powerful instabilities. This will allow us to study situations in which the nonlinear terms do not become very large and where the analysis is tractable. Of course, the case of the static solar model is one with highly unstable modes so, later, we shall have to do some rationalizing of this choice.

On entering the parameter region of interest from the completely stable side, we locate a point at the onset of instability within the parameter regime we have decided to work in. There may be several of those and we choose one with the greatest number of marginal modes. That is the most singular place in the neighborhood so it is best to deal with it first by expanding from there. If we then nudge the parameter values into the unstable regime,

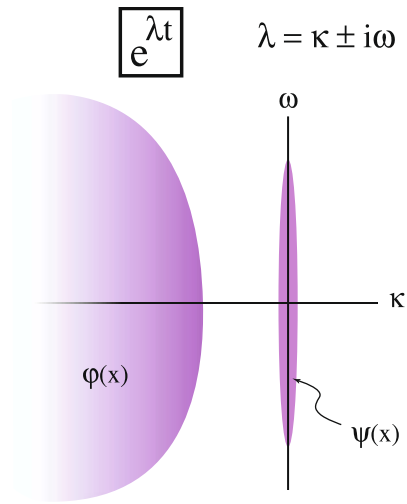


Fig. 13 A slightly unstable situation with the eigenvalues lying in the *shaded regions*

we get a situation something like that in Fig. 13. Here there are two kinds of modes, slow modes denoted as $\boldsymbol{\psi}$ (for pslow) that either grow or decay very slowly in linear theory and modes with relatively fast decay that we call $\boldsymbol{\varphi}$. They lie in their respective shaded regions.

If we do not exceed the marginal condition by too much, the gap between the two kinds of modes remains open, as in Fig. 13. We are skipping over the nuances that may be introduced by the non-normal (or generalized) modes as those are not of immediate interest here though they may be needed to complete the set of basis eigenfunctions. Thus we expand the full solution of (8) as

$$\mathbf{u}(\mathbf{x}, t) = \sum_k A_k(t)\boldsymbol{\psi}_k(\mathbf{x}) + \sum_\ell B_\ell(t)\boldsymbol{\varphi}_\ell(\mathbf{x}) \tag{13}$$

where A_k and B_ℓ are the expansion coefficients indexed in an (I hope) evident way. These coefficients can serve as the coordinates of the state space of the system. We could substitute the expansion for $\mathbf{u}$ into (8) and carry out the usual kinds of projections to get a set of coupled ordinary differential equations for the expansion coefficients. Those describe the trajectory of the system in state space. But a simpler possibility is available when the problem has the features that we have been describing. That is what is behind the marvel that I mentioned at the start of this section.

4.3 Amplitude Equations

A system constituted in the manner we have described and in a state of marginal stability is attracted to an invariant subspace whose dimension is equal to the number of marginal modes. To put this in ordinary language, we may say that the B_ℓ are functions of the A_k; subspaces that such relations describe are known as center manifolds. Once a system arrives at a center manifold, its dynamics are governed by ordinary differential equations for the A_k alone—or rather for suitable coordinates in that center manifold related to the A_k. This (loosely stated) result, called the center manifold theorem, is very intuitive, though I have described it in terms of equations. For example, when computing the evolution of a star it is usual to set the abundances of the rapidly reacting nuclear species to their instantaneous equilibrium values while the slowly reacting species carry the evolution of the star along.

This is just the sort of thing we have been considering for what may be regarded as a dynamical system.

In the parameter neighborhood of the place in parameter space where we are centering our expansion and which is at the onset of instability, we then write the equations for the A_k, transformed into coordinates $\mathcal{A}_k$ adapted to the new subspace, as

$$\dot{\mathcal{A}}_k = M_{jk}\mathcal{A}_k + g_k(\boldsymbol{\mathcal{A}}). \tag{14}$$

Here the matrix $\mathbb{M}$ (with elements M_{ij}) is constructed to possess the same eigenvalues as those of the slow modes. Its construction, once those eigenvalues are known from linear theory, is an exercise in linear algebra. The object g_k is a strictly nonlinear function of $\boldsymbol{\mathcal{A}}$ and the method for its construction is not overly difficult in the case of reasonable looking equations such as the standard fluid equations. I do not go into the derivation of g_k here but it is to be found in many places; if you like spectroscopy try Spiegel (1985). It is customary to put g_k in what is called a normal form, one of several possible standard choices dictated by the eigenvalues of the slow modes. The point of interest to us here is that, in these restricted conditions, there is an equation for the dynamics of the slow modes alone that may be derived by standard procedures. The behavior of the fast modes is also derivable once the amplitudes of the slow modes are found.

I have omitted many *caveats* that may occur to you and am painting a carefully designed, rosy picture. Above all, in the case of the sun, as in much of astrophysics, we have to deal with systems that are well beyond the instability threshold. When there are modes with large growth rates in the system, we confront conditions of great disorder that lead into the problems of large nonlinear terms as in turbulence. Yet, even in turbulent systems, there is often some kind of order. This is rationalized by the argument that turbulent motions transport various fluid properties so effectively as to bring the mean state of the fluid close to marginality. That is how the instability is finally contained. The behavior of the moving parcels of a turbulent fluid is then modeled as if they transported material properties as atoms in a gas do. That vision may be promoted into a coarse-grained vision of a turbulent fluid as a gas made up of vigorously moving constituent particles.

The procedure is then to replace (7) by a decomposition into an average part plus a deviation from that average. On the grounds that such a deviation may not be so large, dimensional reduction may be possible in highly unstable situations, at least for a first look. Perhaps what is done in this way is not as bad as what I have described but, even so, it is not very satisfactory. There is not much of an alternative as yet as we grope for a better understanding of the turbulent process of solar hydromagnetics. For some, this is a justification that may be offered for the belief that a dimensional reduction may be possible in some sense even in highly unstable systems.

5 Grand Minima

With a little experience, it is not so difficult to produce a system of equations whose solutions *qualitatively* imitate the variation of the sunspot number during the active periods. This in itself is not a cause for jubilation for, as Dirac has said (and as all know), "Just because the results happen to be in agreement with observation does not prove that one's theory is correct." We would therefore like to do better and the direction of improvement may be indicated by the (relatively) inactive periods known as grand minima that occur every two hundred years or so. Those intermissions are not shown in the data displayed here but let me

remind you of the great dearth of sunspots in the lifetime of Newton. We would like to have a model of the purely temporal variations that includes such episodes of very low activity. Such a model may be of help in deciding what are the key physical processes in driving the solar cycle. Unfortunately systems described by (1)–(2) do not by themselves generally produce the grand minima that are part of the solar hydromagnetic cycle.

If you skipped the previous section (or even if you didn't) you may be skeptical of the idea that simple models can be made of the internal dynamics of highly turbulent objects such as the sun. The notion that this may be possible rests on an old belief about unstable media. The suggestion is that a very unstable medium generates strong activity, such as transport of material properties, and that this activity exerts a negative feedback that tends to weaken the instability. It rests on the view that turbulence so to say renormalizes the mean state of the medium into a state of near neutrality or of only mild instability. That is why, in theories of stellar structure, the specific entropy in fully convective regions in stars is often taken to be constant, for that is the state of convective neutrality. And with such a wishful presumption we go on to ask what is likely to happen in such nearly marginal states.

The first question is, what might the instabilities be? A simple model process that, in a sense, exhibits instability is the kinematic dynamo. In this model, the fluid motions are specified and the aim is to learn whether an initially infinitesimal magnetic field will be amplified. The term kinematic in the name of the model implies that, if the magnetic field does grow, its effects are not felt by the velocity field. This approach to the dynamo problem resembles a conceptual framework that mathematicians call a skew product structure in which certain feedbacks are left out to simplify the system so that we may better anticipate its behavior. The initial growth of the magnetic field in the kinematic dynamo is like an instability.

When the growing field is allowed to feed back on the motion we confront the full dynamo problem. Then, even if the field grows initially, it may become strong enough to inhibit the motions and quench the dynamo process (Vainshtein and Cattaneo 1992). But if the field is spatially intermittent, that offers a way to evade the quenching problem. That in turn may cause the region of dynamo action to migrate—in latitude say—but we shall not discuss that issue here.

For our present purposes, we need to isolate from among the many instabilities that may plague the outer layers of the sun at least one that has been tamed by feedback into producing an oscillation that resembles the smoothed solar variation of Fig. 9. Fortunately, even if we have not identified the specific instability mechanism, weakly nonlinear theory tells us generally how to describe the oscillations of mildly unstable systems given only the number of the *slowly* growing and decaying modes and their frequencies in linear theory. The underlying philosophy of the approach is outlined in Sect. 4 and a method it suggests is described in Coullet and Spiegel (1983), for example. That approach also permits us to bypass for now the discussion of the long (and as yet unresolved) problem of the possible instabilities that may provoke solar activity. It tells us what the nature of a suitable simple model may be and so help us to surmise what the underlying physics is.

5.1 A Solar Oscillator

In the simplest instance of instability, an infinitesimal perturbation on a slightly unstable mean state of a fluid gives rise to a single growing mode whose amplitude increases like $\exp(\kappa t)$ in time with an oscillation frequency ω (which may be zero) at the onset of instability. Once the amplitude of the mode has grown sufficiently, the nonlinear feedbacks that may be neglected when the modal amplitudes are still infinitesimal come into play and

keep the growth in check. This is a very general situation in which the (complex) amplitude $\mathcal{A}(t)$ for a single slow mode will be governed by a well known equation called the Landau or Hopf-Landau equation. With the amplitude written as $\mathcal{A} = X + iY$, we find, on keeping nonlinear terms only up to cubic,

$$\dot{X} = \kappa X - \omega Y - (X^2 + Y^2)X, \tag{15}$$

$$\dot{Y} = \omega X + \kappa Y - (X^2 + Y^2)Y \tag{16}$$

where the units have been chosen to make the coefficients of the nonlinear terms equal to unity. With only the linear terms retained, these equations describe a simple linear oscillation. For $\kappa > 0$, which is the case of an unstable mode, the oscillator may be thought of as feeling a negative friction—an effect that may be mimicked by a simple mechanical process. The nonlinear terms in this system that keep the amplitude in check are the leading terms in an expansion in amplitude that may be performed when the instability is weak. A model like this may be derived for any growing oscillation or overstability (in Eddington's terminology) as in a simple radial stellar pulsation, for example.

In numerical solutions of (15)–(16) the quantity X^2 varies in time somewhat as the sunspot number does, though it varies periodically. So we take X^2 as the model sunspot number. (X may be negative, as in the Bracewell artifice.) The regularity of those solutions make them unsuitable for descriptions of the aperiodic solar cycle but, as we saw in the very beginning of this discussion, it is not difficult for an unstable oscillator to be made to behave chaotically. If a suitable temporal variation in κ were introduced, as in the chaotic oscillator discussed at the outset, we would have a situation not too different from that of the system discussed in Sect. 3. But such systems generally do not produce semblances of grand minima, so we should enlarge the models. The idea is that the oscillatory mechanism underlying the solar cycle is driven by the main convective dynamo of the sun to be chaotic and intermittent.

5.2 On-Off Intermittency

If, in the oscillator equations (13)–(14), κ varies slowly, we may view the system as provisionally stable or unstable at any time according to the instantaneous value of κ. If the durations of the periods with a given sign are long enough, we may expect to observe intermittency between bursts of oscillatory behavior and occurrences of low activity when the oscillation may seem to have ceased. This mechanism is called on-off intermittency (Platt et al. 1993a) and it was in fact devised with the grand minima in mind. It ascribes irregular occurrences of the observed grand minima to chaotic variations in κ in the context of the model.

For the solar application, it would be well to couple the oscillator equations to a model representing the global dynamo action of the sun. For a schematic version of that approach we may use as a model solar dynamo the shunted disk dynamo (Malkus 1972). As we saw, its dynamics is described by a system of equations that may be put into the form of the third-order equation system described in Sect. 3.3. That is, we drive the oscillator with the system given by (1), (2) and (4). Then we identify κ of (15)–(16) with the variable parameter, α. The five equations (1), (2), (3), (13) and (14) with $\kappa = \alpha$ constitute the model presented in Platt et al. (1993b). From it, we find results like those in Fig. 14a, a plot of X^2, representing the level of solar activity as it varies in time (with arbitrary units); Fig. 14b shows a blowup of one of the eras of high activity of Fig. 14a. The activity represented in this way is episodic and the duty cycle of the active eras may be controlled by the choice of parameter values.

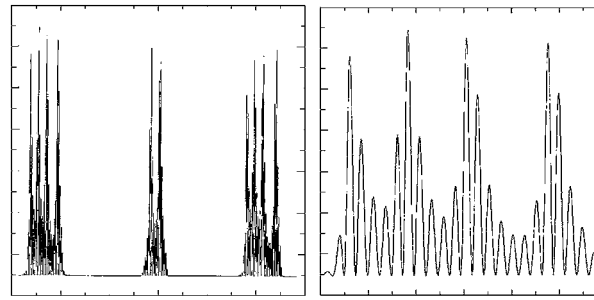


Fig. 14 A plot of X^2 vs. time as in Platt et al. (1993b) with a blowup of one burst

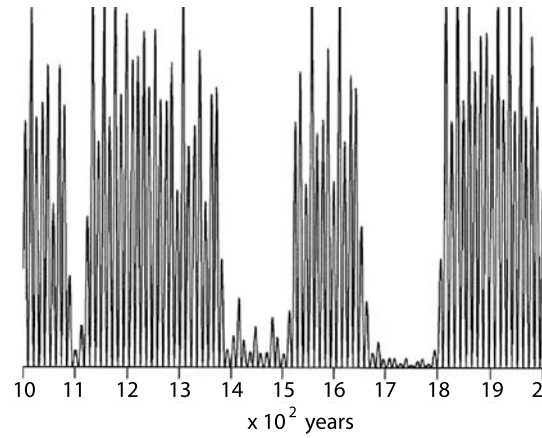


Fig. 15 Model with activity during a grand minimum after Pasquero (1995)

The first models of on-off intermittency consisted of two nonlinear oscillators coupled through their mutual dependence on κ (Spiegel 1981; Platt 1993). The result was a chaotic variation of κ produced by the competitive inputs of the two oscillators. The oscillators were on an equal footing and the model was a sort of dissipative version of the well known Hénon-Heiles system. In the application to the sun, the idea was that one of the oscillators, representing the global convective dynamo, was always on while the other one, representing the solar cycle was to be intermittent. At the time of construction of the model, I (the astronomer in the team) was under the (mis)impression that the cycle switched off completely during the grand minima. The desired asymmetry between the behavior of the two oscillators could be achieved by making the feedback of the oscillation of X on the driver very weak. In Platt et al. (1993b), we suppressed the feedback entirely. In that case, activity during quiescent periods is not seen in plots like those in Fig. 14. When it was realized that during the Maunder minimum, in the time of Louis XIV, there were occasional spots (Ribes and Nesmes-Ribes 1993; Beer et al. 1998) feedback was included in the desired amount in the thesis of Pasquero (1995). When an additional term proportional to X^2 was introduced into (1) to allow feedback on the driver, activity was seen during the minima as in Fig. 15 computed by Pasquero. Thus, the results, seen in Fig. 15 were obtained by replacing (4) with

$$V = \frac{1}{4}x^4 - \frac{\alpha}{2}x^2 - \frac{1}{2}qX^2 \quad \text{and} \quad g = a(x^2 - 1) \tag{4'}$$

where q is a parameter that controls the strength of the feedback.

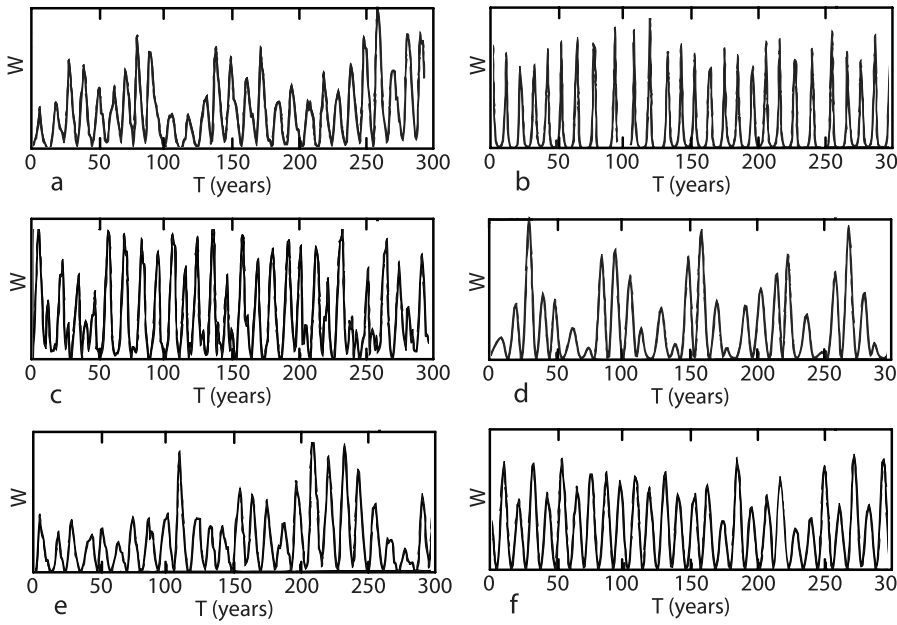


Fig. 16 **a** Wolf number from real data, **b** Tobias et al. (1995), **c** Jones et al. (1984), **d** Feudel et al. (1995), **e** Barnes et al. (1980), **f** Platt et al. (1993b)

Other means of producing grand minima such as modulation effects in dynamo models (Tobias 1996) and nonlinear filtering (Barnes et al. 1980) have been suggested. A comprehensive discussion of various approaches has been given by Pasquero (1995). In this, the challenge is that detailed comparison of theory and observation is not apt because one is, in most instances, dealing with sensitive systems. Hence comparisons of only general behavior can be very meaningful. However, Pasquero's thesis makes interesting comparisons among the various models available at the time of her work. Figure 16 from Pasquero et al. (1995) shows plots of the sunspot number vs. time together with the results of several models.

An issue that must be confronted in the models discussed here and by Pasquero is how to best choose the values of the parameters that go into the models. The group of L.A. Smith at LSE have been studying just this issue and that will figure in cases like the solar cycle with several parameters and complicated solutions.

6 Ruminations

6.1 Simple Oscillators

There is by now plenty to read in the literature about the connection of chaos to the solar cycle. I have provided some examples but have made no attempt at reviewing it all. Rightly or wrongly I have taken to heart the wisdom of Voltaire who wrote that "The secret of being a bore is to tell everything." So I have limited myself to those aspects of this subject that have intrigued me most. Moreover I aimed to make this a qualitative introduction. A good coverage of the range of models available just a decade ago was provided in the thesis of

Pasquero (1995), which was never published. (I must share the blame for that with Antonello Provenzale.) What is discussed both there and herein is the prospect of modeling the observed variation of a measure of solar activity such as the sunspot number. Our view and that of the authors cited above is that this variation can be qualitatively reproduced by the solutions of relatively simple systems of ordinary differential equations.

If you are unfamiliar with the subject of chaos, you may be wondering why I have not shown any detailed comparisons of theory and observation. The answer to that question lies in the term "sensitive systems." Chaotic models are not good predictors of details, except for very short times, and so direct comparison of the details found in theory and observation does not provide a very critical test of a model, especially if the observed process is itself chaotic. As I mentioned, we do not have enough data on the solar variation to decide in a conclusive way whether the solar cycle represents a chaotic process. However, I have no doubt that it is one. Although chaos is not turbulence, fluid turbulence is chaotic (Spiegel 1987). The main question that this poses is what is the best kind of model to describe the solar cycle. I fear that there is no definitive answer to that question; the model to use depends on what we are trying to do with it.

The method of delay coordinates provides a way to get an idea of what an empirical attractor for a given system may be like. It has been used mainly to try to discover the number of degrees of freedom needed to make a meaningful model of the system's behavior. I have said already that there are not enough data on the solar variability to pass muster with many specialists of the method. Still, for the solar case, the method does suggest a form for the underlying attractor (and suggests that its dimension is somewhere between five and six). That, together with nonlinear bifurcation theory, gives us an idea of what a reasonable model might be. Then, as in certain methods of sculpture, once a satisfactory final product is cast, the original model may be put aside. Indeed, being discarded may well be the fate of the simplest models I have been discussing. However, like the first models of sculpture, scientific models are sometimes worth saving for the insight they may provide. The case of the solar magnetic cycle illustrates the point.

The model equations I have focused on here describe a local instability process of an oscillator driven by a chaotic, third-order system that is a stand-in for the main convective dynamo of the sun. This was an extension of work done with D.W. Moore in the sixties on what came to be called chaos. We had been thinking that the irregularity of the solar cycle (and other astrophysical variations) was just another manifestation of that process. But when Eddy (1972) made it clear that the Maunder minimum was a real event that could not be ignored in thinking of the solar cycle we saw that our idea was too naive. The chaotic systems we knew about did not show such behavior so we began to worry about our original interpretation. It seemed likely that we needed to enlarge the model and the idea of a composite system seemed a good prospect.

Our earlier work on chaos was based on a parametric driving of an oscillator as in the third-order systems discussed above. The on-off mechanism was an enlargement of this idea. We already knew of a layer that had appeared in discussions of the spindown of the solar rotation caused by the hydromagnetic torques exerted by the solar wind. I had called that layer the tachycline (a misnomer devised to tease a friend). Though we had met this layer only as a transient in the solar spindown process (Howard et al. 1967), we posited (with no real justification) that it had a long-lived analogue to serve our purposes. The issue of how such a layer can persist below the convection zone is still being discussed (Hughes et al. 2007). Nevertheless, this analogous layer has subsequently been observed. The long-lived version has been named the tachocline with inspiration from the usages of oceanography (Spiegel and Zahn 1992). The idea is that the convection zone feeds the magnetic field into

this layer. The field in the tachocline is then sheared out by differential rotation and erupts in grand arches that twist as they rise upward. This dynamo process, which Elsasser used to call Parker's bathtub mechanism, in its mathematical incarnation is nowadays referred to as the $\alpha - \omega$ dynamo. For its justification by asymptotic analysis of magnetofluid equations, see Childress and Gilbert (1995) and Weiss and Thompson, this issue.

In describing this background I am hoping to suggest that by seeking a mathematical description (in the sense of *applied* mathematics) of an observed process, we may be able to learn from the mathematical description something about the underlying physical workings that we do not understand. In modeling the solar magnetic variation, we were led to suspect the existence of the solar tachocline. Though this is but one small example of an astromathematical procedure, I believe it is worth considering further. In the solar case, if by mathematical modeling of the cycle we can get an idea of the nature of the active instabilities, we can better grapple with the problem of trying to identify their physical origin. Perhaps this is the sort of thing that Dirac had in mind when said that "Mathematics can lead us in a direction we would not take if we only followed up physical ideas by themselves."

6.2 Spatio-Temporal Aspects

So far we have been concerned with models of global or total measures of solar magnetic activity—what engineers called lumped models. Yet we know that the process varies in space as well as in time. Figure 17 is a portion of the so-called butterfly diagram indicating where in time and latitude sunspots are found. (The colors indicate daily sunspot area averaged over individual solar rotations.) In this case, however, the usual plotting style has been very slightly altered. Here, time is the vertical coordinate and the latitudes of the spots detected are shown in the horizontal coordinate. This spacetime plot may be interpreted as showing the progress of a series of solitary waves propagating from mid-latitudes to the equatorial region of the sun. The relatively narrow widths of these waves bespeaks an origin in a narrow layer and I presume that is the tachocline. The spectrum of modes of such a relatively thin layer is likely to be closely spaced so that, for practical purposes, it may be regarded as continuous. From this, one may make propagating wave packets such as seem to be suggested by the butterfly diagram (Proctor and Spiegel

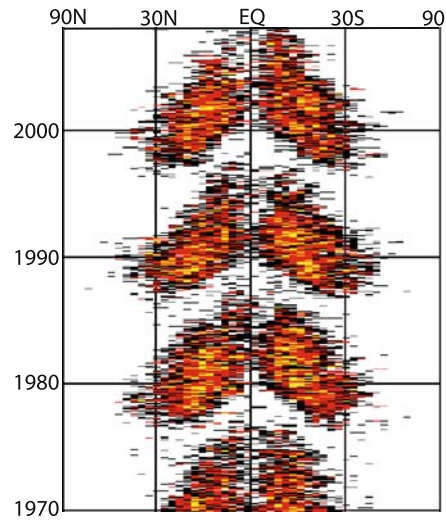


Fig. 17 A spacetime plot of the locus of solar activity. Time is vertical, latitude is horizontal. Adapted from http://solarscience.msfc.nasa.gov/ (courtesy of David Hathaway)

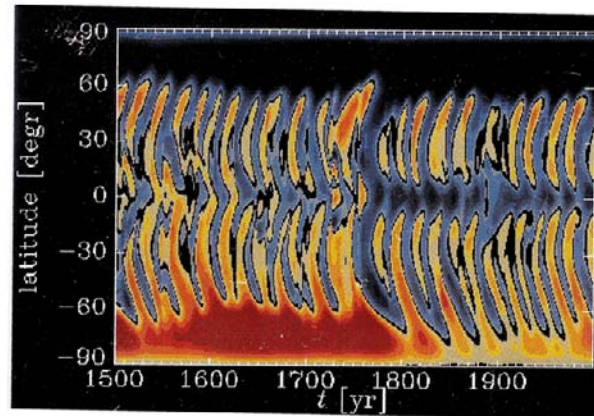


Fig. 18 The butterfly diagram calculated by the Brandenburg program with on-off intermittency included

1991). The mathematical procedures for doing this are well developed (Whitham 1974). These waves may be nonlinear versions of the dynamo waves (Parker 1955) that appear in hydromagnetic simulations based on the $\alpha - \omega$ dynamo process (Brandenburg 2005; Rüdiger and Brandenburg 1995).

When a version of the on-off intermittency mechanism is introduced into calculations of the solar cycle by way of noise in the governing parameters they produce somewhat erratic magnetic cycles. (By "noise" I mean disturbances produced with the aid of a random number generator.) An example of what results from doing that is shown in Fig. 18 (Brandenburg and Spiegel 2008), and this illustrates only a mild version of the process.

A wave train, being a spatio-temporal process, is generally describable by nonlinear partial differential or integro-differential equations. Thus, *prima facie*, its state space has an uncountable number of dimensions. However, the description of a wave train such as we see in the butterfly diagram, or in an analogous nonlinear wave model, can frequently be treated by a form of dimensional reduction when the waves overlap only slightly. The idea is that solitary waves can be thought of as particles in interaction. Their field equations can be reduced to equations of motion for the individual wave packets. These equations involve interactions among the packets as described in the review by Balmforth (1995). When these (quasi)particles are widely separated, the interactions are weak and their governing equations may often be reduced to systems of ordinary differential equations or even algebraic ones in the simplest cases. In other words, a wave train like that in the butterfly diagram may, to a large extent, be modeled by the kind of simple chaotic system we have been concerned with here. Here is another rationale for seeking simple models for the solar cycle that allow for the spatio-temporal aspects.

6.3 The End

Blaise Pascale wrote (in *Pensées*) to a correspondent that if he had more time he would have written a shorter letter. I must plead a similar dilemma. There is more to be said on this topic but time and Voltaire cause me to end here. I close then with the thought that though the sunspot data are limited, there are the surrogate data provided by the spoor of solar activity left on the earth in tree rings, ice cores (Weiss and Thompson, this issue) and records of aurorae (Stothers 1979; Solow 2005). So the continuation is open.

There are also data suggesting magnetic activity on other cool stars than the sun (Baliunas et al. 1995). Nigel Weiss (personal communication) has provided this concise statement of the observational situation:

There are about a dozen slowly rotating G and K stars, like the sun, that display similar periodic behavior. Rapid rotators are much more active and may perhaps exhibit several periodicities; they tend to have polar spots together with activity at low latitudes. It is not known whether fully convective stars show periodic activity, though they are certainly magnetic.

Some of those stars may be fully convective and so lack tachoclines. However, in the solar case, there is a shear layer at the top of the convection zone (Basu and Antia 2001; Antia et al. 2008) that is not much deeper than a granule. Would stars with larger granular structure form deeper shear layers of this kind and produce cycles in them? Such a layer at the top of a convection zone is not fed by plunging plumes, but it might produce some sort of coherent structures nonetheless. The upper shear layer of the sun may be too shallow to play a big role in the solar cycle but such a layer may be responsible for cycles in fully convective stars. Modeling informed by chaos theory may help us unravel such questions as we move to extend the domain of our understanding of stellar hydromagnetics to other stars than the sun.

Acknowledgements I am indebted to Nigel Weiss for and to Antonello Provenzale for their comments. Andre Balogh was a fine editor who even improved some figures. A lost Fig. 2 was handsomely remade by J.-L. Thiffeau.

References

H.M. Antia, S. Basu, S.M. Chitre, Astrophys. J. **681**, 680 (2008)
A. Arneodo, P.H. Coullet, E.A. Spiegel, Geophys. Astrophys. Fluid Dyn. **31**, 1 (1985)
S.L. Baliunas et al., Astrophys. J. **438**, 269 (1995)
N.J. Balmforth, Annu. Rev. Fluid Mech. **27**, 335 (1995)
N.J. Balmforth, R.V. Craster, Chaos **7**, 738 (1997)
J.A. Barnes, H.H. Sargent III, P.V. Tryon, in *The Ancient Sun*, ed. by R.O. Pepin, J.A. Eddy, R.B. Merrill (Pergamon Press, New York, 1980), p. 159
S. Basu, H.M. Antia, Mon. Not. R. Astron. Soc. **324**, 498 (2001)
J. Beer, S. Tobias, N. Weiss, Sol. Phys. **181**, 237 (1998)
J. Bernoulli, Ars conjectandi, in *World of Mathematics*, vol. 3 (1713), p. 1453
R.N. Bracewell, Aust. J. Phys. **38**, 1009 (1985)
A. Brandenburg, Astrophys. J. **625**, 529 (2005)
A. Brandenburg, E.A. Spiegel, Astron. Nachr. **329**, 351 (2008)
S. Childress, A. Gilbert, *Stretch, Twist, Fold: The Fast Dynamo*. Lecture Notes in Physics (Springer, Berlin, 1995)
P.H. Coullet, E.A. Spiegel, SIAM J. Appl. Math. **43**, 774 (1983)
J.P. Eckman, D. Ruelle, Rev. Mod. Phys. **57**, 617 (1985)
J.A. Eddy, Science **192**, 1189 (1972)
U. Feudel, W. Jansen, J. Kurths, Int. J. Bifurc. Chaos **3**, 131 (1995)
A.M. Fraser, H.L. Swinney, Phys. Rev. A **33**, 1134 (1986)
B. Friedman, *Principles and Techniques of Applied Mathematics* (Wiley, New York, 1956)
R. Hoffman, V. Torrence, The Sciences, 29 November–December 1993
L.N. Howard, D.W. Moore, E.A. Spiegel, Nature **214**, 1297 (1967)
D.W. Hughes, R. Rosner, N.O. Weiss (eds.), *The Solar Tachocline* (Cambridge University Press, Cambridge, 2007)
C.A. Jones, N.O. Weiss, F. Cattaneo, Physica D **14**, 161 (1984)
J.B. Keller, Am. Math. Mon. **93**, 191 (1986)
C. Letellier, L.A. Aguirre, J. Maquet, R. Gilmore, Astron. Astrophys. **449**, 2006 (2006)
G.E.R. Lloyd, in *Ancient Cosmologies*, ed. by C. Blackes, M. Loewe (George Allen & Unwin, London, 1975), p. 200
W.V.R. Malkus, E.O.S. Trans. Am. Geophys. Union **53**, 617 (1972)
B.B. Mandelbrot, *The Fractal Geometry of Nature* (Freeman, New York, 1997)
C.J. Marzec, E.A. Spiegel, SIAM J. Appl. Math. **38**, 387 (1980)

E.N. Parker, Astrophys. J. **122**, 293 (1955)
C. Pasquero, Tesi de Laurea, Univ. di Torino, 1995
C. Pasquero, A. Provenzale, E.A. Spiegel, Unpublished manuscript, 1995
N. Platt, in *Proceedings of 1990 GFD Program*. WHOI Tech. Rep. (WHOI-91-93) (1993), p. 327
N. Platt, E.A. Spiegel, C. Tresser, Phys. Rev. Lett. **70**, 279 (1993a)
N. Platt, E.A. Spiegel, C. Tresser, Geophys. Astrophys. Fluid Dyn. **73**, 146 (1993b)
M.R.E. Proctor, E.A. Spiegel, in *The Sun and the Cool Stars*, ed. by D. Moss, G. Rüddiger, I. Tuominen (Springer, Berlin, 1991), p. 117
O. Regev, *Chaos and Complexity in Astrophysics* (Cambridge Univ. Press, Cambridge, 2006)
J.C. Ribes, E. Nesmes-Ribes, Astron. Astrophys. **276**, 549 (1993)
K.A. Robbins, Math. Proc. Camb. Philos. Soc. **82**, 309 (1997)
G. Rüdiger, A. Brandenburg, Astron. Astrophys. **296**, 557 (1995)
D. Ruelle, F. Takens, Commun. Math. Phys. **20**, 167 (1971)
L.A. Smith, *Chaos: A Very Short Introduction* (Oxford University Press, New York, 2007)
L.A. Smith, C. Ziehmann, K. Fraedrich, Q. J. R. Meteorol. Soc. **125**, 2855 (1999)
R.A. Solow, Earth Planet. Sci. Lett. **232**, 67 (2005)
C. Sparrow, *The Lorenz Equations: Bifurcations, Chaos and Strange Attractors* (Springer, New York, 1982)
E.A. Spiegel, Ann. N.Y. Acad. Sci. **357**, 305 (1981)
E.A. Spiegel, in *Chaotic Behavior in Astrophysics*, ed. by R. Buchler, J. Perdang, E.A. Spiegel (Reidel, Dordrecht, 1985), p. 91
E.A. Spiegel, Proc. R. Soc. Lond. **413**, 87 (1987)
E.A. Spiegel, in *Past and Present Variability of the Solar-Terrestrial System: Measurement, Data Analysis and Theoretical Models*, ed. by G. Cini (IOS Press, Amsterdam, 1997), p. 311
E.A. Spiegel, A.N. Wolf, Ann. N.Y. Acad. Sci. **497**, 55 (1987)
E.A. Spiegel, J.P. Zahn, Astron. Astrophys. **265**, 106 (1992)
R.B. Stothers, Astrophys. J. **77**, 121 (1979)
F. Takens, in *Dynamical Systems and Turbulence*, ed. by D.A. Rand, L.-S. Young. Lecture Notes in Physics, vol. 898 (Springer, Berlin, 1981), p. 366
J. Thurber, *The New Yorker*, March 1 (1947), p. 26
S.M. Tobias, Astron. Astrophys. **307**, L21 (1996)
S.M. Tobias, N.O. Weiss, V. Kirk, Mon. Not. R. Astron. Soc. **273**, 1150 (1995)
G.E. Uhlenbeck, G.W. Ford, *Lectures in Statistical Mechanics* (Am. Math. Soc., Providence, 1963)
S.I. Vainshtein, F. Cattaneo, Astrophys. J. **393**, 165 (1992)
G.B. Whitham, *Linear and Nonlinear Waves* (Wiley-Interscience, New York, 1974)
A. Wolf, J.B. Swift, H.L. Swinney, J.A. Vastanio, Physica D **16**, 285 (1985)

The Solar Dynamo

N.O. Weiss · M.J. Thompson

Originally published in the journal Space Science Reviews, Volume 144, Nos 1–4, 53–66.
DOI: 10.1007/s11214-008-9435-z

Abstract It is generally accepted that the strong toroidal magnetic fields that emerge through the solar surface in sunspots and active regions are formed by the action of differential rotation on a poloidal field, and then stored in or near the tachocline at the base of the Sun's convection zone. The problem is how to explain the generation of a reversed poloidal field from this toroidal flux—a process that can be parametrised in terms of an α-effect related to some form of turbulent helicity. Here we first outline the principal patterns that have to be explained: the 11-year activity cycle, the 22-year magnetic cycle and the longer term modulation of cyclic activity, associated with grand maxima and minima. Then we summarise what has been learnt from helioseismology about the Sun's internal structure and rotation that may be relevant to our subject. The ingredients of mean-field dynamo models are differential rotation, meridional circulation, turbulent diffusion, flux pumping and the α-effect: in various combinations they can reproduce the principal features that are observed. To proceed further, it is necessary to rely on large-scale computation and we summarise the current state of play.

Keywords Sun · Sunspots · Magnetic fields · Dynamos

1 Introduction

The principal task of dynamo theory, as applied to the Sun, is to explain the systematic global behaviour of solar magnetic fields. In this broad-brush review, we begin by summarising the key properties of magnetic fields that are observed at and above the surface of the Sun, including not only the cyclic variation of solar activity but also its modulation on much longer

N.O. Weiss
DAMTP, Centre for Mathematical Sciences, Wilberforce Road, Cambridge CB3 0WA, UK
e-mail: now@damtp.cam.ac.uk

M.J. Thompson (✉)
School of Mathematics & Statistics, University of Sheffield, Sheffield S3 7RH, UK
e-mail: Michael.Thompson@sheffield.ac.uk

M.J. Thompson et al. (eds.), *The Origin and Dynamics of Solar Magnetism*.
DOI: 10.1007/978-1-4419-0239-9_4

time-scales. It is clear that these fields are generated below the photosphere, at levels where they have not as yet been directly measured, but helioseismology has succeeded in revealing the possibly associated patterns of motion in the solar interior, including both differential rotation and meridional flows. In Sect. 3 we describe these results, with some emphasis on the slender solar tachocline, which is generally agreed to play an essential part in the dynamo process (Tobias and Weiss 2007a). In the next section we briefly review the current state of mean field dynamo theory, before going on, in Sect. 5, to discuss the various models that have been proposed in order to describe the solar dynamo. Then, in Sect. 6, we consider progress beyond mean field models and towards direct numerical simulation of stellar dynamos. In the final section we comment briefly on attempts to predict the future amplitude of solar activity and try to estimate the expected lifetime of the current grand maximum. Solar and stellar dynamos have already been the subject of several recent reviews (e.g. Tobias 2002; Ossendrijver 2003; Choudhuri 2003; Rüdiger and Hollerbach 2004; Charbonneau 2005; Solanki et al. 2006).

2 Magnetic Activity on the Sun

Sunspots are the most striking manifestations of solar activity (Thomas and Weiss 2008) and their incidence, as measured by area occupied or the traditional sunspot number R, follows the irregular 11-year activity cycle that is best demonstrated by the butterfly diagram in Fig. 1. The true nature of solar activity was only revealed when Hale discovered that dark spots are the sites of strong magnetic fields. Sunspots typically occur as a pair, aligned approximately with a parallel of latitude but with the leading spot (in the sense of the solar rotation) usually somewhat closer to the equator. The leading and following spots have

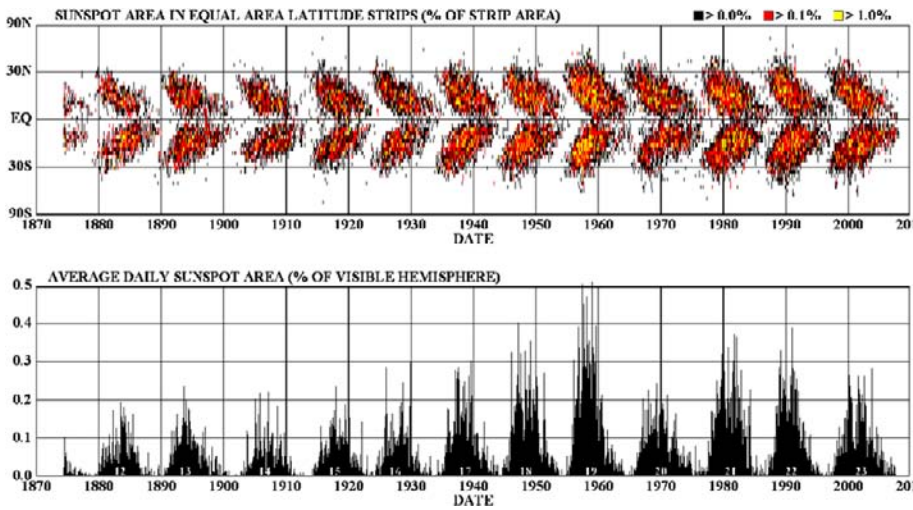


Fig. 1 Cyclic activity on the Sun since 1874. The *lower panel* shows the daily sunspot area, as a percentage of the visible hemisphere that is covered by sunspots and averaged over individual solar rotations, which varies quasiperiodically with an 11-year period. The butterfly diagram in the *upper panel* shows the corresponding incidence of sunspots as a function of latitude and time. At the beginning of a new cycle, spots appear around latitudes of $\pm 30°$. The activity zones spread until they extend to the equator, and then gradually die away, disappearing at the equator as the first spots of the next cycle appear at higher latitudes. (Courtesy of D.H. Hathaway)

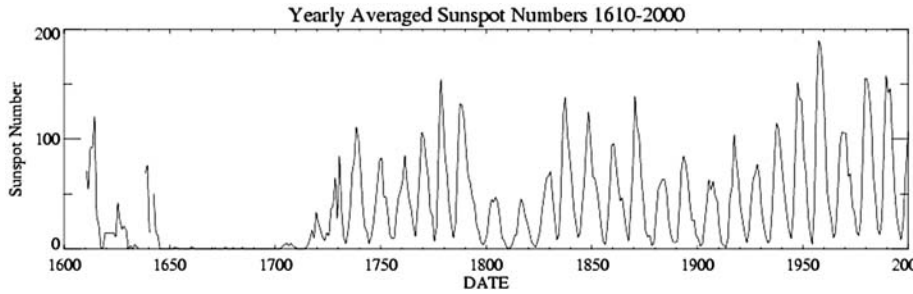


Fig. 2 Annual values of the sunspot number from the first telescopic observations in 1610 to 2000. The amplitude of cyclic activity has varied irregularly, with a prominent interval of inactivity in the seventeenth century—the Maunder Minimum. (Courtesy of D.H. Hathaway)

opposite magnetic polarities, and the appearance of sunspot pairs at the photosphere is best understood as caused by the emergence of a toroidal flux tube from deep in the underlying convection zone. Hale found that the polarities of the spots are consistent in each hemisphere but antisymmetric about the equator. Moreover, these polarities reverse at the end of each 11-year activity cycle, so that there is a magnetic cycle with a period of 22 years. The sunspots are associated with active regions, and there are smaller scale fields on a wide range of scales all over the solar surface. Near the poles there are weaker unipolar fields that are most prominent at sunspot minimum, and reverse at sunspot maximum.

The cycles in Fig. 1 are asymmetric, with the sunspot area, and the corresponding sunspot number R, rising more rapidly than it falls; it is also apparent that stronger cycles rise faster and are shorter. The longer record of R in Fig. 2 shows that the cycle's amplitude has varied widely during the past 300 years: the peak value of R approached 200 in 1958, but in the early nineteenth century the maxima reached no more than 50. Far more striking is the interval from 1645 to 1715—the Maunder Minimum (Eddy 1976; Ribes and Nesme-Ribes 1993)—when sunspots almost completely disappeared.

To investigate the long-term modulation of solar activity we must turn to proxy data—and fortunately such datasets exist. The magnetic fields that are carried out into the heliosphere by the solar wind deflect galactic cosmic rays. When these energetic particles impinge on the earth's atmosphere they give rise to the production of radioactive isotopes such as ^{14}C (which is preserved in trees) and ^{10}Be (which is preserved in polar icecaps) whose abundances therefore vary in antiphase with solar magnetic activity. Tree rings can be dated and seasonal variations are apparent in ice cores; thus the 11-year activity cycle can be followed back for hundreds of years, while its envelope has been established for many millennia. Figure 3 shows how this envelope of solar magnetic activity has varied over the past 9000 years. The apparently chaotic pattern of modulation, giving rise to grand maxima and grand minima, persists throughout this period. Power spectra nevertheless reveal persistent periodicities in both records (Damon and Sonett 1991; Stuiver and Braziunas 1993; Beer 2000; Wagner et al. 2001): in addition to the basic 11 year (Schwabe) cycle, there are well-defined peaks corresponding to periods of 88 years (the Gleissberg cycle), 210 years (the de Vries cycle) and 2300 years (the Hallstatt cycle). As Figs. 1 and 2 show, we happen to be living during an episode of exceptionally high activity—but this episode is not unique, and there have been several comparable intervals within the past millennium.

It is also worth pointing out that similar patterns of cyclic activity can be detected in other slowly rotating late-type stars with deep convection zones, like the Sun. There are also

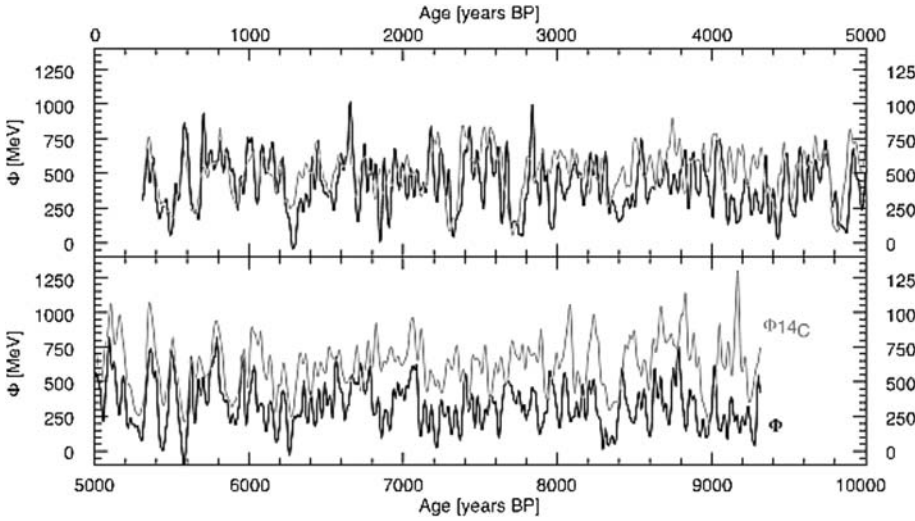


Fig. 3 Modulation of solar activity over a 9000 yr interval from 304 to 9315 BP (1646 AD to 7365 BC), as shown by the solar modulation potential Φ derived from ^{10}Be abundances in the GRIP ice core (*black line*) and that derived from the ^{14}C production rate in tree rings (*grey line*). Both records are high-pass filtered to reduce the effects of changes in the geomagnetic field and other long-term variations. The two records are initially in close agreement, though they gradually drift apart. (From Vonmoos et al. (2006))

examples of similar stars that are magnetically quiescent, and perhaps undergoing grand minima.

3 Helioseismology and Internal Properties of the Sun

Helioseismology uniquely is capable of imaging the solar interior. Waves that are essentially acoustic (modified by the stratification and by magnetic fields) are generated by turbulent motions in the upper part of the convection zone. These waves set up resonant global modes of the whole Sun, and the frequencies of these modes are used to infer properties of the solar interior. The local properties of the wave fields can also be used to infer subsurface conditions under localised features such as sunspots. In this section we summarise the helioseismic findings relevant to the solar dynamo.

Solar models indicate that in the outer envelope of the Sun the temperature gradient required to transport the observed luminosity by radiation is too steep to be stable, thus requiring this region to be convectively unstable. The transport of heat by convection is efficient and so the resulting stratification throughout the bulk of the convection zone in such models is very close to being adiabatic. These model predictions are confirmed by helioseismology. (A good review of the helioseismological results on the solar structure and on internal abundances is that of Basu and Antia (2007).) By measuring the gradient of sound speed, which changes rather abruptly where the heat transport changes between radiation and convection, Christensen-Dalsgaard et al. (1991) located the base of the convective envelope at a fractional radius $r/R_\odot = 0.713 \pm 0.003$. More recent work has confirmed this result with even higher precision (Basu 1998). Moreover, the location of the base of the convection zone appears to be independent of latitude (Basu and Antia 2001).

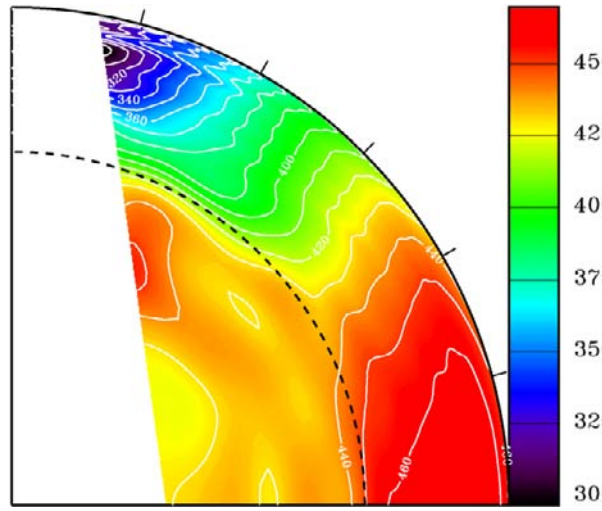


Fig. 4 The internal rotation of the Sun, as determined from observations by the MDI instrument on board the SOHO satellite. The solar equator is along the *horizontal axis*, the pole along the *vertical axis*. Values of $\Omega/2\pi$ are shown, in nHz. The *dashed line* indicates the base of the convection zone, and *tick marks* are at 15° intervals in latitude. (From Thompson et al. (2003))

This measurement actually refers to the extent of the essentially adiabatically stratified region: this may include a region of convective overshooting insofar as the overshoot region is also adiabatically stratified. Simplistic models of convective overshooting at the base of the solar convection zone would indicate that beneath the adiabatically stratified region there would then be a rather sharp adjustment to the subadiabatic gradient required to transport flux in the radiative interior. Such an abrupt transition would give rise to a characteristic signature in the p-mode frequencies, which can be used to measure the location and sharpness of the transition. Analyses of the observed frequencies have indicated that the extent of overshooting of this nature, if any, is small, no more than about one tenth of a pressure scale-height (Basu and Antia 1994; Monteiro et al. 1994; Christensen-Dalsgaard et al. 1995). One possibility is that the overshoot region is corrugated, which would appear smoother when spherically averaged as it is in this seismic analysis. Another possibility is that the overshoot produces a gentler subadiabatic transition: such models have been realised by Rempel (2004).

A major achievement of helioseismology has been to determine the internal rotation of the Sun over much of the solar interior (Fig. 4). It has long been known that at the surface the Sun rotates faster at the equator than at high latitudes: helioseismology has shown that this latitudinal differential rotation persists through the convection zone, with contours of isorotation mostly aligned nearly radially. There are two regions of prominent radial shear: a near-surface shear layer and a layer now known as the tachocline between the convection zone and the radiative interior beneath. The results for the radiative interior are consistent with solid-body rotation, though the rotation of the inner, energy-generating core is still uncertain. Hence the tachocline is not only a region of radial shear but also a transition layer from latitudinal differential rotation to latitudinally independent rotation.

The helioseismic results for the tachocline region are summarised in the review by Christensen-Dalsgaard and Thompson (2007). Kosovichev (1996) obtained the first quantitative results on the location and thickness of the tachocline. He adopted a functional dependence $\Phi(r)$ for the transition in depth of the latitudinal differential rotation of the form

$$\Phi(r) = \frac{1}{2}\left[1 + \mathrm{erf}\left(2(r - r_c)/w\right)\right], \tag{1}$$

where erf is the error function. This provides for a continuous step function varying monotonically from zero to one, with a characteristic width w and centred on radial location $r = r_c$. Kosovichev found that the tachocline thus defined was centred at $r_c/R_\odot = 0.692 \pm 0.05$ and had width $w/R_\odot = 0.09 \pm 0.04$. More recent determinations are summarised by Christensen-Dalsgaard and Thompson (2007). The modern result of Basu and Antia (2003), converted to the same functional form as used by Kosovichev, is that the tachocline location varies from $r_c/R_\odot = 0.692 \pm 0.002$ at low latitudes to $r_c/R_\odot = 0.710 \pm 0.002$ at 60° latitude. At the same latitudes the corresponding tachocline widths are $w/R_\odot = 0.033 \pm 0.007$ and $w/R_\odot = 0.076 \pm 0.010$, respectively. Thus the tachocline is prolate and its width is greater at high latitudes: at low latitudes the tachocline is essentially wholly beneath the base of the convection zone, whereas at high latitudes it is roughly centred upon it.

Since aspects of the Sun vary with the solar cycle, it may be expected that the frequencies of the solar oscillations may vary also, at some level. Indeed the frequencies do vary with the solar cycle, and the variability has a very strong correlation with surface magnetic activity. Moreover, analyses of where the frequency variations originate indicate that the predominant causes of variability are located in the very superficial outer layers of the Sun (Antia et al. 2001). There has been no accepted direct seismological detection of magnetic fields in the tachocline or deeper interior. There are only upper bounds from helioseismology on the possible field strength in and near the tachocline, and these are of order several hundred kilogauss since that is the strength of field required at those depths to get any sensible frequency variation (Roberts and Campbell 1986). The presence of a magnetic field may be detectable through its effect on the thermal structure or on the bulk fluid motions. There have been hints of temporal variations in the wave-speed in the region (Chou and Serebryanskiy 2002; Baldner and Basu 2008). There have also been detections of modulations in the rotation rate just above and beneath the base of the tachocline, with an apparent quasi-periodic 1.3-year oscillation (Howe et al. 2000a). More certain are the relatively weak banded zonal flows, sometimes called torsional oscillations, that were first detected in the surface rotation rate and have since been revealed by helioseismology to penetrate at least one third of the way down through the convection zone, and possibly even to its base (Howe et al. 2000b; Vorontsov et al. 2002; Antia et al. 2008). At mid- and lower latitudes the banded flows migrate towards the equator over the solar cycle, in step with the active latitudes. At higher latitudes there is a poleward branch whose strength waxes and wanes with the cycle.

There is a greater possibility to detect magnetic fields directly in the upper convection zone than in the tachocline, since their dynamical importance is likely to be greater there. There is a persistent suggestion of a wave-speed anomaly at about 60° latitude and at a fractional radius of about $r/R_\odot = 0.9$, which could be caused by a fractional thermal perturbation of about 10^{-4} or a magnetic field strength of about 50 kG (Antia et al. 2000; Dziembowski et al. 2000). However, one might expect such a feature, if magnetic, to vary with the solar cycle, and little variation is discernable (Antia et al. 2003). Local helioseismic techniques such as ring analysis and time-distance helioseismology have detected flows and apparent thermal or magnetic anomalies under sunspots and active regions (e.g. Zhao and Kosovichev 2003; Haber et al. 2004).

Local techniques have also succeeded in revealing the subsurface meridional flow in the upper convection zone. Longitudinally and temporally averaged over a couple of months, the flows are poleward and largely steady from one epoch to another. Over the outer 15 Mm or so the flow speed is also largely independent of depth, at about 20–30 m s^{-1} (Haber et al. 2002). There are indications that in the approach to the last solar maximum the meridional flow in the northern hemisphere developed a counter-cell at mid-latitudes, with poleward flow persisting at those latitudes only in a superficial surface layer. Unfortunately the evidence from helioseismology for what the meridional flow is at greater depths is as yet inconclusive.

4 Dynamo Theory

Transient amplification of magnetic fields by sheared transverse flows (for example, by differential rotation) is distinct from dynamo action (Mestel 1999). There is also an important distinction between small-scale dynamo action, which allows the generation of disordered magnetic fields, averaging to zero, by turbulent convection, for instance in the solar photosphere (Cattaneo 1999; Vögler and Schüssler 2007), and the large-scale global dynamos with which we are concerned. In a planet like the Earth, with a resistive decay time $\tau_\eta \approx 7000$ yr in its core, a dynamo is needed to maintain a magnetic field that has been present for at least 3.5×10^9 yr and reverses on a time scale much longer than τ_η. In the Sun, by contrast, $\tau_\eta \approx 10^{10}$ yr, and yet there is a rapidly oscillating field. A straightforward hydromagnetic oscillator can be ruled out, for it would require variations in velocity with a 22 yr period, and there is no sign of them. The challenge to dynamo theory is then to explain both the cyclic variations of solar magnetic activity and their longer term modulation.

Cowling's theorem prohibits the maintenance of a purely axisymmetric magnetic field. Most dynamo models have treated azimuthally averaged fields, which can be split into a poloidal (meridional) component $\mathbf{B}_\mathrm{P} = \boldsymbol{\nabla} \times (A\mathbf{e}_\phi)$ and a toroidal (azimuthal or zonal) component $\mathbf{B}_\mathrm{T} = B_\phi \mathbf{e}_\phi$. Then we expect the overall field to exhibit dipole symmetry, with B_ϕ antisymmetric about the equator. The principal mechanisms responsible for maintaining these fields are differential rotation, which generates $\mathbf{B}_\mathrm{T}$ from $\mathbf{B}_\mathrm{P}$ (as first realised by Ferraro, Walén and Cowling) and cyclonic eddies, which give rise to the α-effect (first introduced by Parker) and can generate a reversed $\mathbf{B}_\mathrm{P}$ from $\mathbf{B}_\mathrm{T}$. These processes have to compete with enhanced turbulent diffusion, the β-effect.

These models rely on mean field dynamo theory (Moffatt 1978; Parker 1979; Krause and Rädler 1980; Roberts 1994; Mestel 1999) in its simplest form. The idea here is to separate the magnetic field $\mathbf{B}$ and the velocity $\mathbf{U}$ into mean and fluctuating parts and to define a suitable averaging procedure, so that

$$\mathbf{B} = \overline{\mathbf{B}} + \mathbf{b}, \qquad \mathbf{U} = \overline{\mathbf{U}} + \mathbf{u}, \tag{2}$$

where $\langle \mathbf{b} \rangle = \langle \mathbf{u} \rangle = 0$. Then the averaged induction equation takes the form

$$\partial \overline{\mathbf{B}} / \partial t = \boldsymbol{\nabla} \times (\overline{\mathbf{U}} \times \overline{\mathbf{B}}) + \boldsymbol{\nabla} \times \boldsymbol{\mathcal{E}} + \eta \nabla^2 \overline{\mathbf{B}}\,, \tag{3}$$

where $\boldsymbol{\mathcal{E}} = \langle \mathbf{u} \times \mathbf{b} \rangle$ and the magnetic diffusivity η is assumed to be uniform. The procedure then is to assume a separation of length scales so that we can write

$$\mathcal{E}_i = \alpha_{ij} B_j + \beta_{ijk} \frac{\partial B_j}{\partial x_k} + \cdots. \tag{4}$$

If we separate α_{ij} into a symmetric part $\overline{\alpha}_{ij}$ and an antisymmetric part $\gamma_k \epsilon_{ijk}$, and consider pseudo-isotropic (non-mirror symmetric) turbulence we may then set $\overline{\alpha}_{ij} = \alpha \delta_{ij}$ and $\beta_{ijk} = \beta \epsilon_{ijk}$, with the pseudo-scalar $\alpha \neq 0$, whence (3) becomes

$$\partial \overline{\mathbf{B}} / \partial t = \boldsymbol{\nabla} \times [(\overline{\mathbf{U}} + \boldsymbol{\gamma}) \times \overline{\mathbf{B}}] + \boldsymbol{\nabla} \times (\alpha \overline{\mathbf{B}}) + (\beta + \eta) \nabla^2 \overline{\mathbf{B}}. \tag{5}$$

We expect that $\beta \gg \eta$, though molecular diffusion is none the less an essential part of the dynamo process. The α-effect typically depends on the presence of 'gyrotropic' turbulence with a net kinetic helicity $H = \langle \mathbf{u} \cdot \boldsymbol{\nabla} \times \mathbf{u} \rangle$ in the small-scale motion. By further making a quasilinear approximation (first order smoothing), and assuming that the magnetic Reynolds

number $R_m \gg 1$ but the correlation time $\tau_c \ll \ell/u$, where u, ℓ are typical values of the velocity and length scale, it can be shown that $\alpha \approx -\frac{1}{3}\tau_c H$, while $\beta \approx \frac{1}{3}u^2\tau_c$ (e.g. Mestel 1999). Analogous results hold if $R_m \ll 1$, with τ_c replaced by the Ohmic decay time ℓ^2/η. If neither of these conditions is satisfied, there is no straightforward relationship between H and α.

The mean field induction (5) is widely used, though it should be borne in mind that this set of approximations can only be justified if there is indeed a separation of scales and either the magnetic Reynolds number $R_m \ll 1$ or $\tau_c \ll \ell/u$. Neither of the latter conditions is valid in the Sun. It follows then that we can only regard the mean field coefficients as physically plausible parametrisations of turbulent processes in the convection zone of a star.

5 Solar Dynamo Models

Mean field dynamo models rely on a number of ingredients, only two of which are constrained by observations. The most obvious contribution is from differential rotation, especially in the tachocline, where the angular velocity Ω is known from helioseismology, as explained in Sect. 3, and the shear gives rise to an ω-effect. Next is meridional flow, revealed by both helioseismic and surface measurements, which show an average poleward flow near the surface; continuity requires that this flow must reverse at some depth—and it may reverse more than once (Mitra-Kraev and Thompson 2007)—but its speed at the base of the convection zone is not determined. Then there is pumping of magnetic flux down the gradient of turbulent intensity and into the tachocline (the γ-effect: Tobias et al. 2001; Dorch and Nordlund 2001). (The actual roles of rotation, shear and penetration are discussed by Tobias elsewhere in these proceedings.) Finally come turbulent diffusion (the β-effect) and, most importantly, the crucial α-effect. Traditionally, this last has been ascribed to helicity produced by the effect of the Coriolis force acting on turbulent convection. Unfortunately, recent numerical studies by Cattaneo and Hughes (2006, 2008) have shown that α is negligible for convection in a rotating Boussinesq layer (though small-scale dynamo action is readily found). A further complication is that α is liable to catastrophic quenching in the nonlinear regime, so that its effective value becomes

$$\alpha_{\rm eff} = \frac{\alpha}{1 + R_m^q B^2/B_0^2}\ , \tag{6}$$

with $0 < q \le 2$, where B_0 is the equipartition field strength (Vainshtein and Cattaneo 1992; Diamond et al. 2005; Hughes 2007a, 2007b). In the Sun, where $R_m \gg 1$, this would imply that α is quenched when the mean field B is less than 1 G. Numerical experiments on turbulence driven by helical forcing (Cattaneo and Hughes 1996) and on rotating compressible magnetoconvection (Ossendrijver et al. 2001) provide support for such catastrophic quenching, with $q = 1$. As an alternative to a distributed α-effect, it has therefore been proposed that the Coriolis force, acting on instabilities driven by magnetic buoyancy at or near the tachocline, may generate helicity and so produce a net α-effect (Ferriz-Mas et al. 1994; Schmitt et al. 1996; Brandenburg and Schmitt 1998; Thelen 2000a, 2000b).

It is generally accepted that the strong toroidal fields that emerge in active regions must be stored at the base of the convection zone, in or near the stably stratified tachocline, where the effects of turbulent pumping and flux expulsion can hold large-scale fields down against magnetic buoyancy. The tachocline is also the obvious site of the ω-effect, though it should be noted that the radial and latitudinal gradients of Ω are of comparable importance there

if, as seems likely, $|B_\theta| \gg |B_r|$. The near-surface radial gradient of Ω may also be at least locally significant.

Opinions differ as to the site and origin of the α-effect. Some hold that it is distributed throughout the convection zone, though this meets with the difficulties cited above, and Brandenburg (2005) has argued in favour of a near surface dynamo—see his contribution to these proceedings. Others incorporate a surface α-effect into a flux-transport dynamo, developed from Babcock's (1961) phenomenological model, as extended by Leighton (1969). The crucial component is the observed equatorward inclination of sunspot groups and active regions, caused presumably by Coriolis effects as flux rises upward through the convection zone. Combined with surface diffusion, ascribed to supergranular motion, and the observed meridional flow, this tilt can explain the observed evolution of photospheric magnetic fields, including reversals of the fields at the poles. The same meridional flows, acting as a conveyor belt and assisted by turbulent diffusion, then somehow manage to transport the reversed poloidal fields down to the tachocline, to serve as a seed for the next cycle (e.g. Dikpati and Charbonneau 1999; Choudhuri 2003). Dikpati and Gilman describe the most developed model of this process (Dikpati et al. 2004) in their contribution to these proceedings.

The alternative, following Parker (1993), is to adopt an interface model, with all the essential dynamo processes concentrated around the tachocline, where α can arise either from local convective motion or, more likely, from the nonlinear development of magnetic buoyancy instabilities (Tobias and Weiss 2007a). Photospheric fields then serve only as indicators of the action down below.

The solar dynamo is intrinsically nonlinear: in a nonlinear dynamo model, growth of the field must be limited by some dynamical process, whether by non-catastrophic α-quenching or modification of the ω-effect by the Lorentz force. Evidence for the latter process comes from the observed zonal shear flows (torsional oscillations), with a period of 11 years, described in Sect. 3, which seem to extend throughout the convection zone. Some models invoke microdynamic effects (Λ-quenching) of the Lorentz force (e.g. Kitchatinov et al. 1994; Rüdiger and Hollerbach 2004), while others rely on the macrodynamic Malkus-Proctor effect, driven by the mean field (e.g. Covas et al. 2004, Bushby 2005). Figure 5 shows the results of an idealised calculation, where the change in the sign of $\partial\omega/\partial r$ at mid-latitudes leads to a poleward branch of dynamo waves at high latitudes, as well as the stronger equatorward waves that give rise to sunspots; the high- and low-latitude branches of zonal shear flows are present in the observations. Periodic and aperiodic modulation of cyclic activity has been found in several nonlinear models with spherical geometry (e.g. Küker et al. 1999; Pipin 1999; Bushby 2006).

What then is the current status of mean-field dynamo models of the solar cycle? On the one hand, it is clear that there is a plethora of different models that yield plausible results when the arbitrary parameters are carefully tuned; it is reassuring that the various codes are now being benchmarked and compared (Jouve et al. 2008). On the other hand, although these models are certainly *instructive*, they are only *illustrative*.

6 Beyond Mean-Field Dynamos

The obvious route ahead is via direct numerical simulation but the alternative is to adopt a minimal formulation of the problem. Low-order models can exhibit generic patterns of behaviour, although they lack any detailed predictive power. The third-order normal form equations for a saddle-node/Hopf bifurcation are structurally stable and describe the relevant bifurcation sequence (Tobias et al. 1995). As an appropriate control parameter (the dynamo

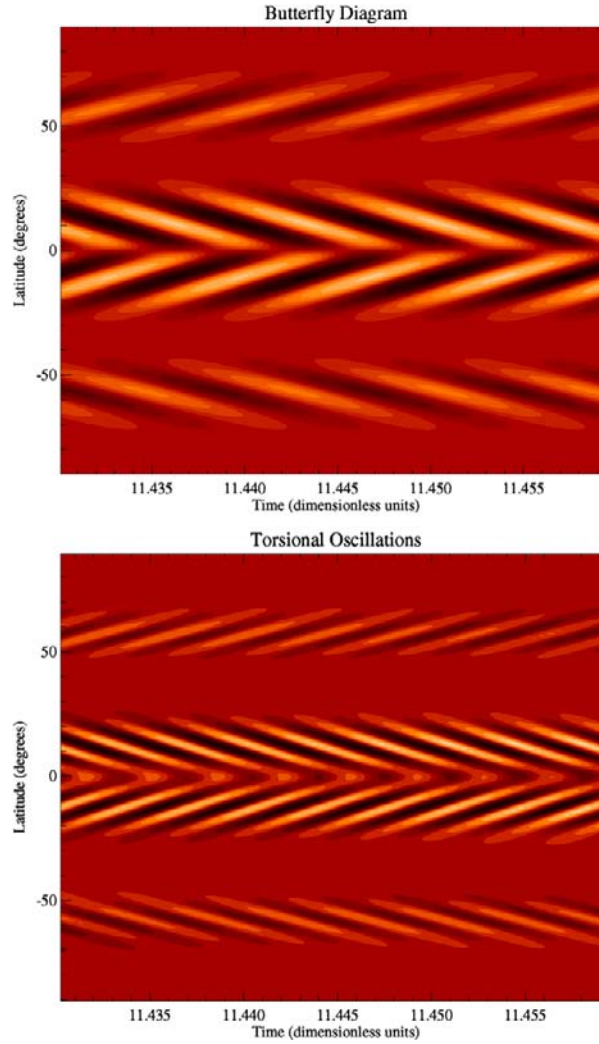


Fig. 5 Nonlinear cyclic behaviour for an idealised model of a spherical interface dynamo, incorporating the macrodynamic Malkus-Proctor effect. *Upper panel*: butterfly diagram showing toroidal fields of opposite signs (with dipole symmetry) at the base of the convection zone. *Lower panel*: the corresponding zonal shear flows (torsional oscillations) with twice the frequency of the magnetic cycle. Note the presence of a polar branch in each panel. (From Bushby (2005))

number D) is increased, there is a Hopf bifurcation from a purely hydrodynamic field-free state to a periodic solution, with trajectories that lie on a limit cycle in phase space; this is followed by a second bifurcation, leading to doubly periodic modulated solutions, with trajectories lying on a two-torus; as D is further increased, the torus is destroyed and a complicated series of bifurcations result in the appearance of chaotically modulated oscillations. The same pattern of *in-out* intermittency (Ashwin et al. 1999; Covas et al. 2001) appears in mean-field dynamo models too (Tobias 1996, 1997; Beer et al. 1998). There it is further complicated by transitions between solutions with dipole and quadrupole symmetry (i.e. with B_ϕ symmetric or antisymmetric about the equator), which can be represented in an extended sixth-order system (Knobloch et al. 1998). In this low-order model the cycles are spatially asymmetric as the dynamo emerges from a grand minimum; the same effect appears in mean-field models (Beer et al. 1998)—and this is just what happened at the end of the Maunder Minimum in 1700 (Ribes and Nesme-Ribes 1993). In-out intermittency differs

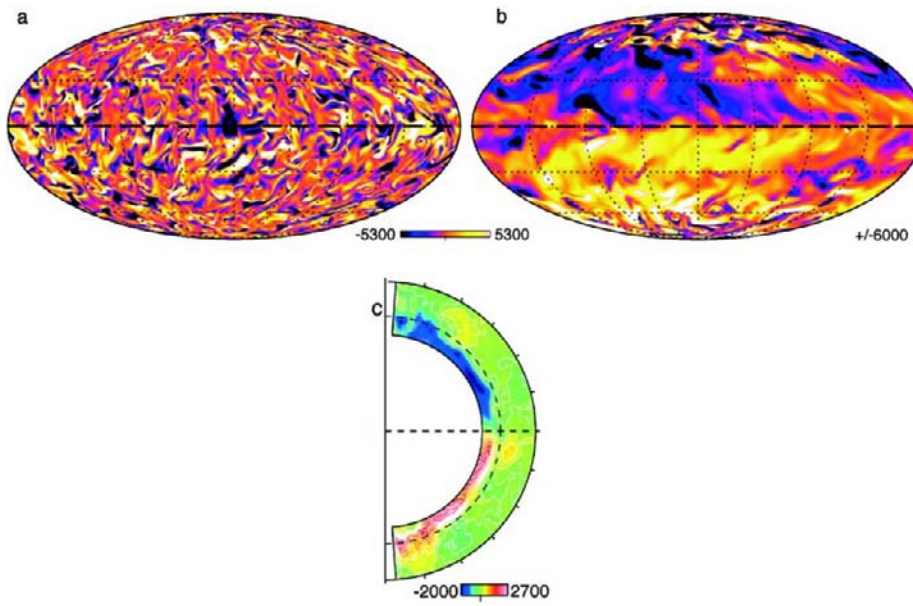


Fig. 6 The toroidal field in a recent computational model of the solar dynamo. Images of the azimuthal field (**a**) in the middle of the convection zone and (**b**) in the region of convective overshoot, both as Mollweide projections, with (**c**) the temporally averaged axisymmetric component of the toroidal field, which has approximate dipole symmetry. (From Browning et al. (2006))

from *on-off* intermittency (Tobias and Weiss 2007b); in the latter case the amplitude of the aperiodic magnetic cycle is controlled by an independent chaotic or stochastic oscillator that is unaffected by magnetic fields (Platt et al. 1993a, 1993b; Brandenburg and Spiegel 2008), as discussed by Spiegel elsewhere in these proceedings.[1]

For more detailed and realistic models one must turn to large-scale computation. This approach has successfully reproduced many key features of the geodynamo (as explained by Christensen et al. in these proceedings). Numerical studies of the solar dynamo were pioneered by Gilman (1983) a quarter-century ago. Although his Boussinesq models could not reproduce the butterflies in Fig. 1, they did establish the generic sensitivity of dynamos to the choice of parameters in the equations: small changes can switch solutions from steady to oscillatory behaviour, from poleward to equatorward travelling waves, or from dipole to quadrupole symmetry. The anelastic approximation was first included by Glatzmaier (1985) and since then anelastic models of the convection zone and of the solar dynamo have grown increasingly elaborate and sophisticated (e.g. Brun et al. 2004; Browning et al. 2006). So far, however, these numerical simulations have all been carried out at parameter values that are far from those that prevail in the solar interior. The current state of the art is illustrated in Fig. 6. Although this model does not exhibit oscillatory behaviour, there is a predominantly azimuthal field, with dipole symmetry, at the base of the convection zone. This is clearly the future: soon we may expect to see more developed models that actually reproduce the known behaviour of the solar cycle.

[1] Mathematically, the latter systems are said to have normal parameters and skew-product structure.

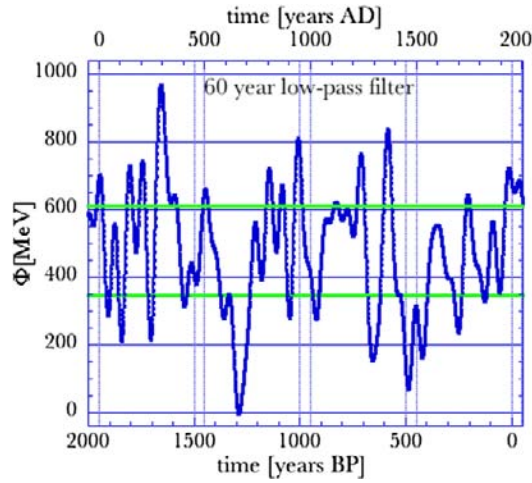


Fig. 7 A composite time series for the solar modulation potential Φ over the past 2000 years. Direct measurements of cosmic rays from neutron monitors (Usoskin et al. 2005) have been merged with ^{10}Be records from the South Pole (McCracken et al. 2004) and Greenland (Vonmoos et al. 2006) and the combined series has been detrended and smoothed (Abreu et al. 2008). *Horizontal lines* define levels of grand maxima and grand minima. The current grand maximum is clearly visible, as is the Maunder Minimum. (Courtesy of J.A. Abreu)

Given such a successful computational model, with both the velocity and the magnetic field known, we might attempt to relate it to mean-field dynamo theory by asking the following three questions. Can one compute a meaningful α from the statistics of the known velocity field **u**? Can one construct a mean-field dynamo model that mimics the results? Does the α-effect indeed capture the essential physics? We forecast that, although the answer to the first question will be *no*, the other two will be answered in the affirmative.

7 Predicting the Future

Solar activity is apparently chaotic, and thus it is intrinsically difficult to predict its future behaviour. Most efforts have been concerned with the immediate short-term problem of forecasting the next cycle, relying on precursor methods, or on reconstructing the attractor, or a combination of the two. These approaches discussed by Hathaway are in these proceedings.

The long-term problem involves attempts to forecast trends in the envelope of cyclic activity. For the past 80 years the Sun has been experiencing an episode of extreme activity, as shown by the ^{14}C and ^{10}Be proxy records (e.g. Solanki et al. 2004; Steinhilber et al. 2008). The GRIP ice-core provides a ^{10}Be record, already illustrated in Fig. 3, that extends back for more than 9 millennia, and Fig. 7 shows a smoothed composite record for the past 2000 yr. The current grand maximum is unusually high but it is certainly not unprecedented: Steinhilber et al. (2008) have identified 25 comparable events in the GRIP record. It is natural to ask how long this episode can be expected to last. Given its present lifetime of 80 yr, Abreu et al. (2008) estimate a total life expectancy of 95 yr, implying that this grand maximum will terminate within the next few cycles. There is then a possibility that it may be followed by a grand minimum as deep as the Maunder Minimum of the seventeenth century. If so, we may expect a detectable cooling effect on the Earth's climate, but one too small to compensate for global warming caused by anthropogenic greenhouse gases.

References

J.A. Abreu, J. Beer, F. Steinhilber, S.M. Tobias, N.O. Weiss, Geophys. Res. Lett. (2008, in press)

H.M. Antia, S.M. Chitre, M.J. Thompson, Astron. Astrophys. **360**, 335 (2000)
H.M. Antia, S. Basu, F. Hill, R. Howe, R.W. Komm, J. Schou, Mon. Not. R. Astron. Soc. **327**, 1029 (2001)
H.M. Antia, S.M. Chitre, M.J. Thompson, Astron. Astrophys. **399**, 329 (2003)
H.M. Antia, S. Basu, S.M. Chitre, Astrophys. J. **681**, 680 (2008)
P. Ashwin, E. Covas, R. Tavakol, Nonlinearity **12**, 562 (1999)
H.W. Babcock, Astrophys. J. **133**, 572 (1961)
C.S. Baldner, S. Basu, Astrophys. J. (2008, in press)
S. Basu, Mon. Not. R. Astron. Soc. **298**, 719 (1998)
S. Basu, H.M. Antia, Mon. Not. R. Astron. Soc. **269**, 1137 (1994)
S. Basu, H.M. Antia, Mon. Not. R. Astron. Soc. **324**, 498 (2001)
S. Basu, H.M. Antia, Astrophys. J. **585**, 553 (2003)
S. Basu, H.M. Antia, Phys. Rep. **457**, 217 (2007)
J. Beer, Space Sci. Rev. **94**, 53 (2000)
J. Beer, S.M. Tobias, N.O. Weiss, Sol. Phys. **181**, 237 (1998)
A. Brandenburg, Astrophys. J. **625**, 539 (2005)
A. Brandenburg, D. Schmitt, Astron. Astrophys. **338**, L55 (1998)
A. Brandenburg, E.A. Spiegel, Astron. Nachr. **329**, 351 (2008)
M.K. Browning, M.S. Miesch, A.S. Brun, J. Toomre, Astrophys. J. **648**, L157 (2006)
A.S. Brun, M.S. Miesch, J. Toomre, Astrophys. J. **613**, 1253 (2004)
P.J. Bushby, Astron. Nachr. **326**, 218 (2005)
P.J. Bushby, Mon. Not. R. Astron. Soc. **371**, 772 (2006)
F. Cattaneo, Astrophys. J. **515**, L39 (1999)
F. Cattaneo, D.W. Hughes, Phys. Rev. E **54**, 4532 (1996)
F. Cattaneo, D.W. Hughes, J. Fluid Mech. **553**, 401 (2006)
F. Cattaneo, D.W. Hughes, J. Fluid Mech. **594**, 495 (2008)
P. Charbonneau, Living Rev. Sol. Phys. **2**, 2 (2005). www.livingreviews.org/lrsp-2005-2
D.-Y. Chou, A. Serebryanskiy, Astrophys. J. **578**, L157 (2002)
A.R. Choudhuri, in *Dynamic Sun*, ed. by B.N. Dwivedi (Cambridge University Press, Cambridge, 2003), p. 103
J. Christensen-Dalsgaard, M.J. Thompson, in *The Solar Tachocline*, ed. by D.W. Hughes, R. Rosner, N.O. Weiss (Cambridge University Press, Cambridge, 2007), p. 53
J. Christensen-Dalsgaard, D.O. Gough, M.J. Thompson, Astrophys. J. **378**, 413 (1991)
J. Christensen-Dalsgaard, M.J.P.F.G. Monteiro, M.J. Thompson, Mon. Not. R. Astron. Soc. **276**, 283 (1995)
E. Covas, R. Tavakol, P. Ashwin, A. Tworkowski, J. Brooke, Chaos **11**, 404 (2001)
E. Covas, D. Moss, R. Tavakol, Astron. Astrophys. **416**, 775 (2004)
P.E. Damon, C.P. Sonett, in *The Sun in Time*, ed. by C.P. Sonett, M.S. Giampapa, M.S. Matthews (University of Arizona Press, Tucson, 1991), p. 360
P.H. Diamond, D.W. Hughes, E. Kim, in *Fluid Dynamics and Dynamos in Astrophysics and Geophysics*, ed. by A.M. Soward, C.A. Jones, D.W. Hughes, N.O. Weiss (CRC Press, Boca Raton, 2005), p. 145
M. Dikpati, P. Charbonneau, Astrophys. J. **518**, 508 (1999)
M. Dikpati, G. de Toma, P.A. Gilman, C.N. Arge, O.R. White, Astrophys. J. **601**, 1136 (2004)
S.B.F. Dorch, A. Nordlund, Astron. Astrophys. **365**, 562 (2001)
W.A. Dziembowski, P.R. Goode, A.G. Kosovichev, J. Schou, Astrophys. J. **537**, 1026 (2000)
J.A. Eddy, Science **192**, 1189 (1976)
A. Ferriz-Mas, D. Schmitt, M. Schüssler, Astron. Astrophys. **289**, 949 (1994)
P.A. Gilman, Astrophys. J. **53**, 243 (1983)
G.A. Glatzmaier, Astrophys. J. **291**, 300 (1985)
D.A. Haber, B.W. Hindman, J. Toomre, M.J. Thompson, Sol. Phys. **220**, 371 (2004)
D.A. Haber, B.W. Hindman, J. Toomre, R.S. Bogart, R.M. Larsen, F. Hill, Astrophys. J. **570**, 855 (2002)
R. Howe, J. Christensen-Dalsgaard, F. Hill, R.W. Komm, R.M. Larsen, J. Schou, M.J. Thompson, J. Toomre, Science **287**, 2456 (2000a)
R. Howe, J. Christensen-Dalsgaard, F. Hill, R.W. Komm, R.M. Larsen, J. Schou, M.J. Thompson, J. Toomre, Astrophys. J. **533**, L163 (2000b)
D.W. Hughes, in *Solar Tachocline*, ed. by D.W. Hughes, R. Rosner, N.O. Weiss (Cambridge University Press, Cambridge, 2007a), p. 275
D.W. Hughes, in *Mathematical Aspects of Natural Dynamos*, ed. by E. Dormy, A.M. Soward (CRC Press, Boca Raton, 2007b), p. 81
L. Jouve, A.S. Brun, R. Arlt, A. Brandenburg, M. Dikpati, A. Bonanno, et al. Astron. Astrophys. **483**, 949 (2008)
L.L. Kitchatinov, G. Rüdiger, M. Küker, Astron. Astrophys. **292**, 125 (1994)
E. Knobloch, S.M. Tobias, N.O. Weiss, Mon. Not. R. Astron. Soc. **297**, 1123 (1998)

A.G. Kosovichev, Astrophys. J. **469**, L61 (1996)
F. Krause, K.-H. Rädler, *Mean-Field Magnetohydrodynamics and Dynamo Theory* (Akademie, Berlin, 1980)
M. Küker, R. Arlt, G. Rüdiger, Astron. Astrophys. **343**, 977 (1999)
R.B. Leighton, Astrophys. J. **156**, 1 (1969)
K.G. McCracken, F.B. McDonald, J. Beer, G. Raisbeck, F. Yiou, J. Geophys. Res. **109**, A12103 (2004)
L. Mestel, *Stellar Magnetism* (Clarendon Press, Oxford, 1999)
U. Mitra-Kraev, M.J. Thompson, Astron. Nachr. **328**, 1009 (2007)
H.K. Moffatt, *Magnetic Field Generation in Electrically Conducting Fluids* (Cambridge University Press, Cambridge, 1978)
M.J.P.F.G. Monteiro, J. Christensen-Dalsgaard, M.J. Thompson, Astron. Astrophys. **283**, 247 (1994)
M.A.J.H. Ossendrijver, Astron. Astrophys. Rep. **11**, 287 (2003)
M. Ossendrijver, M. Stix, A. Brandenburg, Astron. Astrophys. **376**, 713 (2001)
E.N. Parker, *Cosmical Magnetic Fields: Their Origin and Their Activity* (Clarendon Press, Oxford, 1979)
E.N. Parker, Astrophys. J. **408**, 707 (1993)
N. Platt, E. Spiegel, C. Tresser, Phys. Rev. Lett. **70**, 279 (1993a)
N. Platt, E. Spiegel, C. Tresser, Geophys. Astrophys. Fluid Dyn. **73**, 146 (1993b)
V.V. Pipin, Astron. Astrophys. **346**, 295 (1999)
M. Rempel, Astrophys. J. **607**, 1046 (2004)
J.C. Ribes, E. Nesme-Ribes, Astron. Astrophys. **276**, 549 (1993)
B. Roberts, W.R. Campbell, Nature **323**, 603 (1986)
P.H. Roberts, in *Lectures on Solar and Planetary Dynamos*, ed. by M.R.E. Proctor, A.D. Gilbert (Cambridge University Press, Cambridge, 1994), p. 1
G. Rüdiger, R. Hollerbach, *The Magnetic Universe* (Wiley-VCH, Weinheim, 2004)
D. Schmitt, M. Schüssler, A. Ferriz-Mas, Astron. Astrophys. **311**, L1 (1996)
S.K. Solanki, B. Inhester, M. Schüssler, Rep. Prog. Phys. **69**, 563 (2006)
S.K. Solanki, I.G. Usoskin, M. Kromer, M. Schüssler, J. Beer, Nature **431**, 1084 (2004)
F. Steinhilber, J.A. Abreu, J. Beer, Atmos. Space Sci. Trans. **4**, 1 (2008)
M. Stuiver, T.F. Braziunas, Holocene **3**, 289 (1993)
J.-C. Thelen, Mon. Not. R. Astron. Soc. **315**, 155 (2000a)
J.-C. Thelen, Mon. Not. R. Astron. Soc. **315**, 165 (2000b)
J.H. Thomas, N.O. Weiss, *Sunspots and Starspots* (Cambridge University Press, Cambridge, 2008)
M.J. Thompson, J. Christensen-Dalsgaard, M.S. Miesch, J. Toomre, Annu. Rev. Astron. Astrophys. **41**, 599 (2003)
S.M. Tobias, Astron. Astrophys. **307**, L21 (1996)
S.M. Tobias, Astron. Astrophys. **322**, 1007 (1997)
S.M. Tobias, Philos. Trans. R. Soc. Lond. **360**, 2741 (2002)
S.M. Tobias, N.O. Weiss, in *The Solar Tachocline*, ed. by D.W. Hughes, R. Rosner, N.O. Weiss (Cambridge University Press, Cambridge, 2007a), p. 319
S.M. Tobias, N.O. Weiss, in *Mathematical Aspects of Natural Dynamos*, ed. by E. Dormy, A.M. Soward (CRC Press, Boca Raton, 2007b), p. 281
S.M. Tobias, N.O. Weiss, V. Kirk, Mon. Not. R. Astron. Soc. **273**, 1150 (1995)
S.M. Tobias, N.H. Brummell, T.L. Clune, J. Toomre, Astrophys. J. **549**, 1183 (2001)
I.G. Usoskin, K. Alanko-Huotari, G.A. Kovaltsov, K. Mursula, J. Geophys. Res. **110**, A12108 (2005)
S.I. Vainshtein, F. Cattaneo, Astrophys. J. **393**, 165 (1992)
M. Vonmoos, J. Beer, R. Muscheler, J. Geophys. Res. **111**, A10105 (2006)
S.V. Vorontsov, J. Christensen-Dalsgaard, J. Schou, V.N. Strakhov, M.J. Thompson, Science **296**, 101 (2002)
A. Vögler, M. Schüssler, Astron. Astrophys. **465**, L43 (2007)
G. Wagner, J. Beer, J. Masarik, P.W. Kubik, W. Mende, C. Laj, G.M. Raisbeck, F. Yiou, Geophys. Res. Lett. **28**, 303 (2001)
J. Zhao, A.G. Kosovichev, Astrophys. J. **591**, 446 (2003)

Flux-Transport Solar Dynamos

Mausumi Dikpati · Peter A. Gilman

Originally published in the journal Space Science Reviews, Volume 144, Nos 1–4, 67–75.
DOI: 10.1007/s11214-008-9484-3

Abstract Large-scale solar dynamo models were first built by Parker (1955). Over the past half a century these models have evolved significantly. We discuss here the development of a class of large-scale dynamo models which include, along with the α-effect and Ω-effect, an important third process, flux transport by meridional circulation. We present the properties of this 'flux-transport' dynamo, including the crucial role meridional circulation plays in giving this dynamo predictive power.

Keywords Solar activity · Dynamo · Meridional circulation

1 What Is a Flux-Transport Dynamo

Flux-transport dynamos are just the so-called α-Ω dynamos with meridional circulation. Flux-transport dynamos include three basic processes: (i) shearing of the poloidal magnetic fields to produce toroidal fields by the Sun's differential rotation (the Ω-effect), (ii) regeneration of poloidal fields by displacing and twisting the toroidal flux tubes by helical motions (the so-called α-effect), and (iii) advective transport of magnetic flux by meridional circulation, whereas an α-Ω dynamo involves only the first two. Meridional circulation acts as a conveyor belt in this class of models. In these models, this ingredient also plays an important role in determining the dynamo cycle period and in governing the memory of the Sun's past magnetic fields.

This work is partially supported by the NASA LWS grant NNH05AB521.

M. Dikpati (✉) · P.A. Gilman
HAO/NCAR, Boulder, CO 80301, USA
e-mail: dikpati@hao.ucar.edu

P.A. Gilman
e-mail: gilman@hao.ucar.edu

M.J. Thompson et al. (eds.), *The Origin and Dynamics of Solar Magnetism*.
DOI: 10.1007/978-1-4419-0239-9_5

2 A Brief History of Development of Flux-Transport Dynamos

Historically recognition of the need for the advective transport of magnetic flux by meridional circulation came after the observation of the poleward drift of the Sun's large-scale fields, primarily poloidal fields that eventually cause the polar reversal. Magnetogram studies (Babcock and Babcock 1955) revealed that weak, diffuse fields, arising from the decay of active regions, drift poleward, in contrast to the equatorward migration of the spot-zones, with the progress of the solar cycle. In order to account quantitatively for the drift-rate of the large-scale surface poloidal flux towards the poles as well as polar reversal timings, advective transport of this flux by meridional circulation was shown to be necessary (Devore et al. 1984; Wang et al. 1989) in addition to supergranular diffusion (see also Schrijver et al. 2002).

From these simulations that explain the observed evolution of the Sun's large-scale surface poloidal fields, the meridional circulation was soon noted as an important ingredient that needs to be included in solar dynamo processes. Thus, considering a Babcock-Leighton-type surface poloidal field source (which arises from the decay of tilted, bipolar active regions, Leighton 1969), a radial shear at the base of the convection zone and a meridional circulation with a poleward surface flow and an equtorward subsurface flow at the convection zone base, Wang and Sheeley (1991) first built, using a simplified rectangular geometry, a Babcock-Leighton flux-transport dynamo model which gave a new twist to the existing dynamo models.

During the past two decades many flux-transport dynamo models have been built (Choudhuri et al. 1995; Durney 1995; Dikpati and Charbonneau 1999; Küker et al. 2001; Bonanno et al. 2002; Guererro and Munoz 2004; Rempel 2006; Jouve and Brun 2007). These models show that the meridional circulation works as a conveyor belt by transporting flux from the surface to the bottom of the convection zone where the new spot-producing flux is generated. Figure 1 shows a sequence of schematic diagrams that depict qualitatively the succession of processes contained in a flux-transport dynamo solution.

3 Existence of Meridional Circulation

Helioseismic observations reveal details about the Sun's flow fields, in particular that the Sun's azimuthal flow primarily exhibits latitudinal differential rotation in the bulk of the convection zone, while radial differential rotation exists in a thin layer at the base of the solar convection zone (Brown et al. 1989; Dziembowski et al. 1989; Tomczyk et al. 1995; Corbard et al. 1998). The meridional circulation has been detected by various observations, such as Doppler measurements (Duvall 1979; Ulrich et al. 1988; Cavallini et al. 1992; Hathaway et al. 1996), magnetic tracers (Komm et al. 1993), and helioseismic analysis (Giles et al. 1997; Braun and Fan 1998)—all have found a poleward flow in the near-surface layers of 10–20 $\mathrm{m\,s^{-1}}$. Helioseismic inversions have also indicated that the meridional flow remains poleward in the upper half of the convection zone down to about 0.85 $R_\odot$. The equatorward return flow hasn't yet been observed, but it must exist, since mass does not pile up at the solar poles. For flux-transport dynamo simulations in the $r-\theta$ meridional cut, the equatorward return flow is generally constructed by incorporating the observed flow-pattern in the outer envelope of the convection zone and then applying the constraint of mass conservation. The appearance and disappearance of a high-latitude reverse cell from time to time has also been reported (Haber et al. 2002; Basu and Antia 2003; Zhao and Kosovichev 2004; Ulrich and Boyden 2005).

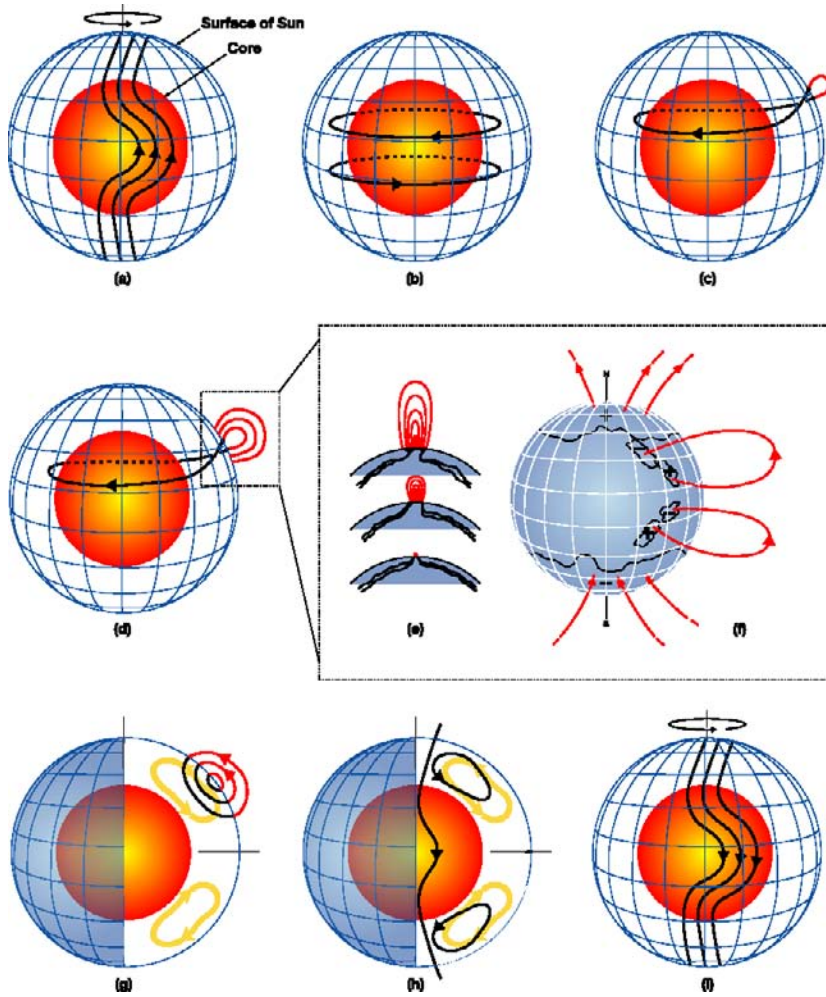


Fig. 1 Schematic of solar flux-transport dynamo processes. *Red inner sphere* represents the Sun's radiative core and *blue mesh* the solar surface. In between is the solar convection zone where dynamo resides. (**a**) Shearing of poloidal field by the Sun's differential rotation near convection zone bottom. The Sun rotates faster at the equator than the pole. (**b**) Toroidal field produced due to this shearing by differential rotation. (**c**) When toroidal field is strong enough, buoyant loops rise to the surface, twisting as they rise due to rotational influence. Sunspots (*two black dots*) are formed from these loops. (**d**,**e**,**f**) Additional flux emerges (**d**,**e**) and spreads (**f**) in latitude and longitude from decaying spots (as described in Fig. 5 of Babcock 1961). (**g**) Meridional flow (*yellow circulation with arrows*) carries surface magnetic flux poleward, causing polar fields to reverse. (**h**) Some of this flux is then transported downward to the bottom and towards the equator. These poloidal fields have sign opposite to those at the beginning of the sequence, in frame (**a**). (**i**) This reversed poloidal flux is then sheared again near the bottom by the differential rotation to produce the new toroidal field opposite in sign to that shown in (**b**)

The theory for the Sun's meridional circulation is still under development (Rempel 2005; Miesch et al. 2008). Meridional flow is the second-order effect of convection in a rotating sphere. The detailed physics behind this theory is quite complicated. In brief, Coriolis force acting on convection creates Reynolds stresses that transport angular momentum towards

the equator. Thus equatorial acceleration is produced. Meridional circulation is then driven by the Coriolis force from the differential rotation—outward radial velocity created from large rotational flow at low latitudes make the fluid particles flow upward from the base of the convection zone and then poleward at the surface.

4 Flux-Transport Dynamo Solutions

In the kinematic regime, flux-transport dynamos include the following major ingredients: (i) differential rotation, (ii) meridional circulation, (iii) Babcock-Leighton poloidal source and (iv) magnetic diffusivity. Although the poloidal field generation in a flux-transport dynamo occurs primarily due to the Babcock-Leighton type source, the presence of some α-effect at the core-envelope interface or at the tachocline helps the parity selection (Dikpati and Gilman 2001; Bonanno et al. 2002). The governing equations can be obtained from the induction equation by using the mean-field formalism (Steenbeck et al. 1966):

$$\frac{\partial A}{\partial t} + \frac{1}{r\sin\theta}(\mathbf{u}\cdot\nabla)(r\sin\theta A) = \eta\left(\nabla^2 - \frac{1}{r^2\sin^2\theta}\right)A + S(r,\theta,B_\phi) + \alpha B_\phi, \qquad (1)$$

$$\frac{\partial B_\phi}{\partial t} + \frac{1}{r}\left[\frac{\partial}{\partial r}(ru_r B_\phi) + \frac{\partial}{\partial\theta}(u_\theta B_\phi)\right]$$

$$= r\sin\theta(\mathbf{B}_p\cdot\nabla)\Omega - \hat{\mathbf{e}}_\phi\cdot\left[\nabla\eta\times\nabla\times B_\phi\hat{\mathbf{e}}_\phi\right] + \eta\left(\nabla^2 - \frac{1}{r^2\sin^2\theta}\right)B_\phi. \qquad (2)$$

Here $B_\phi(r,\theta,t)\,\hat{\mathbf{e}}_\phi$ is the toroidal field, $\nabla\times A(r,\theta,t)\,\hat{\mathbf{e}}_\phi$, the poloidal field, $S(r,\theta,B_\phi)$, the Babcock-Leighton poloidal field source, α, the tachocline α-effect, $\Omega(r,\theta,t)$, the differential rotation, $\mathbf{u} = u_r\,\hat{\mathbf{e}}_r + u_\theta\,\hat{\mathbf{e}}_\theta$, the meridional circulation, and η the depth-dependent magnetic diffusivity. These equations are solved numerically with suitable boundary conditions which are straightforward. B_ϕ is zero on all four boundaries of the pole-to-pole meridional cut extending from $0.6R$ to $1R$ in the radial direction, whereas A is zero in all boundaries except at the upper boundary at which a smooth matching between the interior and exterior poloidal fields is demanded. Initialization can be done with random fields.

The differential rotation does not change much with time. The meridional flow pattern is not known at the lower half of the convection zone. But it has been widely shown that flux-transport type dynamos are not very sensitive to the variations in streamlines in terms of how dense or rare they are near the equator or pole, near the surface or base of the convection zone, as long as the flow is a singe cell pattern in each hemisphere. The model becomes sensitive when the flow becomes multi-cell pattern (Bonanno et al. 2005; Jouve and Brun 2007). The least known ingredient is the magnetic diffusivity profile. Therefore by selecting a single cell flow and tuning the diffusivity profile this dynamo can be calibrated.

Flux-transport dynamos have been successful in reproducing many large-scale solar cycle features including the correct phase relationship between the equatorward-migrating sunspot belts and the poleward-drifting large-scale poloidal fields, a difficult feature to be reproduced using dynamos without meridional circulation. A typical 'butterfly diagram' derived from the model-out put of a calibrated flux-transport dynamo (Dikpati et al. 2004; Dikpati and Gilman 2007) is shown in Fig. 2 and is compared with the observed 'butterfly diagram'.

The α-Ω convection zone dynamos, thin-layer dynamos and interface dynamos can explain the equatorward migration of sunspot zones through the propagating dynamo-wave solution, but the poloidal fields, which are the vector counterparts of those spot-producing

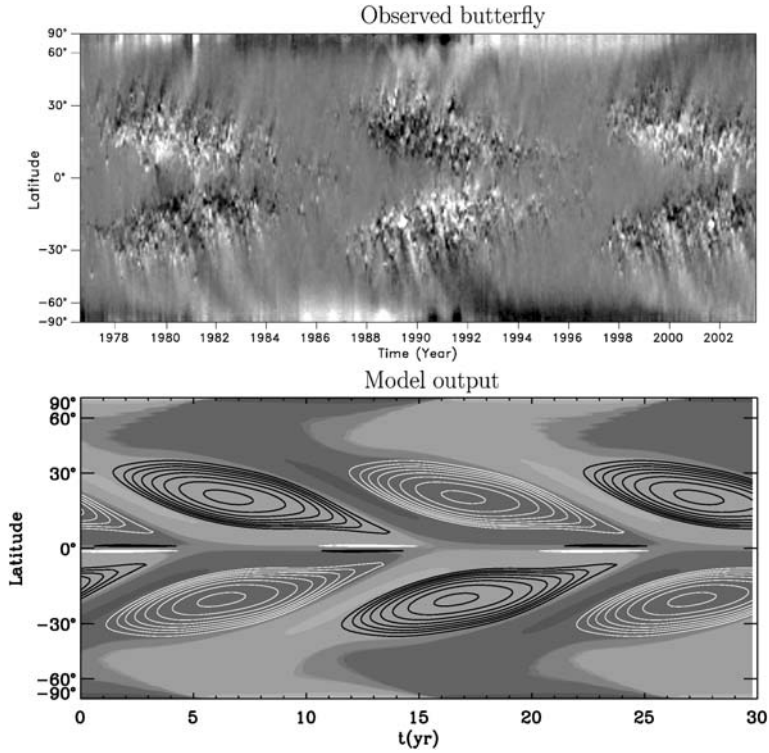


Fig. 2 *Top frame*: NSO synoptic map of observed longitude-averaged photospheric magnetic fields. Equatorward propagating branch represents the evolution of bipolar spots as a function of solar cycle; poleward branch represents polar field evolution. *Bottom frame*: Gray-scale map of $B_\phi|_{r=0.725R}$ (primarily confined below 35° latitude) is superimposed on that of surface radial field in time–latitude plane. *Bright shades* represent positive fields, which means that in low latitudes, toroidal fields run from positive to negative along the direction of rotation and hence follower spots will erupt with positive polarity; in mid-latitudes and high latitudes *bright shades* denote radial fields that are radially outward. *Dark shades* in low and high latitudes respectively represent toroidal field lines opposite to the direction of rotation and inward radial fields. Innermost shade in low-latitude has a value of 100 kG and 3 contours cover an order of magnitude field strength. Radial fields are 3–4 orders of magnitude weaker. Adopted from Figs. 8 and 9 of Dikpati et al. (2004)

toroidal fields, also migrate equatorward instead of poleward, in these models. Without meridional circulation, it is difficult to explain the contrasting evolution of the Sun's two vector components of the global magnetic fields with correct phase relationship. Along with the revival of Babcock-Leighton dynamos with meridional circulation, a parallel attempt was made—a polar-branch dynamo was built (Gilman et al. 1989), which can explain the poleward drift without meridional circulation. However the details remain unexplored yet.

5 Some Unique Properties of a Flux-Transport Dynamo

The dynamo cycle period and the Sun's magnetic memory are primarily governed by the meridional flow speed in this class of models. For a typical solar-like meridional circulation pattern, containing a single flow-cell with maximum surface flow-speed of 15 $\mathrm{m\,s^{-1}}$, the flow-particles are plotted in blue dots in Fig. 3 along a selected streamline in one year intervals. This plot reveals: (i) the latitude-zone of spot-producing fields drift from mid-latitudes

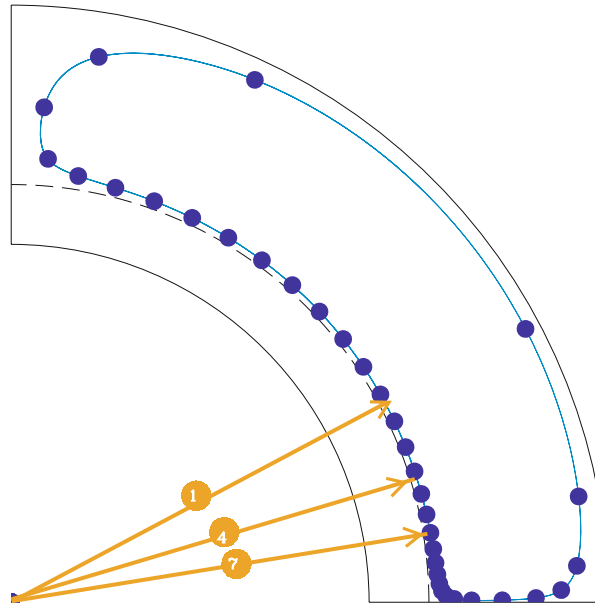


Fig. 3 The *blue flow-dots* are plotted on a streamline in one-year intervals. The number of *dots* from mid-latitudes at the surface to the mid-latitudes at the base of the convection zone indicates the Sun's magnetic memory length in years, and that from mid-latitudes to the equator indicates how this circulation of $\sim$ 15–20 years

to equator in $\sim$ 11 years, indicating how the average length of a sunspot cycle becomes 11 year, (ii) the surface poloidal fields are transported from the high-latitudes at the surface to the mid-latitudes at the bottom in $\sim$ 15–20 years, indicating how the meridional circulation plays a crucial role in determining the Sun's memory about its past magnetic fields. The variations in the meridional flow-speed should cause the variations in the cycle-length from cycle to cycle.

The property of magnetic memory rendered by the meridional circulation provides the predictive power to this class of models (Dikpati 2004; Dikpati et al. 2006). The solar meridional circulation is not a unique circulation in nature. Analogous latitudinal circulations exist in the terrestrial system, such as Hadley cell, polar cell, Ferrel cell in the tropopause, which are known to have influence in weather forecasting. The great ocean conveyor belt, which is a thermohaline circulation, is driven primarily by the formation and sinking of cold water in the Norwegian Sea. This circulation is thought to be responsible for the large flow of upper ocean water from the tropical Pacific to the Indian Ocean through the Indonesian Archipelago. The two counteracting forcings operating in the North Atlantic control the conveyor belt circulation: (1) the thermal forcing (high-latitude cooling and low-latitude heating) which drives a polar southward flow; and (2) haline forcing (net high-latitude freshwater gain and low-latitude evaporation) which moves in the opposite direction.

The common features between the Sun's meridional flow conveyor belt and that of the ocean are: both are persistent slow flow in the turbulent medium and both carry the surface forcing (surface magnetic flux and thermal patterns respectively in the case of the Sun and the ocean) with long memory. Warm North Atlantic meridional overturning circulation is linked to excessive rain-fall over Sahel and western India, while variations in overturning flow in eastern Pacific determine the timing, amplitude and duration of an El Nino event. Similarly, the variations in the meridional flow amplitude and profile determine the timing and shape of a cycle, and amplitude to some extent (details are discussed in this book in the solar cycle prediction chapter by David Hathaway).

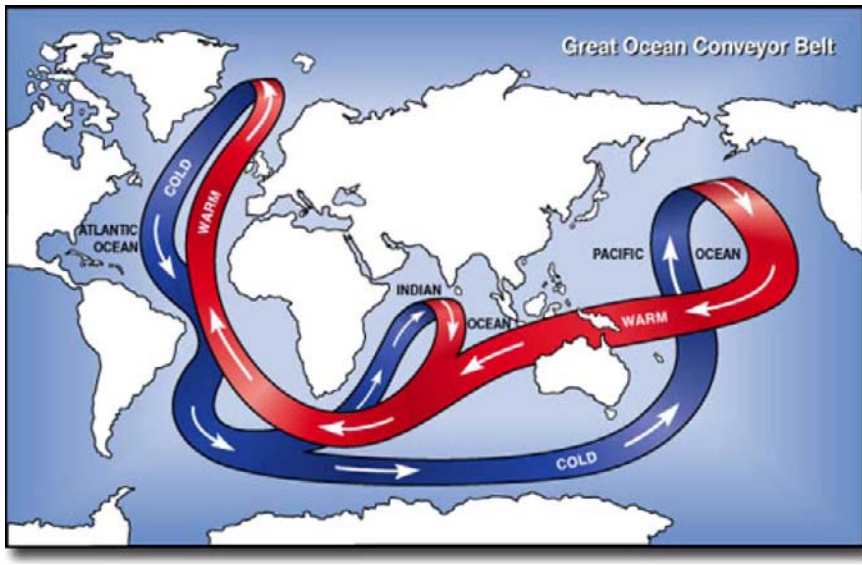


Fig. 4 The great ocean conveyor belt is created due to thermo-haline forcing

6 Discussion and Future Prospects

To be calibrated, a solar dynamo model must be sufficiently realistic to allow inclusion of ingredients known from solar observations. For example, a spherical shell configuration is necessary in order to incorporate the observed differential rotation and meridional circulation. In order to incorporate the observed Babcock-Leighton type poloidal source in the most realistic way, one needs to consider departures from axisymmetry, because this poloidal source arises from the decay of bipolar active regions which are tilted in latitude and longitude. For simplicity, averaging this poloidal source in longitude can be a good starting point, and the amplitude and time-dependence of this source can be derived from observations. The magnetic diffusivity is the least known ingredient, and we have to rely on theoretical arguments.

Kinematic dynamo models that solve only certain forms of the induction equation can be calibrated to many solar cycle features, but adding an equation of motion would allow calibration of the model with the so-called torsional oscillations, which are clearly associated with the solar cycle. Nevertheless, kinematic flux-transport dynamos should work well, since the modulation of differential rotation with cycle is small (Vorontsov et al. 2002), and hence the global effect of $\mathbf{j}\times\mathbf{B}$ forces is limited (see Rempel 2006).

There are practical limits to the inclusion of further realism. Dynamo action in the Sun no doubt occurs on many space and timescales, from the global down to granulation scales (10^{-4} of the solar radius), involving many turbulent processes. Capturing all or even most of these scales and processes in one numerical model is not possible yet. Current 3D global MHD models for solar differential rotation, convection and magnetic fields are truncated at larger spatial scales than supergranulation (10^{-2} of the solar radius), so they must parameterize all smaller-scale turbulent processes. It is therefore not possible to do direct numerical simulations (DNS) for the solar convection zone. Such models show dynamo action (Browning et al. 2006); however the simulation for global reversal, and hence the calibration, are still under progress. The earliest such simulation attempts (Gilman 1983), done with much

coarser resolution, did show field reversals but the resulting butterfly diagrams were not solar-like. Toroidal fields produced in them migrated toward the poles rather than toward the equator.

As a consequence of the above considerations, the large-scale flux-transport dynamos are being formulated in the way described above, although there is no doubt about the need for full 3D MHD models. In both the kinematic and non-kinematic regimes, axisymmetric models are good for explaining certain longitude-averaged solar cycle features. But there are some important longitude-dependent solar cycle features, such as 'active longitudes'. Hence it is necessary to include the third dimension. One way of building 3D models is mentioned above (Browning et al. 2006). The other possible way toward 3D modelling is to generalize flux-transport dynamos by including large-scale longitudinal dependence. This latter approach has some advantage regarding speed and simplicity over full 3D MHD simulations. Although the ultimate goal is to build full 3D MHD dynamos, the 3D generalizations of flux-transport dynamos can still capture important physics of some global effects, and can provide some guidance to full 3D MHD models.

Acknowledgements We thank Keith MacGregor for his thorough review of the entire manuscript. This work is partially supported by NASA grant NNH05AB521. We also acknowledge the support from ISSI, Bern for participating in the solar magnetism workshop, during which interactions with worldwide dynamo theorists were very helpful for the preparation of this paper.

References

H.W. Babcock, The topology of the Sun's magnetic field and the 22-year cycle. Astrophys. J. **133**, 572 (1961)
H.W. Babcock, H.D. Babcock, The Sun's magnetic field, 1952–1954. Astrophys. J. **121**, 349 (1955)
S. Basu, H.M. Antia, Changes in solar dynamics from 1995 to 2002. Astrophys. J. **585**, 553 (2003)
A. Bonanno, D. Elstner, G. Rüdiger, G. Belvedere, Parity properties of an advection dominated solar alpha $^2\Omega$-dynamo. Astron. Astrophys. **390**, 673–680 (2002)
A. Bonanno, D. Elstner, G. Belvedere, G. Rudiger, A flux-transport dynamo with a multi-cell meridional circulation. Astron. Nachr. **326**, 170 (2005)
D.C. Braun, Y. Fan, Helioseismic measurements of the subsurface meridional flow. Astrophys. J. Lett. **508**, L105 (1998)
T.M. Brown, J. Christensen-Dalsgaard, W.A. Dziembowski, P.R. Goode, D.O. Gough, C.A. Morrow, Inferring the sun's internal angular velocity from observed p-mode frequency splittings. Astrophys. J. **343**, 526 (1989)
M. Browning, M.S. Miesch, A.S. Brun, J. Toomre, Dynamo action in the solar convection zone and tachocline: pumping and organization of toroidal field. Astrophys. J. **648**, L157 (2006)
F. Cavallini, G. Ceppatelli, A. Righini, About spectroscopic measurements of the solar meridional motion. Astron. Astrophys. **254**, 381 (1992)
A.R. Choudhuri, M. Schussler, M. Dikpati, The solar dynamo with meridional circulation. Astron. Astrophys. **303**, L29 (1995)
T. Corbard, G. Berthomieu, J. Provost, P. Morel, Inferring the equatorial solar tachocline from frequency splittings. Astron. Astrophys. **330**, 1149 (1998)
C.R. Devore, J.P. Boris, N.R. Sheeley Jr., The concentration of the large-scale solar magnetic field by a meridional surface flow. Sol. Phys. **92**, 1 (1984)
M. Dikpati, Global MHD theory of tachocline and the current status of large-scale solar dynamo. Proc. SOHO 14 / GONG 2004 Workshop ESA SP-559 (2004), p. 233
M. Dikpati, P. Charbonneau, A Babcock-Leighton flux transport dynamo with solar-like differential rotation. Astrophys. J. **518**, 508 (1999)
M. Dikpati, P.A. Gilman, Flux-transport dynamos with alpha-effect from global instability of tachocline differential rotation: A solution for magnetic parity selection in the Sun. Astrophys. J. **559**, 428 (2001)
M. Dikpati, G. de Toma, P.A. Gilman, C.N. Arge, O.R. White, Diagnostics of polar field reversal in solar cycle 23 using a flux transport dynamo model. Astrophys. J. **601**, 1136 (2004)
M. Dikpati, G. de Toma, P.A. Gilman, Predicting the strength of solar cycle 24 using a flux-transport dynamo-based tool. Geophys. Res. Lett. **33**, L05102 (2006)

M. Dikpati, P.A. Gilman, Sol. Phys. **241**, 1 (2007)
B.R. Durney, On a Babcock-Leighton dynamo model with a deep-seated generating layer for the toroidal magnetic field. Sol. Phys. **63**, 3 (1995)
T.L. Duvall Jr., Large-scale solar velocity fields. Sol. Phys. **63**, 3 (1979)
W.A. Dziembowski, P.R. Goode, K.G. Libbrecht, The radial gradient in the Sun's rotation. Astrophys. J. **337**, L53 (1989)
P.M. Giles, T.L. Duvall Jr., P.H. Scherrer, R.S. Bogart, A flow of material from the suns equator to its poles. Nature **390**, 52 (1997)
P.A. Gilman, Dynamically consistent nonlinear dynamos driven by convection in a rotating spherical shell. II – Dynamos with cycles and strong feedbacks. Astrophys. J. Suppl. **53**, 243 (1983)
P.A. Gilman, C.A. Morrow, E.E. Deluca, Angular momentum transport and dynamo action in the Sun – Implications of recent oscillation measurements. Astrophys. J. **338**, 528–537 (1989)
G.A. Guererro, J.D. Munoz, Kinematic solar dynamo models with a deep meridional flow. Mon. Not. R. Astron. Soc. **350**, 317 (2004)
D.A. Haber, B.W. Hindman, J. Toomre, R.S. Bogart, R.M. Larsen, F. Hill, Evolving submerged meridional circulation cells within the upper convection zone revealed by ring-diagram analysis. Astrophys. J. **570**, 855 (2002)
D. Hathaway, P. Gilman, J. Harvey, F. Hill, R. Howard, H. Jones, J. Kasher, J. Leibacher, J. Pintar, G. Simon, GONG observations of solar surface flows. Science **272**, 1306 (1996)
L. Jouve, A.S. Brun, On the role of meridional flows in flux transport dynamo models. Astron. Astrophys. **474**, 239–250 (2007)
R.W. Komm, R.F. Howard, J.W. Harvey, Meridional flow of small photospheric magnetic features. Sol. Phys. **147**, 207–233 (1993)
M. Küker, G. Rüdiger, M. Schultz, Circulation-dominated solar shell dynamo models with positive alpha-effect. Astron. Astrophys. **374**, 301 (2001)
R.B. Leighton, A magneto-kinematic model of the solar cycle. Astrophys. J. **166**, 1 (1969)
M.S. Miesch, A.S. Brun, M.L. de Rosa, J. Toomre, Structure and evolution of giant cells in global models of solar convection. Astrophys. J. **673**, 557 (2008)
M. Rempel, Influence of random fluctuations in the Λ-effect on meridional flow and differential rotation. Astrophys. J. **622**, 1320 (2005)
M. Rempel, Transport of toroidal magnetic field by the meridional flow at the base of the solar convection zone. Astrophys. J. **647**, 662 (2006)
C.J. Schrijver, M.L. de Rosa, A.M. Title, What is missing from our understanding of long-term solar and heliospheric activity? Astrophys. J. **577**, 1006 (2002)
M. Steenbeck, F. Krause, K.-H. Rädler, A calculation of the mean electromotive force in an electrically conducting fluid in turbulent motion under the influence of coriolis forces. Z. Naturforsch. **21**, 369 (1966)
S. Tomczyk, J. Schou, M.J. Thompson, Measurement of the rotation rate in the deep solar interior. Astrophys. J. **448**, L57 (1995)
R.K. Ulrich, J.E. Boyden, The solar surface toroidal magnetic field. Astrophys. J. **620**, L123 (2005)
R.K. Ulrich, J.E. Boyden, L. Webster, S.P. Padilla, H.B. Snodgrass, Solar rotation measurements at Mount Wilson. V – Reanalysis of 21 years of data. Sol. Phys. **117**, 291 (1988)
S. Vorontsov, J. Christensen-Dalsgaard, J. Schou, V.N. Strakhov, M.J. Thompson, Helioseismic measurement of solar torsional oscillations. Science **296**, 101 (2002)
Y.-M. Wang, N.R. Sheeley Jr., Magnetic flux transport and the sun's dipole moment – New twists to the Babcock-Leighton model. Astrophys. J. **375**, 761 (1991)
Y.-M. Wang, A.G. Nash, N.R. Sheeley, Magnetic flux transport on the Sun. Science **245**, 712 (1989)
Y.-M. Wang, A.G. Nash, N.R. Sheeley, A new solar cycle model including meridional circulation. Astrophys. J. **383**, 431 (1991)
J. Zhao, A.G. Kosovichev, Torsional oscillation, meridional flows, and vorticity inferred in the upper convection zone of the Sun by time–distance helioseismology. Astrophys. J. **603**, 776 (2004)

The Solar Dynamo: The Role of Penetration, Rotation and Shear on Convective Dynamos

Steven M. Tobias

Originally published in the journal Space Science Reviews, Volume 144, Nos 1–4, 77–86.
DOI: 10.1007/s11214-008-9442-0

Abstract In this paper I discuss the importance of turbulence, rotation, penetration and shear for solar dynamos (both local and global). An understanding of these processes is vital for progress towards a self-consistent theory for the generation of solar magnetic activity. I discuss the difficulties for large-scale field generation and suggest that large-scale solar magnetic activity may be driven by dynamos that arise owing to instabilities, with these dynamos modified by the presence of turbulence.

Keywords Solar dynamo · Sun · Magnetic fields · Magnetic activity

1 Introduction: Models of Solar Cycle

The mechanism for the generation of magnetic fields in the Sun remains a subject of contentious debate. The solar magnetic field has dynamics on a vast range of spatial and temporal scales and is responsible for important phenomena such as sunspots, flares and coronal mass ejections and heats the corona to such high temperatures.

Although there remains much to understand about the origin of magnetic fields in the Sun, a consensus has developed that these are generated by dynamo action. In a dynamo motion of the electrically conducting plasma acts to produce field against the dissipative effects of ohmic diffusion. Owing to the vast range of spatial scales exhibited by solar magnetic fields, turbulent dynamo theory has traditionally been separated into two strands, small-scale dynamo theory (sometimes termed fluctuation dynamo theory) in which the field is generated on scales smaller (or of the same size) than those of the turbulent eddies, and large-scale dynamo theory, which is concerned with the systematic generation of fields at a scale larger than that of the turbulence. In the solar context, the systematic large-scale field which leads to the solar cycle visible in sunspots and active regions is believed to owe its origin to a large-scale dynamo, whilst the magnetic carpet, which is generated on smaller

S.M. Tobias (✉)
Department of Applied Mathematics, University of Leeds, Leeds, LS2 9JT, UK
e-mail: smt@maths.leeds.ac.uk

M.J. Thompson et al. (eds.), *The Origin and Dynamics of Solar Magnetism*.
DOI: 10.1007/978-1-4419-0239-9_6

scales is the product either of a small-scale dynamo or the reprocessing of large-scale magnetic flux by the turbulent solar convection. In this article I shall not review the observations of solar magnetic fields or go into detail of the various scenarios that have been postulated for their maintenance, since these issues are dealt with in other contributions to this volume (see e.g. Weiss and Thompson 2008) and in recent reviews (see e.g. Ossendrijver 2003; Tobias and Weiss 2007)

It is fair to say that much more is understood about the dynamics of small-scale dynamos than large-scale ones. Dynamo theory has demonstrated that virtually any turbulent flow is capable of generating small-scale magnetic field if the magnetic Reynolds number *Rm*—the non-dimensional measure of the rate of stretching to diffusion—is large enough. This issue is not completely settled, however, with two outstanding questions remaining. There is still some doubt as to whether even a small-scale dynamo can successfully generate magnetic field when the fluid Reynolds number *Re* is much larger than *Rm*, the so-called small magnetic Prandtl number limit—the appropriate limit for the solar interior. Although this issue is not settled (and is exceedingly difficult to settle via numerical computation, see e.g. Iskakov et al. 2007) the indications are that these small-scale dynamos can survive efficiently in this limit (Boldyrev and Cattaneo 2004; Tobias and Cattaneo 2008a). There has also been some discussion about whether strongly stratified turbulence, such as exists at the solar surface can still be efficient as a dynamo (see e.g. Stein and Nordlund 2002) and I shall return to this question below.

Much more contentious is the issue of the origin of the large-scale field that is responsible for the solar cycle. Here no consensus has been reached over which ingredients are important for the generation of large-scale fields and even which locations in the solar interior might be responsible for generating the flux responsible for the solar cycle. Again, a complete review of the proposed mechanisms for large-scale field generation is beyond the scope of this contribution, but it is important to note the many and varied possibilities that have been proposed.

The most conventional proposed dynamo scenario is that the large-scale field is generated by a distributed dynamo located in the convection zone. Here cyclonic turbulence generates poloidal field throughout the convection zone whilst the shear (either latitudinal in the convection zone or a combination of radial and latitudinal in the tachocline) regenerates the poloidal field. This classical picture has been modelled for a number of years via mean-field electrodynamics and more recently has been refined to include the impact of the near-surface shear revealed by helioseismology (Brandenburg 2005). An alternative (either deep-seated or interface) paradigm invokes the base of the solar convection zone as the preferred location for the dynamo. Here the toroidal field is generated via the strong shear in the tachocline whilst the dynamo loop is completed by the regeneration of poloidal field either by large scale convection in the lower convection zone, or via the interaction of the magnetic buoyancy instability with rotation (see e.g. Thelen 2000). In this scenario, the dynamo region is hidden from observations and the active regions that are formed do not play a key role in the regeneration process, but exist merely as a by-product of the dynamo. In a third scenario, the flux that reaches the solar surface does however play a key role. Flux-transport (sometimes called conveyor belt) dynamos still maintain the tachocline as the preferred site of generation for toroidal field, but the poloidal field is regenerated *at the solar surface* via the decay and transport of active regions with a preferred tilt (sometimes termed the Babcock-Leighton mechanism). In these models the turbulent convection zone may not generate any large-scale flux and provides only a weak turbulent diffusion. The spatial separation of the two generation regions requires the existence of a systematic meridional flow that is able to transport the poloidal flux from the solar surface back to the tachocline—such a flow has

been observed at the solar surface and just below, but helioseismology is currently unable to measure the flows at greater depths.

Why is it proving so difficult to come to a consensus on the solar dynamo? The nature of dynamo action is subtle and involves the interaction of magnetic fields with a number of physical elements that may (or may not) play a key role in the dynamo. Usually models are constructed by arguing which of these myriad effects are important and ignoring others which are deemed to be less crucial. In this paper, I shall review the arguments that are usually made as to whether an individual effect is beneficial or detrimental to dynamo action, and also whether it is believed to help the generation of large-scale systematic fields. I shall then briefly report on some recent numerical simulations that shed light on which of these effects may be important.

2 The Physical Effects That May Play a Role in Dynamo Action

The solar interior consists of a magnetised collisional electrically conducting plasma, which is well described by the equations of magnetohydrodynamics (MHD)—unlike the solar corona where plasma effects beyond the MHD framework may need to be included. This plasma undergoes a number of interactions with (self-)gravity and rotation and so the environment for dynamo action is a magnetised, moderately rotating, compressible, stratified plasma. The interaction with rotation leads naturally to the formation of shear flows, whilst energy is transported either by radiation (in the inner regions of the Sun) or turbulent convection (in the outer 30% by radius). The presence of turbulence ensures that both the fluids and magnetic Reynolds numbers ($Re = UL/\nu$ and $Rm = UL/\eta$ respectively) are extremely large, whilst the density and temperature of the ionised plasma imply that the magnetic Prandtl number $Pm = Rm/Re$ is extremely small.

Numerical simulations of dynamo action in such extreme parameter regimes are currently not feasible, even with numerical codes optimised for use on massively parallel computers. Global simulations *are* beginning to yield insights into the nature of the interaction of turbulence, shear flows and magnetic fields (see e.g. Browning et al. 2006) but most current progress arises from gaining an understanding of the basic physics through theory and careful computation.

For each underlying physical process that contributes to the solar dynamo it is often possible to postulate reasonable hypotheses as to whether that process helps or hinders dynamo action in general and, perhaps more importantly, whether it promotes the generation of large scale magnetic fields. Some of the arguments are listed below.

- Rotation
  - *Advantageous for dynamo action:* Rotation may help dynamo action by lending increasing coherence to the fluid flow. Turbulence has both a random and coherent component and it is now believed that the presence of long-lived coherent structures amid the turbulence is beneficial for dynamo action (see e.g. Tobias and Cattaneo 2008a, 2008b). In addition the presence of rotation introduces a preferred direction into the turbulence, which may act so as to encourage sustained stretching and amplification of the field. Finally the presence of rotation can decrease the length-scale of the flow, for example decreasing the distance between strong plumes in turbulent convection. If the strong plumes are largely responsible for dynamo action then this effect will act so as to increase the filling factor of the magnetic field.

  - *Disadvantageous for dynamo action:* In convective dynamos rotation is known to decrease the vigour of the turbulence at fixed thermal driving, and so the presence of rotation may lead to the decrease of the local *Rm* and consequently the dynamo efficiency.
  - *Advantageous for* large-scale *dynamo action:* It is widely believed that the presence of cyclonic turbulence—i.e. turbulence for which there is a net handedness—is advantageous for large-scale dynamo action. In this picture, mean magnetic fields may emerge by averaging over many cyclonic small-scale interactions of field and flow (see e.g. Parker 1955; Steenbeck et al. 1966). This picture seems to work well at low *Rm* (although there may be problems at high *Rm*—see below) with large-scale fields emerging efficiently. Rotation naturally engenders a preferred handedness to the turbulence and so should act to promote the generation of large-scale fields. In addition, the presence of rotation introduces a preferred direction in the flow; this may be important in reducing small-scale fluctuations and promoting large-scale, systematic magnetic fields.
  - *Disadvantageous for* large-scale *dynamo action*: Not applicable.
- Turbulence
  - *Advantageous for dynamo action:* A successful dynamo is one where stretching of magnetic field within the fluid successfully overcomes the dissipative effects of ohmic diffusion. It is therefore apparent that, ceteris paribus, flows with good stretching properties are more likely to be good dynamos. Turbulent flows are in general chaotic and have large regions of strong stretching. Hence turbulence is generally believed to generate small-scale magnetic fields on an advective timescale with great efficiency.
  - *Disadvantageous for dynamo action:* The generation of magnetic fields on small-scales does lead to increased dissipation of the magnetic field. This does have a negative impact on dynamo action though it is believed that the enhanced diffusion is not enough to switch off dynamo action if *Rm* is high enough, i.e. all sufficiently turbulent flows will act as dynamos at high *Rm*.
  - *Advantageous for* large-scale *dynamo action:* As noted above, the interaction of rotation with turbulence is sometimes thought to lead to the generation of large-scale magnetic fields. The small-scale velocity fields interact with the small-scale magnetic fields to yield a mean electromotive force via the alpha-effect of mean field electrodynamics.
  - *Disadvantageous for* large-scale *dynamo action:* There are many reasons for supposing that the presence of turbulence may be disadvantageous for large-scale dynamo action. The presence of turbulence suggests that chaotic flows exist on a wide range of scales, from large to small. These flows will, as argued above, (kinematically) generate small-scale magnetic field very easily with an eigenfunction that is peaked at small scales. The more turbulent the flow, the more effective the small-scale dynamo. This dynamo competes with any large-scale dynamo via correlations between the magnetic field and the flow. Furthermore the presence of turbulence leads to spatial and temporal *decorrelations* in the flow and field, which reduces the prospects of generating large-scale magnetic field via mean inductive effects. Finally the generation of strong small-scale fields may saturate the mechanism for generation of large-scale fields when the mean field is still exceptionally weak. This *catastrophic quenching* (Vainshtein and Cattaneo 1992; see also Brandenburg and Subramanian 2005) may have serious consequences for mean field generation in stellar interiors.
- Stratification
  - *Advantageous for dynamo action:* Stratification enables the dynamo to access regions with different underlying thermal and hydrodynamic properties, and therefore to generate field in the region in which dynamo action is most efficient. For example the

presence of strong stratification can lead to a gradient in the intensity of the turbulent flows and then dynamo action may occur at the intensity for which dynamo action is optimised. The strong gradient can also lead to the transport of magnetic fields between different dynamo regions, allowing significant magnetic energy to be stored away from the generation region. This may help in softening the back-reaction of the Lorentz force on the generating turbulent eddies.

- *Disadvantageous for dynamo action:* It has been argued (Stein and Nordlund 2002) that the presence of significant density stratifications can be fatal for small-scale dynamos, driven by convection. In this scenario, the strong stratification leads to strong turbulent flows where the density is small (usually at the top of the layer) and weaker flows where the density is large. These convective flows are characterised by fast narrow sinking plumes and weaker broader upwellings. The fast downflows are at higher Rm and are largely responsible for the field generation. However, because at high Rm the magnetic field in these downflows is tied to the fluid, it can be argued that it shares the same fate as the fluid in the downflows, which, because of the strong stratification, is never returned to the less dense layer above. Hence the dynamo becomes increasingly inefficient and eventually switches off as the density is increased. Stein and Nordlund (2002) argue that this places significant restrictions on the possible mode of operation of a dynamo driven by convection at the solar surface.
- *Advantageous for* large-scale *dynamo action:* The presence of a stable layer has long been thought to be advantageous for large-scale dynamo action. A relatively quiescent stably stratified layer, beneath a layer of strong convection is a suitable place in which to store large-scale toroidal magnetic fields (Spiegel and Weiss 1980). Moreover, storage of magnetic flux and large-scale magnetic helicity away from a the generation region of rotating convection is believed to help remedy the *catastrophic quenching* described above (Parker 1993; Charbonneau and MacGregor 1997; Tobias 2005).
- *Disadvantageous for* large-scale *dynamo action:* The loss of large-scale poloidal magnetic flux (whether radial or latitudinal) from the region of generation in the convection zone may inhibit large-scale dynamo action. Radial field may be lost via turbulent diffusion whilst latitudinal field may be pumped out of the convection zone by radial flows. A mechanism for returning poloidal flux to this generation region, such as magnetic buoyancy (which acts primarily on toroidal field, but in the presence of rotation this can be transformed into poloidal field) may help in this regard.

- Shear
  - *Advantageous for dynamo action:* The presence of a consistent strong shear flow is known to be beneficial to dynamo action. Shear flows by their nature amplify magnetic field via the $\mathbf{B}\cdot\nabla\mathbf{u}$ term in the induction equation. This coherent stretching will systematically amplify both large and small-scale magnetic fields, leading to efficient dynamo action. For this reason sites of strong shear are often believed to be good candidates for dynamo action.
  - *Disadvantageous for dynamo action:* Not applicable.
  - *Advantageous for* large-scale *dynamo action:* The presence of shear has also been argued to favour the generation of large-scale magnetic fields over small-scale fields (in addition to the simple amplification of a large-scale field via the ω-effect). A number of mechanisms have been postulated for this. Recent interest has focussed on the interaction of shear flows with turbulence that leads to a fluctuating α-effect. The presence of shear introduces correlations to the turbulence that on a long time-scale may lead to the generation of large-scale field (Vishniac and Brandenuburg 1997; Proctor 2007). Another mechanism for the preferential generation of large-scale fields

is the *shear-current effect* which arises naturally via an expansion of the turbulent electromotive force in mean-field electrodynamics. Finally shear flows are believed to enhance diffusion, preferentially dissipating small-scale fields and leaving the larger scale fields relatively undamped.

– *Disadvantageous for* large-scale *dynamo action:* Not applicable.

All of these arguments given above are plausible, yet some are contradictory. In the next two sections I briefly summarise two numerical experiments which investigate the importance of penetration (stratification), rotation and shear for dynamo action in turbulent flows at large *Rm*.

3 The Role of Penetration on Compressible Dynamos

In principle, the importance of penetration for dynamo action can be investigated within the framework of Boussinesq models (see below). However, the importance of density stratification—with particular emphasis to the strong stratification at the solar surface—has been discussed largely with reference to compressible calculations. In this section, we summarise the results of a large-scale numerical computation to determine the effect of strong stratification and penetration on models of compressible dynamo action. The interested reader may find further details in Brummell et al. (2008). We consider a local Cartesian domain, with x and y representing longitude and latitude and z representing depth. As in Tobias et al. (2001) a penetrative formulations is achieved by stacking two plane polytropes on top of each other with the unstable layer occupying $0 \leq z \leq 1$ and the stable layer occupying $1 < z \leq z_{\max}$. The stability of the layers to convection is assured by selecting different values for the adiabatic index m and thermal diffusivity in the two layers and ensuring that the heat flux in the basic polytrope is continuous. Here we ensure that the upper layer is weakly unstable to convection and the lower layer is very stable. We solve the fully compressible equations for MHD at resolutions of up to $512 \times 512 \times 768$ for the penetrative case. The fully compressible dynamo equations with strong stratification are notoriously difficult to integrate owing to the evacuation of swirling downflows near the top boundary. It is for this reason that such a high resolution is needed.

Initially the hydrodynamic equations are integrated until a statistically steady state is achieved in a non-penetrative calculation (i.e. $z_{\max}$ was set to unity). The convection is heavily influenced by the density variation and is highly asymmetric with broad upwellings and narrower strong downflows. Once the statistically steady hydrodynamic state is achieved, a small amplitude, random seed magnetic field is added and the full magnetohydrodynamic equations are stepped forward in time. The magnetic field initially grows kinematically (exponentially on average) on an advective timescale and saturates in a state of dynamic MHD turbulence. The magnetic fields are generated primarily by the strong downflows and strong fields are focussed in the narrow downwards sinking plumes. As noted above, this concentration of magnetic field is the cause of the numerical difficulty in integrating the fully compressible equations with strong density contrasts, and makes high resolutions necessary. An alternative approach is to use sub-grid scale (SGS) models rather than direct numerical simulations (DNS) and this is the approach taken by Vögler and Schüssler (2007).

Once the saturated state of nonlinear MHD turbulence is achieved in the non-penetrative case, a penetrative layer is added underneath the convecting region and the calculation is restarted. This is to determine the effects of penetration on the nature of the solutions. As this penetrative MHD state is evolved, both kinetic and magnetic energy begin to overshoot the nominal base of the convection zone and are transported into the stable layer. In this

manner the turbulence removes magnetic energy into the stable layer whilst continuing to generate field in the convecting layer. Here the sole effect of including a penetrative layer is to provide a storage location for the magnetic field, which continues to be efficiently generated throughout the convection zone. Indeed, with the inclusion of the stable layer the total magnetic energy of the calculation increases from that with the convection alone — although the ratio of kinetic to magnetic energy remains roughly constant between the two cases.

The above discussion (and the more detailed exposition included in Brummell et al. 2008) would seem to indicate that penetration and stratification do not have a significant effect on the efficiency of dynamo action, and that no significant restriction can be placed on the operation of a dynamo at the solar surface (*pace* the results of Stein and Nordlund 2002).

4 Turbulent Boussinesq Dynamos with Penetration, Rotation and Shear

Turbulent compressible dynamos, as noted above, require computations run with extremely high resolutions for accurate solutions to be obtained. This requirement precludes the possibility of performing a parameter survey of the effects of penetration, rotation and shear within the compressible framework. In a recent paper, Tobias et al. (2008) performed such a survey within the Boussinesq framework and we highlight the important results here.

The local model again consists of a Cartesian domain (x, y, z) where z points downwards. The domain is of total depth λd, where d is the depth of a convective layer that lies above a lower convectively-stable layer. The evolution of the dynamics is described by the standard equations of Boussinesq magnetohydrodynamics (see e.g. Tobias et al. 2008) and therefore by four non-dimensional parameters. Firstly, the Rayleigh number $Ra(z)$ measures the strength of thermal buoyancy relative to dissipation. Secondly, the Taylor number Ta measures the importance of rotation compared to viscous effects. Finally, the kinetic and magnetic Prandtl numbers Pr and Pm represent the ratios of the thermal and magnetic diffusion timescales relative to the viscous diffusion timescale respectively. Penetrative convection is achieved by selecting the Rayleigh number to vary from being large and positive in the upper layer $0 \leq z \leq d$ and negative in the lower layer $d \leq z \leq z_{\max}$. Note, in this formulation the variation of Rayleigh number can be thought of as arising from variation of the coefficient of thermal expansion with depth. An additional shear-flow $\boldsymbol{u} = U(z)\hat{\mathbf{y}}$ can be added by including the relevant advective effects in the temperature, momentum, and induction equations.

In this configuration, the dynamo properties of a number of flows—and the role of turbulence, rotation, penetration and shear can be investigated in a systematic manner. We summarise the results below, and leave the interested reader to discover the details in Tobias et al. (2008). In all cases the convecting solutions were integrated to statistically steady hydrodynamic state before dynamo action was investigated. The addition of a seed field was followed by fast exponential growth—i.e. on an advective timescale—of magnetic energy followed by saturation of the magnetic energy density.

- *Penetration:* As for the compressible case, the inclusion of penetration has little effect on the dynamo properties of the flow. As the degree of penetration is increased by reducing the stability of the lower layer, the hydrodynamic state allows downflows to overshoot further into the stable layer. Thus an increasing fraction of the kinetic energy density is contained in the stable layer. The dynamo properties are largely unaffected by the presence of a stable layer, with the ratio of magnetic energy to kinetic energy remaining

largely unchanged with degree of penetration and the form of the field being largely unaffected. In all of these cases the dynamo generated field is small-scale, though there is some tendency to generate larger scale field in the stable layer. Even here though the magnetic energy is dominated by small scales.

- *Turbulence:* As expected, increasing the level of turbulence has a significant effect on the nature of dynamo action. As *Ra* is increased (with other parameters held fixed) the nature of the convection becomes more disordered and irregular. Increase in *Ra* increases both the kinetic and magnetic energy of the final saturated state, with the ratio between the two remaining largely unchanged. As the level of turbulence is increased, the generated magnetic field becomes more and more dominated by the small-scales, with the ratio of the energy in the fluctuating field to that in the large-scale field increasing with *Rm*.
- *Rotation:* The presence of rotation has three main effects on the form of convection and these feed into the dynamo properties of the flows. The addition of rotation decreases the supercriticality of the convective flow at fixed *Ra* and so the convection becomes more ordered and laminar as the Taylor number is increased. Second the horizontal scale of rotation is decreased with increasing *Ta*, which enables more convective cells to be fitted into a computational domain of fixed aspect ratio. The third, and possibly most interesting effect, is that the presence of rotation and penetration allows for the generation of a net helicity in the computational domain, which is absent in the rotating non-penetrative cases considered by Cattaneo and Hughes (2006). Frustratingly, although the penetrative rotating dynamos are all exceptionally good small-scale dynamos, no significant large-scale magnetic field is generated. Increasing the Taylor number leads to an initial increase and then a decrease in the saturation level of the magnetic energy. This arises owing to a competition between the advantageous effects of rotation increasing the packing and coherence of the dynamo plumes and the detrimental effect of rotation decreasing the supercriticality of the convection. Despite the presence of a net helicity, no significant mean field could be detected.
- *Shear:* The addition of depth-dependent shear, with a mean flow comparable with the strongest downflows, does increase the efficiency of dynamo action. The persistent stretching provided by the shear is an efficient amplifier of the field and this leads to strong generation of field in the shear layer. However in this configuration, the shear does not appear to lead to systematic generation of magnetic field with a large-scale component. The strong magnetic field is still dominated by small-scales. This disappointing result can be understood by noting that the shear is only able to amplify the magnetic field that is presented to it by the convection. As the convection only produces unsystematic large-scale components that vary on an advective timescale, the shear acts only to amplify this to produce strong but unsystematic large-scale components for the magnetic field. It has been suggested (Brandenburg private communication) that altering the spatial dependence of the shear so that it is a function of x rather than z will improve the prospects for large-scale field generation, and this idea will be tested in a subsequent paper.

5 Discussion

In this paper, I have discussed some of the issues faced by theorists in trying to construct models of dynamo action in turbulent regions such as the solar convection zone. Model construction often relies on parameterising the effects of turbulence, penetration, rotation and shear—yet the effects of these on dynamo action remain poorly understood. In local Cartesian domains the role of these in modifying the dynamo properties can be investigated

systematically in both compressible and Boussinesq models. Penetration does not appear to alter dynamo properties significantly—so there is little problem with locating a local dynamo in the strongly convecting layer just below the solar photosphere. In Boussinesq models, rotation and shear improve the efficiency of dynamo action, but they do not appear to lead to the generation of systematic large-scale magnetic fields.

Given the difficulties that turbulent dynamos encounter in generating a large-scale magnetic field, it is important to consider the opposite paradigm where the dynamo is generated by instabilities of a strong magnetic field. Such instabilities have been demonstrated to lead to systematic generation of magnetic field and are even capable of reproducing cyclic activity (see e.g. Brandenburg and Schmitt 1998; Cline et al. 2003). It is therefore of interest to determine whether these laminar dynamos continue to be as efficient in the presence of turbulence, which would inevitably be present in a stellar interior.

Acknowledgements I would like to thank the anonymous referee and Nigel Weiss for helpful comments. I would also like to acknowledge the support of the Leverhulme Trust. This paper was written at the dynamo program at the Kavli Institute for Theoretical Physics and was supported in part by the National Science Foundation under Grant No. PHY05-51164.

References

S. Boldyrev, F. Cattaneo, Magnetic-field generation in Kolmogorov turbulence. Phys. Rev. Lett. **92**, 144501 (2004)

A. Brandenburg, The case for a distributed solar dynamo shaped by near-surface shear. Astrophys. J. **625**, 539–547 (2005)

A. Brandenburg, D. Schmitt, Simulations of an alpha-effect due to magnetic buoyancy. Astron. Astrophys. **338**, L55–L58 (1998)

A. Brandenburg, K. Subramanian, Astrophysical magnetic fields and nonlinear dynamo theory. Phys. Rep. **417**, 1–209 (2005)

M.K. Browning, M.S. Miesch, A.S. Brun, J. Toomre, Dynamo action in the solar convection zone and tachocline: pumping and organization of toroidal fields. Astrophys. J. **648**(2), L157–L160 (2006)

N.H. Brummell, F. Cattaneo, S.M. Tobias, The role of penetration in compressible dynamos (2008, in preparation)

F. Cattaneo, D.W. Hughes, Dynamo action in a rotating convective layer. J. Fluid Mech. **553**, 401–418 (2006)

P. Charbonneau, K.B. MacGregor, Solar interface dynamos II. Linear kinematic models in spherical geometry. Astrophys. J. **486**, 502 (1997)

K.S. Cline, N.H. Brummell, F. Cattaneo, Dynamo action driven by shear and magnetic buoyancy. Astrophys. J. **599**, 1449–1468 (2003)

A.B. Iskakov, A.A. Schekochihin, S.C. Cowley, J.C. McWilliams, M.R.E. Proctor, Numerical demonstration of fluctuation dynamo at low magnetic Prandtl numbers. Phys. Rev. Lett. **98**, 208501 (2007)

M. Ossendrijver, The solar dynamo. Astron. Astrophys. Rev. **11**, 287 (2003)

E.N. Parker, Hydromagnetic dynamo models. Astrophys. J. **122**, 293 (1955)

E.N. Parker, A solar dynamo surface wave at the interface between convection and nonuniform rotation. Astrophys. J. **408**, 707 (1993)

M.R.E. Proctor, Effects of fluctuation on $\alpha - \omega$ dynamo models. Month. Not. R. Astron. Soc.: Lett. **382**, L39–L42 (2007)

E.A. Spiegel, N.O. Weiss, Magnetic activity and variations in solar luminosity. Nature **287**, 616 (1980)

M. Steenbeck, F. Krause, K.-H. Rädler, Z. Naturforsch **21a**, 364 (1966)

R.F. Stein, A. Nordlund, Solar surface magneto-convection and dynamo action, in *SOLMAG 2002. Proceedings of the Magnetic Coupling of the Solar Atmosphere Euroconference and IAU Colloquium 188, 11–15 June 2002, Santorini, Greece*, ed. by H. Sawaya-Lacoste. ESA SP-505, vol. 89 (ESA, Noordwijk, 2002), pp. 83–89

J.-C. Thelen, Non-linear $\alpha - \omega$-dynamos driven by magnetic buoyancy. Mon. Not. R. Astron. Soc. **315**, 165–183 (2000)

S.M. Tobias, The solar tachocline: Formation, stability and its role in the solar dynamo, in *Fluid Dynamics and Dynamos in Astrophysics and Geophysics*, ed. by A.M. Soward, C.A. Jones, D.W. Hughes, N.O. Weiss (CRC, Boca Raton, 2005), pp. 193

S.M. Tobias, N.H. Brummell, T.L. Clune, J. Toomre, Transport and storage of magnetic field by overshooting turbulent compressible convection. Astrophys. J. **549**, 1183–1203 (2001)
S.M. Tobias, N.O. Weiss, Stellar dynamos, in *Mathematical Aspects of Natural Dynamos*, ed. by E. Dormy, A.M. Soward (CRC, Boca Raton, 2007), pp. 281–312
S.M. Tobias, F. Cattaneo, Dynamo action in complex flows: the quick and the fast. J. Fluid Mech. **601**, 101–122 (2008a)
S.M. Tobias, F. Cattaneo, On the limited role of spectra in dynamo theory. Phys. Rev. Lett. **101**, 125003 (2008b)
S.M. Tobias, F. Cattaneo, N.H. Brummell, Dynamo action with penetration, rotation and shear. Astrophys. J. **685**, 596 (2008)
S.I. Vainshtein, F. Cattaneo, Nonlinear restrictions on dynamo action. Astrophys. J. **393**, 165–171 (1992)
E. Vishniac, A. Brandenuburg, An incoherent alpha-omega dynamo in accretion disks. Astrophys. J. **475**, 263 (1997)
A. Vögler, M. Schüssler, A solar surface dynamo. Astron. Astrophys. **465**, L43–L46 (2007)
N.O. Weiss, Thompson, Space Sci. Rev. (2008, this issue)

Advances in Theory and Simulations of Large-Scale Dynamos

Axel Brandenburg

Originally published in the journal Space Science Reviews, Volume 144, Nos 1–4, 87–104.
DOI: 10.1007/s11214-009-9490-0

Abstract Recent analytical and computational advances in the theory of large-scale dynamos are reviewed. The importance of the magnetic helicity constraint is apparent even without invoking mean-field theory. The tau approximation yields expressions that show how the magnetic helicity gets incorporated into mean-field theory. The test-field method allows an accurate numerical determination of turbulent transport coefficients in linear and nonlinear regimes. Finally, some critical views on the solar dynamo are being offered and targets for future research are highlighted.

Keywords Solar dynamo · Sun · Magnetic fields · Magnetic activity

1 Introduction

Over the past 50 years significant progress has been made in understanding the origin of the solar magnetic field. In an important paper, Parker (1955) introduced the idea of mean magnetic fields and identified the α effect as the crucial ingredient of large-scale dynamos. He also proposed and solved an explicit one-dimensional mean-field model and found the migratory Parker dynamo wave. This provided an important tool for understanding the effects of α and shear, and it led to useful estimates for the excitation conditions, the cycle period, and the direction of field migration in solar and stellar dynamo models. However, Parker's work appeared at a time when it was still unclear whether homogeneous fluid dynamos really exist. These are dynamos of uniformly conducting matter, without insulating wires that are thus susceptible to "short circuits". In the years following Cowling's (1933) theorem, it remained doubtful whether the Sun's magnetic field can be explained in terms of dynamo theory, as originally anticipated by Larmor (1919).

In the paper on his famous theorem, Cowling (1933) concluded "The theory proposed by Sir Joseph Larmor, that the magnetic field of a sunspot is maintained by the currents it induces in moving matter, is examined and shown to be faulty; the same result also applies

A. Brandenburg (✉)
Nordita, Roslagstullsbacken 23, 10691 Stockholm, Sweden
e-mail: brandenb@nordita.org

M.J. Thompson et al. (eds.), *The Origin and Dynamics of Solar Magnetism*.
DOI: 10.1007/978-1-4419-0239-9_7

for the similar theory of the maintenance of the general field of Earth and Sun." Larmor (1934) responded that "the self-exciting dynamo analogy is still, so far as I know, the only foundation on which a gaseous body such as the Sun could possess a magnetic field: so that if it is demolished there could be no explanation of the Sun's magnetic field even remotely in sight."

Although the first qualitative ideas on homogeneous dynamos were proposed nearly a hundred years ago, the resistance was immense; for historical accounts see the reviews by Krause (1993) and Weiss (2005). An important existence proof for homogeneous self-excited dynamos was that of Herzenberg (1958), who showed, using asymptotic theory, that dynamos work in a conducting medium where two rotors spin about axes that lie in planes perpendicular to their direction of separation, and inclined relative to each other by an angle between 90° and 180°. Such systems were later realized experimentally by Lowes and Wilkinson (1963, 1968). In their experiments oscillations commonly occurred. Those where thought to be some kind of nonlinear relaxation oscillations. However, they used angles of less than 90°. Indeed, when the relative angle between the rotors is between 0° and 90°, oscillatory solutions are expected even from linear theory (Brandenburg et al. 1998). Those solutions were not captured by the original analysis of Herzenberg (1958), because he only looked for steady solutions.

The next important steps came with the development of mean-field electrodynamics by Steenbeck et al. (1966), who used the first order smoothing approximation (or second order correlation approximation) as a rigorous tool to compute α effect and turbulent diffusivity in limiting cases. Steenbeck and Krause (1969) later produced global mean-field models in spherical geometry and computed synthetic butterfly diagrams. For an introduction to mean-field theory we refer to the article by N.O. Weiss in this issue.

The technical tools made available by mean-field theory have stimulated much of the research in the field during the 1970s. However, during the 1980s a number of problems were discussed. For example, doubts were raised whether turbulent magnetic diffusion still works at large magnetic Reynolds numbers, $R_{\rm m}$; see work by Knobloch (1978), Layzer et al. (1979) and Piddington (1981). This problem applies equally to kinematic and nonlinear cases. Regarding the kinematic α effect, Childress (1979) found that in steady convection α decreases with increasing $R_{\rm m}$ like $R_{\rm m}^{-1/2}$. This result is now understood to be a common feature of steady flows (Rädler et al. 2002; Rädler and Brandenburg 2009), and is generally not shared by unsteady (e.g. turbulent) flows (Sur et al. 2008).

The nonlinear problem was a focus of much of the work on dynamos during the 1990s, and started with the work of Cattaneo and Vainshtein (1991, hereafter referred to as CV91) who showed, using two-dimensional turbulence simulations, that for $\overline{\boldsymbol{B}}^2 \approx B_{\rm eq}^2$, $\eta_{\rm t}$ decreases like $R_{\rm m}^{-1}$. It was expected that a similar relation applies also to α (Vainshtein and Cattaneo 1992, hereafter VC99), but this required three-dimensional considerations. Indeed, using uniform imposed fields, Cattaneo and Hughes (1996, hereafter CH96) showed that α decays with increasing $R_{\rm m}$ like $R_{\rm m}^{-1}$. These results were later understood to be due to the presence of conservation laws for the mean squared vector potential, $\langle \boldsymbol{A}^2 \rangle$, in two dimensions and the magnetic helicity, $\langle \boldsymbol{A} \cdot \boldsymbol{B} \rangle$, in three dimensions Gruzinov and Diamond (1994, hereafter GD94, 1995). Here, $\boldsymbol{A}$ is the magnetic vector potential with $\boldsymbol{B} = \boldsymbol{\nabla} \times \boldsymbol{A}$. However, these conservation laws only tell us how much small-scale magnetic field is being produced as the mean-field dynamo produces large-scale field, such that $\langle \boldsymbol{A}^2 \rangle$ (in two dimensions) or $\langle \boldsymbol{A} \cdot \boldsymbol{B} \rangle$ (in three dimensions) remain unchanged. One still needs a theory that relates the corresponding small-scale mean squared vector potential to the turbulent magnetic diffusion in two dimensions or the small-scale magnetic helicity to the turbulent diffusivity or the α

Table 1 Summary of results obtained over the years. The key to the references is given at the end of Sect. 1

Result	Details	Reference
$\eta_t \sim R_m^{-1}$	2-D periodic, $\overline{B} \sim \sin kx$	CV91
$\alpha \sim R_m^{-1}$	phenomenology, $\langle \boldsymbol{A} \cdot \boldsymbol{B} \rangle$ conservation, simulations with $\overline{B} = \text{const}$	VC92, GD94, CH96
$\overline{B}^2/B_{eq}^2 \sim R_m^{-1}$	helical turbulence, normal field b.c.	GD94, BD01
$\overline{B}^2/B_{eq}^2 \sim k_f/k_1$	helical turbulence, periodic domain	B01
$\overline{B}^2/B_{eq}^2 \gg k_f/k_1$	helical turb. with shear, periodic	BBS01, BB02
$\overline{B}^2/B_{eq}^2 \sim 0.5$	helical turb. with horizontal shear, normal field b.c.	B05
$\overline{B}^2/B_{eq}^2 \ll 0.5$	convection with vertical shear, normal field b.c.	TCB08
$\overline{B}^2/B_{eq}^2 \sim 0.5$	convection with horizontal shear, normal field b.c.	KKB08

effect in three dimensions. This can be done using a corresponding mean-field equation for these quantities.

The effect of such nonlinear dependencies of turbulent transport coefficients on the dynamo can be quite dramatic. In three dimensions, Gruzinov and Diamond (1995) showed that in the case of pseudo-vacuum boundary conditions the saturation field strength of a dynamo with just helicity is of the order of $R_m^{-1/2} B_{eq}$. This was also confirmed by simulations (Brandenburg and Dobler 2001, hereafter BD01; Brandenburg and Subramanian 2005a). In the special case of periodic boundary conditions, however, the field strength does not decline, but remains of the order of $(k_f/k_1)^{1/2} B_{eq}$ (Brandenburg 2001, hereafter B01). This is now well understood as being a consequence of magnetic helicity evolution, which was soon applied to cases with shear (Brandenburg et al. 2001, hereafter BBS01; Blackman and Brandenburg 2002, hereafter BB02) in domains with periodic as well as open boundary conditions (Brandenburg 2005, hereafter B05). Magnetic helicity evolution has also been invoked to understand recent simulations of convection by Tobias et al. (2008, hereafter; TCB08) and Käpylä et al. (2008a, hereafter KKB08).

Table 1 summarizes a number of results that have been obtained over the years. These results may appear conflicting at first sight, but they are in fact all explained by modern dynamo theory that takes magnetic helicity evolution into account, and that allows for magnetic helicity changes in the presence of losses through boundaries. In the following we restrict ourselves to cases in Cartesian geometry, but we note that important progress is now also being made in spherical shell geometry where large-scale fields have been seen when rotation is sufficiently rapid (Brown et al. 2007).

2 Saturation Phenomenology in a Periodic Box

During the early phase of a strongly helical dynamo there can be a phase during which the magnetic energy of the large-scale field is still subdominant. However, at later times the magnetic energy can redistribute itself from small to large scales. The fields that suffer minimal back-reaction from the Lorentz force tend to be force-free at large scales. Force-free fields are generally referred to as Beltrami fields. Qualitatively speaking, the helical driving produces a helical field at the driving scale, but because magnetic helicity cannot change, helical field of opposite helicity must emerge at some other scale. Simple arguments show that this can only happen at a larger scale (Frisch et al. 1975; see also Brandenburg and Subramanian 2005b). To explain the evolution of the resulting large-scale magnetic field, let

us begin with the evolution equation of magnetic helicity,

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle \boldsymbol{A}\cdot\boldsymbol{B}\rangle = -2\eta\mu_0\langle \boldsymbol{J}\cdot\boldsymbol{B}\rangle, \tag{1}$$

where angular brackets denote volume averages, η is the microscopic magnetic diffusivity, μ_0 is the vacuum permeability, and $\boldsymbol{J}=\boldsymbol{\nabla}\times\boldsymbol{B}/\mu_0$ is the current density. Next, we introduce horizontal averages denoted by overbars. The direction over which we take these averages depends of course on the direction in which the mean magnetic field chooses to align itself. There are three equivalent possibilities, so let us assume that the field shows a large-scale modulation in the z direction. In a periodic box the Beltrami field with the smallest wavenumber is then of the form

$$\overline{\boldsymbol{B}} = \overline{\boldsymbol{B}}(z,t) = \hat{B}(t)\,(\cos k_1 z, \sin k_1 z, 0)\,, \tag{2}$$

where we have ignored the possibility of an arbitrary phase shift in the z direction. Note that $\overline{\boldsymbol{J}}(z,t) = -k_1\overline{\boldsymbol{B}}/\mu_0$ and $\overline{\boldsymbol{A}}(z,t) = -k_1^{-1}\overline{\boldsymbol{B}}$, so the current and magnetic helicities have negative sign at large scales. This is the situation when the small-scale driving has positive helicity.

Note that the definition of averaging automatically defines small-scale (or fluctuating) magnetic fields as $\boldsymbol{b}=\boldsymbol{B}-\overline{\boldsymbol{B}}$, and likewise for $\boldsymbol{a}=\boldsymbol{A}-\overline{\boldsymbol{A}}$ and $\boldsymbol{j}=\boldsymbol{J}-\overline{\boldsymbol{J}}$. We can then split (1) into contributions from large scales and small scales, reorganize the equations in terms of $\langle\overline{\boldsymbol{B}}^2\rangle$ and $\langle\boldsymbol{b}^2\rangle$, assume that, after the end of the kinematic phase ($t=t_\mathrm{s}$), $\langle\boldsymbol{b}^2\rangle$ is approximately constant in time (approximately equal to $\mu_0\langle\rho\boldsymbol{u}^2\rangle$). This yields (B01)

$$k_1^{-1}\frac{\mathrm{d}\langle\overline{\boldsymbol{B}}^2\rangle}{\mathrm{d}t} = 2\eta k_\mathrm{f}\langle\boldsymbol{b}^2\rangle - 2\eta k_1\langle\overline{\boldsymbol{B}}^2\rangle, \tag{3}$$

which has the solution

$$\langle\overline{\boldsymbol{B}}^2\rangle = \langle\boldsymbol{b}^2\rangle\frac{k_\mathrm{f}}{k_1}\left[1-\mathrm{e}^{-2\eta k_1^2(t-t_\mathrm{s})}\right]. \tag{4}$$

Thus, $\langle\overline{\boldsymbol{B}}^2\rangle$ saturates on a time scale $(2\eta k_1^2)^{-1}$, i.e. the microscopic diffusion time based on the scale of the box. This equation reproduces extremely well the saturation behavior in a periodic box. This equation also shows what happens if either the fluctuating field or the mean field are not fully helical (Brandenburg et al. 2002). For example, if the large-scale field is no longer fully helical, then the ratio $\mu_0|\langle\overline{\boldsymbol{J}}\cdot\overline{\boldsymbol{B}}\rangle|/\langle\overline{\boldsymbol{B}}^2\rangle$ will be less then k_1, so we say that the *effective* value of k_1 will be smaller. (Later on we refer to this value as k_m.) Thus, if the large-scale field is not fully helical, but the small-scale field is still fully helical, then the effective value of k_1 in the denominator of (4) decreases and $\langle\overline{\boldsymbol{B}}^2\rangle$ can be even somewhat higher than for periodic boundary conditions. This is indeed the case for perfectly conducting boundary conditions, which do not permit (2) as a solution. This is the reason why the effective value of k_1 is smaller, and hence $\langle\overline{\boldsymbol{B}}^2\rangle$ is larger (Brandenburg and Dobler 2002), Conversely, if the small-scale field is not fully helical, the effective value of k_f is smaller, and so $\langle\overline{\boldsymbol{B}}^2\rangle$ is smaller (Maron and Blackman 2002; Brandenburg et al. 2002).

We emphasize that in the considerations in this section we did not invoke mean-field theory at all. The slow-down during the final saturation stage is rather general and it should be possible to describe this by a sufficiently detailed mean-field theory. This will be discussed briefly in the following section.

3 Mean-Field Theory and Transport Coefficients

In mean-field theory one considers the averaged induction equation. The cross-product of the correlation of the fluctuations $\boldsymbol{u} = \boldsymbol{U} - \overline{\boldsymbol{U}}$ and $\boldsymbol{b} = \boldsymbol{B} - \overline{\boldsymbol{B}}$, i.e. the mean electromotive force, $\overline{\boldsymbol{\mathcal{E}}} = \overline{\boldsymbol{u} \times \boldsymbol{b}}$, provides an important term in the averaged induction equation,

$$\frac{\partial \overline{\boldsymbol{B}}}{\partial t} = \boldsymbol{\nabla} \times \left(\overline{\boldsymbol{U}} \times \overline{\boldsymbol{B}} + \overline{\boldsymbol{\mathcal{E}}} - \eta \mu_0 \overline{\boldsymbol{J}} \right). \tag{5}$$

A central goal of mean-field theory is to find expressions for $\overline{\boldsymbol{\mathcal{E}}}$ in terms of mean-field quantities. Quadratic correlations such as $\overline{\boldsymbol{\mathcal{E}}}$ are obtained using evolution equations for the fluctuations, $\boldsymbol{u} \equiv \boldsymbol{U} - \overline{\boldsymbol{U}}$ and $\boldsymbol{b} = \overline{\boldsymbol{B}} - \boldsymbol{b}$. A range of different approaches can be used to calculate the functional form of the mean electromotive force, $\overline{\boldsymbol{\mathcal{E}}} = \overline{\boldsymbol{u} \times \boldsymbol{b}}$, including the second order correlation approximation (SOCA), the τ approximation, and the renormalization group procedure. Common to both the SOCA and the τ approximation is the fact that the linear terms in the evolution equations for the fluctuations are solved exactly. However, there is an important difference in that the τ approximation starts by computing the time evolution of $\overline{\boldsymbol{\mathcal{E}}}$, so one begins with

$$\partial \overline{\boldsymbol{\mathcal{E}}} / \partial t = \overline{\dot{\boldsymbol{u}} \times \boldsymbol{b}} + \overline{\boldsymbol{u} \times \dot{\boldsymbol{b}}}, \tag{6}$$

whereas under SOCA one uses primarily the induction equation by computing $\overline{\boldsymbol{\mathcal{E}}} = \overline{\boldsymbol{u} \times \boldsymbol{b}}$, where $\boldsymbol{u}$ is assumed given and $\boldsymbol{b}$ is being solved using the Green's function for the induction equation. In simple terms, this reduces to solving for $\overline{\boldsymbol{\mathcal{E}}} = \overline{\boldsymbol{u} \times \int \dot{\boldsymbol{b}} \, \mathrm{d}t}$. This distinction is important because under the τ approximation the term $\overline{\dot{\boldsymbol{u}} \times \boldsymbol{b}}$ leads immediately to a term of the form $\overline{(\boldsymbol{j} \times \overline{\boldsymbol{B}}) \times \boldsymbol{b}}$ owing to the Lorentz force. This expression leads to an important feedback by attenuating the α effect by a term α_M, where, under the assumption of isotropy, $\alpha_M = \frac{1}{3} \tau \overline{\boldsymbol{j} \cdot \boldsymbol{b}}$ is the magnetic α effect. Another important difference is that there is a natural occurrence of a time derivative of $\overline{\boldsymbol{\mathcal{E}}}$. Thus, compared with SOCA, which leads to

$$\overline{\mathcal{E}}_i = \alpha_{ij} \overline{B}_j + \eta_{ijk} \overline{B}_{j,k} \,, \tag{7}$$

one now has

$$\tau \partial \overline{\mathcal{E}}_i / \partial t + \overline{\mathcal{E}}_i = \alpha_{ij} \overline{B}_j + \eta_{ijk} \overline{B}_{j,k} \,, \tag{8}$$

where τ is a relaxation time, and a comma between indices denotes a spatial derivative. In (8) the origin of the $\tau \partial \overline{\boldsymbol{\mathcal{E}}} / \partial t$ term is clear in view of (6), and it is instead the $\overline{\boldsymbol{\mathcal{E}}}$ term that is due to retaining nonlinear terms in the evolution equations for $\boldsymbol{u}$ and $\boldsymbol{b}$. In both cases these terms lead to the triple correlations that are then approximated by $-\overline{\boldsymbol{\mathcal{E}}}/\tau$ on the right hand side. After multiplying by τ, this leads to the $\overline{\boldsymbol{\mathcal{E}}}$ term in (8).

In the expressions above we have used the more general tensorial forms of α effect and turbulent diffusion. Scalar transport coefficients used before denote the isotropic contributions of the α_{ij} and η_{ijk} tensors, i.e. $\alpha = \frac{1}{3} \delta_{ij} \alpha_{ij}$ and $\eta_{\mathrm{t}} = \frac{1}{6} \varepsilon_{ijk} \eta_{ijk}$.

Both SOCA and the τ approximation are rather primitive and their merits has been discussed in some detail in the recent literature (Rädler and Rheinhardt 2007; Sur et al. 2007). The emergence of the $\overline{\boldsymbol{j} \cdot \boldsymbol{b}}$ term is qualitatively a new feature that leads to a quantitative description of the saturation of large-scale dynamos in periodic domains (Field and Blackman 2002, BB02). Furthermore, the emergence of an additional time derivative in (8) has been confirmed qualitatively using simulations (Brandenburg et al. 2004). However, there is now

also evidence for the occurrence of even higher time derivatives in some cases (Hubbard and Brandenburg 2008).

The time derivative in (8) suppresses changes of mean-field properties on timescales shorter than the turnover time τ of the turbulence. This is analogous to the occurrence of the Faraday displacement current in the Maxwell equations, except that there the limiting velocity is the speed of light, whereas here it is the rms velocity of the turbulence. This changes the parabolic nature of the diffusion and dynamo equations into hyperbolic wave equations (Blackman and Field 2003; Brandenburg et al. 2004). This property is physically appealing, because it retains causality, which means here that no mean-field pattern can propagate faster than the rms velocity of the turbulence.

Similar to the suppression of fast temporal variations discussed above, there is also a suppression of spatial variations on short length scales. Indeed, (8) takes the more accurate form

$$\tau \partial \overline{\mathcal{E}}_i / \partial t + \overline{\mathcal{E}}_i = \hat{\alpha}_{ij} \circ \overline{B}_j + \hat{\eta}_{ijk} \circ \overline{B}_{j,k}, \tag{9}$$

where $\hat{\alpha}_{ij}$ and $\hat{\eta}_{ijk}$ are the components of integral kernels and the circles denote a convolution. Recent numerical work has now established that for driven turbulence the integral kernels have an exponential form with a width given by the inverse wavenumber of the energy-carrying eddies (Brandenburg et al. 2008b).

This implies that mean-field theory should never produce rapid spatial or temporal variations. Conversely, the more complicated kernel formulation in (9) can be avoided if the solutions are sufficiently smooth in space and time. However, this is not always guaranteed, especially near boundaries.

Let us at this point also highlight the occurrence of another time derivative in the mean-field equations. Under the τ approximation, the $\overline{(\boldsymbol{j} \times \boldsymbol{B})} \times \boldsymbol{b}$ term leads to the emergence of a magnetic contribution to the α effect. The full α effect is then written as $\alpha = \alpha_{\rm K} + \alpha_{\rm M}$, where $\alpha_{\rm K}$ is related to the kinetic helicity and $\alpha_{\rm M}$ is related to the current helicity. The latter obeys an evolution equation where the omission of the time-derivative is often problematic, especially when $R_{\rm m}$ is large and the mean divergence of current helicity fluxes vanishes. Therefore, the more complete quenching formula with extra effects included takes the form (see, e.g., Brandenburg 2008),

$$\alpha = \frac{\alpha_0 + R_{\rm m}\left(\eta_{\rm t} \frac{\mu_0 \overline{\boldsymbol{J}} \cdot \overline{\boldsymbol{B}}}{B_{\rm eq}^2} - \frac{\boldsymbol{\nabla} \cdot \overline{\mathcal{F}}_{\rm C}}{2 k_{\rm f}^2 B_{\rm eq}^2} - \frac{\partial \alpha / \partial t}{2 \eta_{\rm t} k_{\rm f}^2}\right)}{1 + R_{\rm m} \overline{\boldsymbol{B}}^2 / B_{\rm eq}^2}. \tag{10}$$

Although this equation can be written as an evolution equation, in practice there is a computational advantage in solving the time-derivative term implicitly; see Brandenburg and Käpylä (2007). The properties of such a "dynamical" α quenching formula have been studied in a number of recent papers including Kleeorin et al. (2000), Field and Blackman (2002), BB02, and Brandenburg and Subramanian (2005a).

4 The Test-Field Method

In the last few years a new and reliable method for calculating the α_{ij} and η_{ijk} tensor coefficients has become available. This method is known as the test-field method and was developed by Schrinner et al. (2005, 2007) to calculate all tensor components from snapshots of simulations of the geodynamo in a spherical shell. This method was later applied

to time-dependent turbulence in triply-periodic Cartesian domains, both with shear and no helicity (Brandenburg 2005; Brandenburg et al. 2008a) as well as without shear, but with helicity (Sur et al. 2008; Brandenburg et al. 2008b), and also with both (Mitra et al. 2008a).

4.1 The Essence of the Test-Field Method

In the test-field method one solves an additional set of three-dimensional partial differential equations for vector fields $\boldsymbol{b}^{pq}$, where the labels $p = 1, 2$ and $q = 1, 2$ correspond to different pre-determined one-dimensional test fields $\overline{\boldsymbol{B}}^{pq}$. The evolution equations for $\boldsymbol{b}^{pq}$ are derived by subtracting the mean-field evolution equation from the evolution equation for $\boldsymbol{B}$. These equations are *distinct* from the original induction equation in that the curl of the resulting mean electromotive force is subtracted.

The test-field method has recently been criticized by Cattaneo and Hughes (2008) on the grounds that the test fields are arbitrary pre-determined mean fields. They argue that the resulting turbulent transport coefficients will only be approximations to the true values unless the test fields are close to the actual mean fields. Mitra et al. (2008a) have reviewed arguments supporting the validity of the test-field method: (i) the test-field method correctly reproduces a vanishing growth rate in saturated nonlinear cases (Brandenburg et al. 2008c); (ii) in the time-dependent case, the test-field method correctly reproduces also a non-vanishing growth rate. In that case one must write (7) as a convolution in time (Hubbard and Brandenburg 2008); (iii) for the Roberts flow with a mean field of Beltrami type, the α_{ij} tensor is anisotropic and has an additional component proportional to $\overline{B}_i \overline{B}_j$ that tends to quench the components of the isotropic part of α_{ij}. The same α_{ij} tensor also governs the evolution of a mean passive vector field. It turns out that the fastest growing passive vector field is then phase-shifted by 90 degrees relative to the one that caused the quenching and thus the quenched form of α_{ij}. This result has been confirmed both numerically and using weakly nonlinear theory (Tilgner and Brandenburg 2008). We discuss this case further in Sect. 4.4.

4.2 $R_{\rm m}$-Dependence of the Kinematic Values of α and $\eta_{\rm t}$

Using the test-field method it has, for the first time, become possible to obtain reliable estimates not only for the α effect, but in particular also for the turbulent magnetic diffusivity. Restricting ourselves to the case of horizontal (xy) averages, the mean fields depend only on z and t. All components of $\overline{B}_{j,k}$ can therefore be expressed in terms of those of $\overline{\boldsymbol{J}}(z,t)$, and the relevant components of η_{ijk} reduce to a rank-2 tensor, η_{ij}. In that case, $\eta_{\rm t} = \frac{1}{2}(\eta_{11} + \eta_{22})$. We present the $R_{\rm m}$ dependences of α and $\eta_{\rm t}$ in normalized forms using the SOCA results for homogeneous isotropic turbulence as reference values,

$$\alpha_0 = -\frac{1}{3}\tau \overline{\boldsymbol{\omega}\cdot\boldsymbol{u}}, \qquad \eta_{\rm t0} = \frac{1}{3}\tau \overline{\boldsymbol{u}^2} \quad \text{(SOCA, linear)}. \tag{11}$$

It turns out that, in the kinematic regime, α_0 and $\eta_{\rm t0}$ are remarkably close to the numerically determined values of α and $\eta_{\rm t}$ in the range $1 < R_{\rm m} < 200$ considered in the study of Sur et al. (2008); see Fig. 1. For $R_{\rm m} < 1$, both α and $\eta_{\rm t}$ increase linearly with $R_{\rm m}$. In the cases considered here we have assumed that the turbulence is fully helical, so $\overline{\boldsymbol{\omega}\cdot\boldsymbol{u}} \approx k_{\rm f}\overline{\boldsymbol{u}^2}$, and that the Strouhal number, ${\rm St} \equiv \tau u_{\rm rms} k_{\rm f}$ is approximately equal to unity (Brandenburg and Subramanian 2005c, 2007). Of course, for $R_{\rm m} < 1$ this is not the case and then $\tau \approx (\eta k_{\rm f}^2)^{-1}$ is a better estimate. This explains the linear increase of α and $\eta_{\rm t}$ for $R_{\rm m} < 1$.

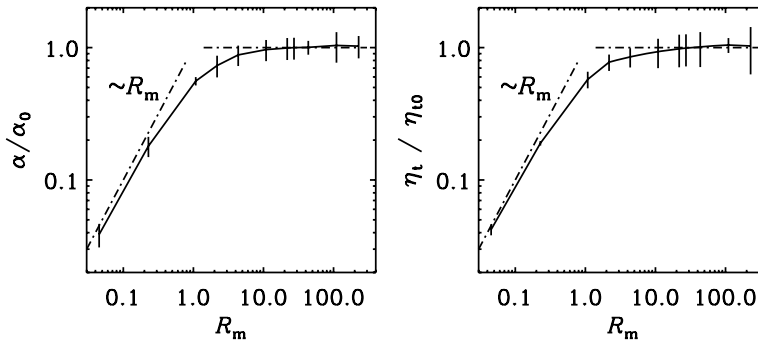


Fig. 1 Dependence of the normalized values of α and η_t on R_m for Re $= 2.2$. The *vertical bars* denote twice the error estimated by averaging over subsections of the full time series. The run with $R_m = 220$ (Re $= 2.2$) was done at a resolution of 512^3 meshpoints. Adapted from Sur et al. (2008)

4.3 Scale-Dependence of α and η_t

Using the test-field method, it has now also been possible to determine what happens if there is poor scale separation, for example if the scale of the mean field is only 2–5 times bigger than the scale of the energy-carrying eddies. In that case one can not longer write the electromotive force in terms of products of α and the mean field or η_t and the mean current density, but one has to write them as convolutions with corresponding integral kernels (e.g. Brandenburg and Sokoloff 2002). In Fourier space, a convolution corresponds to a multiplication. In the test-field method we use only harmonic test fields with a single wavenumber, so we can use this method to calculate α and η_t separately for each wavenumber and obtain the integral kernels via Fourier transformation.

Not surprisingly, it turns out that α and η_t decrease with decreasing scale, i.e. with increasing values of k/k_f, where k is the wavenumber of a particular Fourier mode of the field. In fact, by calculating α and η_t for test-fields of different wavenumber k, one finds that for isotropic turbulence, α and η_t have Lorentzian profiles of the form

$$\alpha(k) = \frac{\alpha_0}{1 + (a_\alpha k/k_f)^2}, \qquad \eta_t(k) = \frac{\eta_{t0}}{1 + (a_\eta k/k_f)^2}, \tag{12}$$

where a_α and a_η are factors of order unity; Brandenburg et al. (2008b) find $a_\alpha \approx 1$ and $a_\eta \approx 0.5$. However, for shear-flow turbulence Mitra et al. (2008a) find $a_\alpha \approx a_\eta \approx 0.7$.

In periodic domains the Fourier transforms of $\alpha(k)$ and $\eta_t(k)$ correspond to the integral kernels introduced in (9). They are of exponential form, i.e.,

$$\hat{\alpha}(z - z') = \frac{1}{2} a_\alpha \alpha_0 k_f \sim \exp(-k_f |z - z'|/a_\alpha) \tag{13}$$

and

$$\hat{\eta}_t(z - z') = \frac{1}{2} a_\eta \eta_{t0} k_f \sim \exp(-k_f |z - z'|/a_\eta). \tag{14}$$

It is important to realize that the test-field method is a tool to analyze the velocity field that is giving rise to α and η_t effects. By applying the test-field method to the case where the induction equation is solved together with the momentum and continuity equations, one can analyze the nonlinear case for one specific value of $\overline{\boldsymbol{B}}$. We emphasize that the test field does not enter the momentum equation in any way. This will be discussed next.

4.4 Quenching for Equipartition-Strength Fields

Once the magnetic field has become sufficiently strong, α and η_t will become anisotropic, even though the turbulence was originally isotropic. If the anisotropy is only due to $\overline{\boldsymbol{B}}$, the tensors α_{ij} and η_{ijk} are of the form

$$\alpha_{ij}(\overline{\boldsymbol{B}}) = \alpha_1(\overline{\boldsymbol{B}})\delta_{ij} + \alpha_2(\overline{\boldsymbol{B}})\hat{\overline{B}}_i\hat{\overline{B}}_j, \tag{15}$$

$$\eta_{ij}(\overline{\boldsymbol{B}}) = \eta_1(\overline{\boldsymbol{B}})\delta_{ij} + \eta_2(\overline{\boldsymbol{B}})\hat{\overline{B}}_i\hat{\overline{B}}_j, \tag{16}$$

where $\hat{\overline{\boldsymbol{B}}} = \overline{\boldsymbol{B}}/|\overline{\boldsymbol{B}}|$ is the unit vector of the mean field.

For equipartition-strength fields, $|\overline{\boldsymbol{B}}| = O(B_{\rm eq})$, the $R_{\rm m}$ dependence of α_1, α_2, η_1, and η_2 has been determined by Brandenburg et al. (2008c). It turns out that α_1 and α_2 have opposite signs (Fig. 2), so when α_{ij} is applied to the actual mean field we have

$$\alpha_{ij}B_j = (\alpha_1 + \alpha_2)B_i. \tag{17}$$

This shows that the α effect is magnetically quenched by the suppressing effect of α_2 on α_1 due to its opposite sign. However, even though the value of $\alpha_1 + \alpha_2$ decreases with increasing values of $R_{\rm m}$, it is only quenched down to values comparable to the value of $\eta_1 k_1$ if $|\overline{\boldsymbol{B}}| = O(B_{\rm eq})$; see Fig. 2. This becomes obvious by looking at the expression for the linear growth rate,

$$\lambda = (\alpha_1 + \alpha_2)k_{\rm m} - (\eta + \eta_1 + \eta_2)k_{\rm m}^2, \tag{18}$$

where $k_{\rm m} = \mu_0\langle\overline{\boldsymbol{J}}\cdot\overline{\boldsymbol{B}}\rangle/\langle\overline{\boldsymbol{B}}^2\rangle$ is the effective wavenumber of the mean field. We note however that the use of λ is only permissible because $\overline{\boldsymbol{B}}^2$ and $\overline{\boldsymbol{J}}\cdot\overline{\boldsymbol{B}}$ are spatially uniform for α^2 dynamos in a periodic domain. For a forcing function with positive helicity we have $k_{\rm f} > 0$, and so $k_{\rm m} < 0$. Moreover, for fully helical mean fields we have $k_{\rm m} = -k_1$. In the saturated state, the growth rate must be zero, which means then that $\alpha_1 + \alpha_2$ must become comparable to $(\eta + \eta_1 + \eta_2)k_{\rm m}$.

The occurrence of the $\hat{\overline{B}}_i\hat{\overline{B}}_j$ term in (15) and the negative sign of α_2/α_1 have been confirmed independently by observing that the velocity field of a saturated dynamo can itself lead to dynamo action for a passive vector field obeying a kinematic induction equation.

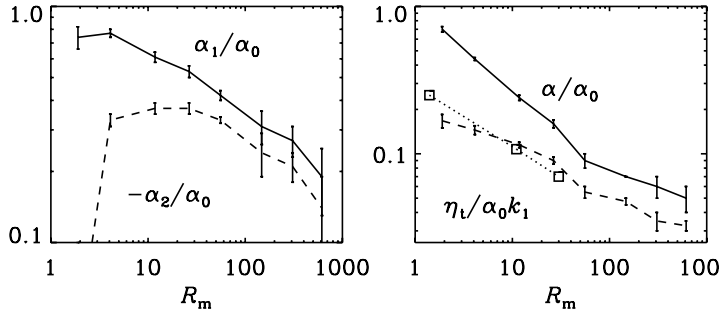


Fig. 2 $R_{\rm m}$ dependence of α_1 and $-\alpha_2$ (*left*) and of α and $\eta_t k_1$ (*right*) for equipartition-strength fields, $|\overline{\boldsymbol{B}}| = O(B_{\rm eq})$. The mutual approach of α_1 and $-\alpha_2$ illustrates how α quenching is accomplished, and the mutual approach of α and $\eta_t k_1$ illustrates by how much the quenching has to proceed

Such an observation was first made by Cattaneo and Tobias (2008) in a convection-driven small-scale dynamo and later by Tilgner and Brandenburg (2008) for the Roberts (1972) flow dynamo, where $\boldsymbol{u} = k_{\rm f}\psi\hat{\boldsymbol{z}} + \boldsymbol{\nabla}\times\psi\hat{\boldsymbol{z}}$ with $\psi = (u_0/k_1)\cos k_1 x \cos k_1 y$ and $k_{\rm f} = \sqrt{2}k_1$. As in the case of helical isotropic turbulence in a triply-periodic domain the solutions for $\overline{\boldsymbol{B}}$ are also here Beltrami fields of the form $\overline{\boldsymbol{B}} = (\cos k_1 z, \sin k_1 z, 0)$, where k_0 is the horizontal wavenumber of the helices of the Roberts flow.

The resulting matrix $\overline{B}_i\overline{B}_j$ has eigenvalues 1 and 0. In the saturated state, the eigenfunction corresponding to eigenvalue 0 is $\tilde{\overline{\boldsymbol{B}}} = (\sin k_1 z, -\cos k_1 z, 0)$ and has the growth rate $\lambda = \alpha_1(\overline{\boldsymbol{B}})k_1 - [\eta_1(\overline{\boldsymbol{B}}) + \eta]k_1^2$, which is positive, even after $\overline{\boldsymbol{B}}$ has reached saturation. This corresponds to continued exponential growth of $\tilde{\overline{\boldsymbol{B}}}$, which confirms the original finding based on the test-field method.

The results obtained using the test-field methods should of course be of predictive value to be useful. The application to a passive vector field discussed above is one example where the result for the full nonlinear α tensor was used to predict the evolution of the passive vector field. Another example is the case of rigidly rotating convection. Using the test-field method, Käpylä et al. (2008b) noticed that with increasing rotation rate α increases and $\eta_{\rm t}$ decreases. This led to the prediction that there should be α^2 dynamo action (i.e. without any shear!) for sufficiently rapid rotation. This was later confirmed using direct simulations (Käpylä et al. 2008c).

5 Three Paradigm Shifts Revisited

Let us now turn attention to the Sun. Solar dynamo theory has experienced arguably three major paradigm shifts since its broad initial acceptance during the 1970s. Inevitably, these paradigm shifts have brought the modelling further away from the original ideas that were based on dynamo theory. At the same time solar dynamo theory has lost much of its initial rigor that dynamo theory used to be based on, i.e. the profiles of α and $\eta_{\rm t}$ are no longer calculated, but are considered freely adjustable. The same is true of the magnetic quenching properties of these profiles. It its therefore important that the motivation for such departures from the original theory are well justified. In the following we discuss and comment on each of the three paradigm shifts.

5.1 Magnetic Buoyancy: from Distributed Dynamos to the Overshoot Layer

In an influential paper by Spiegel and Weiss (1980), a number of different aspects led to the suggestion that the solar dynamo operates at the base of the convection zone. One of the arguments concerned the rapid rise of magnetic flux tubes from the bulk of the convection zone. Subsequent simulations, however, have demonstrated a strongly opposing effect due to turbulent magnetic pumping (Brandenburg and Tuominen 1991; Nordlund et al. 1992; Brandenburg et al. 1996; Tobias et al. 1998). It appears, therefore, that magnetic buoyancy might not constitute a problem for the dynamo, even though its effects are clearly visible in regions where the field is strong. An example is Fig. 10 of Brandenburg et al. (1996), where the strongest tube is just "hovering" at the same height in a balance between magnetic buoyancy and downward pumping.

5.2 Helioseismology: Overshoot Layer and Flux-Transport Dynamos

The idea of dynamos operating in the overshoot layer was soon reinforced when it became evident that in the bulk of the convection zone the *radial* differential rotation, which is

important for the mean-field dynamo, is small. At the time, the strongest shear was believed to occur at the bottom of the convection zone. The positive value of the radial differential rotation in this layer, which is now called the tachocline, together with an α effect of opposite sign relative to what it is in the bulk of the convection zone, could explain the equatorward migration of the sunspot belts (DeLuca and Gilman 1986, 1988; Rüdiger and Brandenburg 1995). However, as with all models that have a positive radial angular velocity gradient, also these models have the wrong phase relation, i.e. the radial and toroidal mean fields are in phase and not in antiphase, as observed (Yoshimura 1976; Stix 1976). However, the phase relation may not pose a serious problem (Schüssler 2005).

Another possibility is that the dynamo could operate with spatially disjoint induction layers: an α effect with the usual sign near the surface, and positive radial shear at the bottom of the convection zone, coupled by meridional circulation. This led to the now popular idea of flux-transport dynamos where the meridional circulation is chiefly responsible for the equatorward migration of the toroidal flux belts (see Dikpati and Gilman 2009). However, in recent years it became clear that in the outer 5% of the Sun by radius there is strong negative radial shear (Benevolenskaya et al. 1999), which could in principle also explain the equatorward migration in the framework of conventional solar dynamo theory (B05). On the other hand, such a theory also faces problems of its own, for example the latitudinal width of the flux belts is expected to be only a few times bigger than the depths of the supergranulation layer (Brandenburg and Käpylä 2007), which would be too small.

5.3 Catastrophic Quenching: Interface and Flux-Transport Dynamos

The possibility of catastrophic quenching led Parker (1993) to propose the so-called interface dynamo where the magnetic field would be weak in the bulk of the convection zone, so as to avoid catastrophic quenching. However, as discussed in the present paper, catastrophic quenching is *always* a serious possibility, even for interface dynamo, which means that magnetic helicity fluxes are needed to alleviate it. One might well imagine that it is easier to shed magnetic helicity when the dynamo operates closer to the surface. Such models have not yet been investigated in sufficient detail.

In conclusion, there are now reasons to believe that all three paradigm shifts are problematic and may need to be reconsidered. An alternative proposal would be that the solar dynamo operates in the bulk of the convection zone, just as anticipated originally in the 1970s, and that the near-surface shear layer may play an important role in shaping the solar dynamo wave (B05).

6 Implications and Open Problems

In future work it will be important to improve our understanding of the solar dynamo, in particular its location within the Sun, its 22 year period, and the origin of the equatorward migration of the sunspot belts. As discussed in the previous section, current thinking places the solar dynamo in the tachocline, i.e. the bottom of the convection zone where the internal angular velocity turns from nearly uniform in the interior to non-uniform in the convection zone. The idea is that the field strength there exceeds the equipartition value by a factor of 100 (D'Silva and Choudhuri 1993), but such a strong field has not yet been reported based on turbulent three-dimensional dynamo simulations. Observationally not much can be said yet, because such fields would be below current helioseismological detection limits. On the theoretical side, a serious problem is that one assumes a turbulent magnetic Prandtl number

of 100, instead of 1, which is predicted by theory and simulations (Yousef et al. 2003). Such considerations neglect however the turbulent viscosity associated with the Maxwell stress of small-scale magnetic fields. Clearly, any *ad hoc* modifications of the theory are the result of trying to make the models reproduce the observations. However, at the same time such models ignore some important findings regarding the nonlinear behavior of the mean-field dynamo effect at large magnetic Reynolds numbers. Recent research has provided new detailed insights that should be followed up using more realistic settings such as spherical shell geometry.

There are several mechanisms proposed for explaining the cause of the equatorward migration of magnetic activity belts at low solar latitudes. Is it the rather feeble meridional circulation, as assumed in the now popular flux transport models (Dikpati and Charbonneau 1999), even though one has to assume unrealistic values of the *turbulent* magnetic Prandtl number, or is it perhaps the near-surface shear layer, which would have indeed the right sign, as emphasized in B05. To clarify things, future research may proceed along two parallel strands; one is connected with the development and exploitation of models in spherical geometry, and the other one is connected with unresolved problems that can be addressed in Cartesian configurations. In the following we list detailed steps of future research.

Catastrophic quenching in spherical shells. Catastrophic quenching behavior has still not yet been demonstrated convincingly in closed spheres or spherical shell sectors using, e.g., perfectly conducting boundary conditions and forced turbulence. Some work in this direction has already been done (Brandenburg et al. 2007; Mitra et al. 2008b), but the resolution is limited and the results not yet entirely conclusive.

Testfield method in spherical geometry. The test-field method needs to be re-examined in spherical coordinates. Originally the test-field method was developed in connection with full spheres, and then the test fields consisted of field components of constant value or constant slope. However, only afterwards it became clear that the scale (or wavenumber) of the field components must be the same for one set of all tensor components, and so it is necessary to work with spherical harmonic functions as test fields. In other words, constant and linearly varying field components are problematic.

Dynamo in open shells with and without shear. To verify our understanding of the saturation process of large-scale dynamos it is important to calculate, at different magnetic Reynolds numbers, the late stages of magnetic field evolution with open boundary conditions in spherical shells or shell sectors with and without shear. One expects low saturation amplitudes with energies of the mean magnetic field being inversely proportional to the magnetic Reynolds number in the absence of shear, but of order unity in the presence of shear. The shear is here critical, because shear is responsible for the local driving of small-scale magnetic helicity fluxes (Vishniac and Cho 2001; Subramanian and Brandenburg 2004, 2006).

Alpha effect from convection. The calculation of the α effect in convective turbulence is at the moment unclear. For unstratified convection with an imposed field Cattaneo and Hughes (2006) find that α diminishes for large magnetic Reynolds numbers, even for kinematically weak magnetic fields. With stratification, on the other hand, Käpylä et al. (2008b) find values of α that are compatible with those from simple estimates. They used the test-field method while Cattaneo and Hughes (2006) used an imposed field and estimate α as the ratio between the resulting field-aligned electromotive force and the imposed field. However, at large magnetic Reynolds number there is dynamo action producing also a mean field that might exceed the imposed field and thereby modify the estimate for α. Another possible reason for the discrepancy could be related to the presence or absence of stratification, because α is expected to be proportional to the local gradient of density and turbulent velocity

(Steenbeck et al. 1966). In unstratified Boussinesq convection the density is constant and the turbulent velocity only changes near boundaries. However, boundary effects could contribute to driving an α effect (Giesecke et al. 2005). Another problem could be poor scale separation, in which case the electromotive force is not just proportional to α and it becomes mandatory to use the integral kernel formulation instead (Brandenburg et al. 2008b).

Convective dynamos in spherical shells are now widely studied (Brun et al. 2004; Browning et al. 2006; Brown et al. 2007). It would be useful to compare the resulting magnetic fields with corresponding forced turbulence simulations in spherical shells and see whether contact can be made with improved mean-field models. This may require careful considerations of the scale-dependence of the turbulent transport coefficients.

Dynamos driven by magnetic instabilities. There is now quite a number of studies looking at possibilities where the flows driving the dynamo are due to the resulting magnetic field itself, and are driven by magnetic instabilities. Examples include magnetic buoyancy instabilities and the magneto-rotational instability. For example, the turbulence in accretion discs is believed to be driven by the magnetorotational instability. This was one of the first examples showing cyclic dynamo action somewhat reminiscent of the solar dynamo (Brandenburg et al. 1995), and it was believed to be a prototype of magnetically driven dynamos (Brandenburg and Schmitt 1998; Rüdiger and Pipin 2000; Rüdiger et al. 2001; Blackman and Field 2004). In the mean time, another example of a magnetically driven dynamo has emerged, where magnetic buoyancy works in the presence of shear and stratification alone (Brummell et al. 2002; Cline et al. 2003a, 2003b; Cattaneo et al. 2006). This phenomenon may be superficially similar to a magnetically dominated version of the shear–current effect (Rogachevskii and Kleeorin 2003, 2004). With the test-field method one is now in a good position to identify the governing mechanism by determining all components of the α and η_t tensors.

Magnetic flux concentrations near the surface. In the conventional picture, active regions and sunspots are thought to emerge as a result of magnetic flux tubes breaking through the surface. Given that it is difficult to imagine such tubes rising unharmed all the way from the bottom of the convection zone over so many pressure scale heights, one must test alternative scenarios in which the emergence of active regions and sunspots can be explained as the result of flux concentrations from local dynamo action via negative turbulent magnetic pressure effects (Kleeorin and Rogachevskii 1994) or turbulent flux collapse (Kitchatinov and Mazur 2000). Clearly, the underlying effects need to be established numerically and corresponding mean-field models need to be solved to make direct contact with simulations.

CME-like features above the surface. Given that virtually all successful large-scale dynamos at large magnetic Reynolds numbers are now believed to shed small-scale magnetic helicity, it is important to analyze the nature of the expelled magnetic field in simulations that couple to a simplified version of the lower solar wind. It is possible that the magnetic field above the surface and in the lower part of the solar wind might resemble coronal mass ejections (CMEs), in which case more detailed comparisons with actual coronal mass ejections would be beneficial.

Solar cycle forecast. Among the popular applications of solar dynamo theory and solar magnetohydrodynamics are solar cycle predictions, solar subsurface weather, and space weather. Also of interest are predictions of solar activity during its first 500 thousand years. This has great relevance for predicting the loss of volatile elements from the Earth's atmosphere, for example, and for understanding the conditions on Earth during the time when life began colonizing the planet. In this connection it is important to calculate the deflection of cosmic ray particles by the Sun's magnetic field and on the scale of the galaxy which is relevant for galactic cosmic rays (Svensmark 2007a, 2007b). However, such studies would

not be very meaningful unless some of the earlier projects in this list have resulted in a solar dynamo model that is trustworthy from a theoretical and a practical viewpoint.

Applications to laboratory liquid sodium dynamos. Unexpected beneficial insights have come from recent laboratory dynamo experiments. Unlike numerical dynamos, experimental liquid metal dynamos are able to address the regime of rather low values of the magnetic Prandtl number of the order of 10^{-5}, which is interesting in connection with solar and stellar conditions. At the same time the magnetic Reynolds number can be large enough (above 100) to allow for dynamo action. The Cadarache experiment is particularly interesting. Simulations of such a flow have been attempted by various groups using the Taylor-Green flow as a model (Ponty et al. 2004, 2005; Mininni et al. 2005; Brandenburg and Käpylä 2007). Again, the nature of the resulting dynamo effect has not yet been elucidated. It would be useful to analyze the resulting flows using the test-field method. One may hope that such work can teach us important lessons about large-scale and small-scale dynamos at low magnetic Prandtl number (Schekochihin et al. 2005; Iskakov et al. 2007), which is relevant to the Sun, but hard to address numerically with the currently available computing capabilities. Another relevant application is precession-driven dynamos (Tilgner 1999), where it might be useful to consider this process for a range of different geometries.

7 Conclusions

Looking back at some of the problems that dynamo theory was facing during the early years, we can say that a good deal of them have now been solved. For example the issue of turbulent magnetic diffusivity at large magnetic Reynolds numbers has now been addressed rather convincingly for values of $R_{\rm m}$ up to 200. Such a result has only recently become possible with the development of the test-field method. At this point we have no evidence that this result may change for larger values of $R_{\rm m}$. Similar statements can be made about α, where it is now reasonably clear that in the *kinematic* regime α approaches a constant value for $1 \leq R_{\rm m} \leq 200$. It should be emphasized that these results hold for forced turbulence and one must expect them to be different in cases of naturally forced turbulence such as convection or flows driven by magnetic instabilities such as the magneto-rotational instability (Brandenburg 2008) or the magnetic buoyancy instability (Brandenburg and Schmitt 1998; Thelen 2000).

Much larger values of $R_{\rm m}$ of 2×10^5 have been obtained for the special case of the Galloway-Proctor flow for which α shows irregular sign changes with $R_{\rm m}$ (Courvoisier et al. 2006). This flow is a time-dependent version of the Roberts flow where the pattern wobbles in the plane with given amplitude and frequency. Expressions of the form (11) do not apply in this case where the correlation time is infinite (Rädler and Brandenburg 2009). In that sense the Galloway-Proctor flow is quite different from a turbulent flow. Asymptotic behavior for large $R_{\rm m}$ is only possible for sufficiently large amplitude and/or frequency of the wobbling motion.

In the nonlinear case equally dramatic progress has been made in just the past few years. While it has long been clear that in closed domains α will be quenched down to values that depend on the quenched value of $\eta_{\rm t}$ and on the effective wavenumber of the mean field, it remained unclear what the quenched value of $\eta_{\rm t}$ is. Recent evidence points to a suppression by a factor of 5 when $R_{\rm m}$ is increased from 2 to 600 (Brandenburg et al. 2008c). However, this value may depend on circumstances and could be slightly less strong in the presence of shear (Käpylä and Brandenburg 2008).

In open domains there is the possibility that the resulting magnetic field strength can still decrease to catastrophically quenched values unless there is a finite divergence of the magnetic helicity flux (Brandenburg and Subramanian 2005a). Such a flux can be driven efficiently in the presence of shear. In order for this mechanism to operate, the contours of constant shear velocity must cross the boundaries (KKB08, Hughes and Proctor 2009), which explains the lack of large-scale fields in simulations with horizontal shear and periodic boundary conditions in that direction (Tobias et al. 2008).

There is clearly a long way to go before the solar dynamo problem can be addressed in full. There is hardly any doubt that the inclusion of magnetic helicity fluxes will be important, but the precise functional form of the magnetic helicity flux needs to be confirmed numerically. In particular, the possible dependencies of the fluxes on $\overline{\boldsymbol{B}}$ and $R_{\rm m}$ are not well understood at present.

Acknowledgements I would like to acknowledge Sasha Kosovichev and the other members of the team on Subphotospheric Dynamics of the Sun at the International Space Science Institute in Bern for providing an inspiring atmosphere. This work was supported in part by the Swedish Research Council.

References

E.E. Benevolenskaya, J.T. Hoeksema, A.G. Kosovichev, P.H. Scherrer, The interaction of new and old magnetic fluxes at the beginning of solar cycle 23. Astrophys. J. **517**, L163–L166 (1999)
E.G. Blackman, A. Brandenburg, Dynamic nonlinearity in large scale dynamos with shear. Astrophys. J. **579**, 359–373 (2002)
E.G. Blackman, G.B. Field, A simple mean field approach to turbulent transport. Phys. Fluids **15**, L73–L76 (2003)
E.G. Blackman, G.B. Field, Dynamical magnetic relaxation: A nonlinear magnetically driven dynamo. Phys. Plasmas **11**, 3264–3269 (2004)
A. Brandenburg, The inverse cascade and nonlinear alpha-effect in simulations of isotropic helical hydromagnetic turbulence. Astrophys. J. **550**, 824–840 (2001)
A. Brandenburg, The case for a distributed solar dynamo shaped by near-surface shear. Astrophys. J. **625**, 539–547 (2005)
A. Brandenburg, The dual role of shear in large-scale dynamos. Astron. Nachr. **329**, 725–731 (2008)
A. Brandenburg, W. Dobler, Large scale dynamos with helicity loss through boundaries. Astron. Astrophys. **369**, 329–338 (2001)
A. Brandenburg, W. Dobler, Hydromagnetic turbulence in computer simulations. Comp. Phys. Comm. **147**, 471–475 (2002)
A. Brandenburg, P.J. Käpylä, Magnetic helicity effects in astrophysical and laboratory dynamos. New J. Phys. **9**, 305 (2007)
A. Brandenburg, D. Schmitt, Simulations of an alpha-effect due to magnetic buoyancy. Astron. Astrophys. **338**, L55–L58 (1998)
A. Brandenburg, D. Sokoloff, Local and nonlocal magnetic diffusion and alpha-effect tensors in shear flow turbulence. Geophys. Astrophys. Fluid Dyn. **96**, 319–344 (2002)
A. Brandenburg, K. Subramanian, Strong mean field dynamos require supercritical helicity fluxes. Astron. Nachr. **326**, 400–408 (2005a)
A. Brandenburg, K. Subramanian, Astrophysical magnetic fields and nonlinear dynamo theory. Phys. Rep. **417**, 1–209 (2005b)
A. Brandenburg, K. Subramanian, Minimal tau approximation and simulations of the alpha effect. Astron. Astrophys. **439**, 835–843 (2005c)
A. Brandenburg, K. Subramanian, Simulations of the anisotropic kinetic and magnetic alpha effects. Astron. Nachr. **328**, 507–512 (2007)
A. Brandenburg, I. Tuominen, The solar dynamo, in *The Sun and Cool Stars: Activity, Magnetism, Dynamos, IAU Coll. 130*, ed. by I. Tuominen, D. Moss, G. Rüdiger. Lecture Notes in Physics, vol. 380 (Springer, Berlin, 1991), pp. 223–233
A. Brandenburg, J.L. Jennings, Å. Nordlund, M. Rieutord, R.F. Stein, I. Tuominen, Magnetic structures in a dynamo simulation. J. Fluid Mech. **306**, 325–352 (1996)

A. Brandenburg, A. Bigazzi, K. Subramanian, The helicity constraint in turbulent dynamos with shear. Mon. Not. R. Astron. Soc. **325**, 685–692 (2001)
A. Brandenburg, W. Dobler, K. Subramanian, Magnetic helicity in stellar dynamos: new numerical experiments. Astron. Nachr. **323**, 99–122 (2002)
A. Brandenburg, P.J. Käpylä, D. Mitra, D. Moss, R. Tavakol, The helicity constraint in spherical shell dynamos. Astron. Nachr. **328**, 1118–1121 (2007)
A. Brandenburg, P. Käpylä, A. Mohammed, Non-Fickian diffusion and tau-approximation from numerical turbulence. Phys. Fluids **16**, 1020–1027 (2004)
A. Brandenburg, D. Moss, A.M. Soward, New results for the Herzenberg dynamo: steady and oscillatory solutions. Proc. R. Soc. A **454**, 1283–1300 (1998)
A. Brandenburg, Å. Nordlund, R.F. Stein, U. Torkelsson, Dynamo generated turbulence and large scale magnetic fields in a Keplerian shear flow. Astrophys. J. **446**, 741–754 (1995)
A. Brandenburg, K.-H. Rädler, M. Rheinhardt, P.J. Käpylä, Magnetic diffusivity tensor and dynamo effects in rotating and shearing turbulence. Astrophys. J. **676**, 740–751 (2008a)
A. Brandenburg, K.-H. Rädler, M. Schrinner, Scale dependence of alpha effect and turbulent diffusivity. Astron. Astrophys. **482**, 739–746 (2008b)
A. Brandenburg, K.-H. Rädler, M. Rheinhardt, K. Subramanian, Magnetic quenching of alpha and diffusivity tensors in helical turbulence. Astrophys. J. Lett. **687**, L49–L52 (2008c)
B.P. Brown, M.K. Browning, A.S. Brun, M.S. Miesch, N.J. Nelson, J. Toomre, Strong dynamo action in rapidly rotating suns. AIPC **948**, 271–278 (2007)
M.K. Browning, M.S. Miesch, A.S. Brun, J. Toomre, Dynamo action in the solar convection zone and tachocline: pumping and organization of toroidal fields. Astrophys. J. **648**, L157–L160 (2006)
N. Brummell, K. Cline, F. Cattaneo, Formation of buoyant magnetic structures by a localized velocity shear. Mon. Not. R. Astron. Soc. **329**, L73–L76 (2002)
A.S. Brun, M.S. Miesch, J. Toomre, Global-scale turbulent convection and magnetic dynamo action in the solar envelope. Astrophys. J. **614**, 1073–1098 (2004)
F. Cattaneo, D.W. Hughes, Nonlinear saturation of the turbulent alpha effect. Phys. Rev. E **54**, R4532–R4535 (1996)
F. Cattaneo, D.W. Hughes, Dynamo action in a rotating convective layer. J. Fluid Mech. **553**, 401–418 (2006)
F. Cattaneo, D.W. Hughes, Problems with kinematic mean field electrodynamics at high magnetic Reynolds numbers (2008). arXiv:0805.2138
F. Cattaneo, S.M. Tobias, How do dynamos saturate? J. Fluid Mech. (2008). arXiv:0809.1801. See also the talk given at the Kavli Institute for Theoretical Physics "Large and small-scale dynamo action" (http://online.kitp.ucsb.edu/online/dynamo_c08/cattaneo)
F. Cattaneo, S.I. Vainshtein, Suppression of turbulent transport by a weak magnetic field. Astrophys. J. Lett. **376**, L21–L24 (1991)
F. Cattaneo, N.H. Brummell, K.S. Cline, What is a flux tube? On the magnetic field topology of buoyant flux structures. Mon. Not. R. Astron. Soc. **365**, 727–734 (2006)
S. Childress, Alpha-effect in flux ropes and sheets. Phys. Earth Planet. Int. **20**, 172–180 (1979)
K.S. Cline, N.H. Brummell, F. Cattaneo, On the formation of magnetic structures by the combined action of velocity shear and magnetic buoyancy. Astrophys. J. **588**, 630–644 (2003a)
K.S. Cline, N.H. Brummell, F. Cattaneo, Dynamo action driven by shear and magnetic buoyancy. Astrophys. J. **599**, 1449–1468 (2003b)
A. Courvoisier, D.W. Hughes, S.M. Tobias, α-effect in a family of chaotic flows. Phys. Rev. Lett. **96**, 034503 (2006)
T.G. Cowling, The magnetic field of sunspots. Mon. Not. R. Astron. Soc. **94**, 39–48 (1933)
E.E. DeLuca, P.A. Gilman, Dynamo theory for the interface between the convection zone and the radiative interior of a star. Part I. Model Equations and exact solutions. Geophys. Astrophys. Fluid Dyn. **37**, 85–127 (1986)
E.E. DeLuca, P.A. Gilman, Dynamo theory for the interface between the convection zone and the radiative interior of a star. Part II. Numerical solutions of the nonlinear equations. Geophys. Astrophys. Fluid Dyn. **43**, 119–148 (1988)
M. Dikpati, P. Charbonneau, A Babcock-Leighton flux transport dynamo with solar-like differential rotation. Astrophys. J. **518**, 508–520 (1999)
M. Dikpati, P. Gilman, Flux transport solar dynamos. Space Sci. Rev. (2009, this issue)
S. D'Silva, A.R. Choudhuri, A theoretical model for tilts of bipolar magnetic regions. Astron. Astrophys. **272**, 621–633 (1993)
G.B. Field, E.G. Blackman, Dynamical quenching of the α^2 dynamo. Astrophys. J. **572**, 685–692 (2002)
U. Frisch, A. Pouquet, J. Léorat, A. Mazure, Possibility of an inverse cascade of magnetic helicity in hydrodynamic turbulence. J. Fluid Mech. **68**, 769–778 (1975)

A. Giesecke, U. Ziegler, G. Rüdiger, Geodynamo α-effect derived from box simulations of rotating magnetoconvection. Phys. Earth Planet. Int. **152**, 90–102 (2005)
A.V. Gruzinov, P.H. Diamond, Self-consistent theory of mean-field electrodynamics. Phys. Rev. Lett. **72**, 1651–1653 (1994)
A.V. Gruzinov, P.H. Diamond, Self-consistent mean field electrodynamics of turbulent dynamos. Phys. Plasmas **2**, 1941–1947 (1995)
A. Herzenberg, Geomagnetic dynamos. Proc. R. Soc. Lond. **250A**, 543–583 (1958)
A. Hubbard, A. Brandenburg, Memory effects in turbulent transport, Astrophys. J. (2008). arXiv:0811.2561
D.W. Hughes, M.R.E. Proctor, Large-scale dynamo action driven by velocity shear and rotating convection. Phys. Rev. Lett. **102**, 044501 (2009)
A.B. Iskakov, A.A. Schekochihin, S.C. Cowley, J.C. McWilliams, M.R.E. Proctor, Numerical demonstration of fluctuation dynamo at low magnetic Prandtl numbers. Phys. Rev. Lett. **98**, 208501 (2007)
P.J. Käpylä, A. Brandenburg, Turbulent dynamos with shear and fractional helicity. Astrophys. J. (2008, submitted). arXiv:0810.2298
P.J. Käpylä, M.J. Korpi, A. Brandenburg, Large-scale dynamos in turbulent convection with shear. Astron. Astrophys. **491**, 353–362 (2008a)
P.J. Käpylä, M.J. Korpi, A. Brandenburg, Alpha effect and turbulent diffusion from convection. Astron. Astrophys (2008b). arXiv:0812.1792
P.J. Käpylä, M.J. Korpi, A. Brandenburg, Large-scale dynamos in rigidly rotating turbulent convection. Astrophys. J. (2008c). arXiv:0812.3958
L.L. Kitchatinov, M.V. Mazur, Stability and equilibrium of emerged magnetic flux. Sol. Phys. **191**, 325–340 (2000)
N. Kleeorin, I. Rogachevskii, Effective Ampère force in developed magnetohydrodynamic turbulence. Phys. Rev. **50**, 2716–2730 (1994)
N. Kleeorin, D. Moss, I. Rogachevskii, D. Sokoloff, Helicity balance and steady-state strength of the dynamo generated galactic magnetic field. Astron. Astrophys. **361**, L5–L8 (2000)
E. Knobloch, Turbulent diffusion of magnetic fields. Astrophys. J. **225**, 1050–1057 (1978)
F. Krause, The cosmic dynamo: from $t = -\infty$ to Cowling's theorem. A review on history, in *The Cosmic Dynamo*, ed. by F. Krause, K.-H. Rädler, G. Rüdiger (Kluwer, Dordrecht, 1993), pp. 487–499
J. Larmor, How could a rotating body such as the Sun become a magnet. Rep. Brit. Assoc. Adv. Sci. **159** (1919)
J. Larmor, The magnetic field of sunspots. Mon. Not. R. Astron. Soc. **94**, 469–471 (1934)
D. Layzer, R. Rosner, H.T. Doyle, On the origin of solar magnetic fields. Astrophys. J. **229**, 1126–1137 (1979)
F.J. Lowes, I. Wilkinson, Geomagnetic dynamo: a laboratory model. Nature **198**, 1158–1160 (1963)
F.J. Lowes, I. Wilkinson, Geomagnetic dynamo: an improved laboratory model. Nature **219**, 717–718 (1968)
J. Maron, E.G. Blackman, Effect of fractional kinetic helicity on turbulent magnetic dynamo spectra. Astrophys. J. Lett. **566**, L41–L44 (2002)
P.D. Mininni, Y. Ponty, D.C. Montgomery, J.-F. Pinton, H. Politano, A. Pouquet, Dynamo regimes with a nonhelical forcing. Astrophys. J. **853**, 853–863 (2005)
D. Mitra, P.J. Käpylä, R. Tavakol, A. Brandenburg, Alpha effect and diffusivity in helical turbulence with shear. Astron. Astrophys. (2008a, in press). arXiv:0806.1608
D. Mitra, R. Tavakol, A. Brandenburg, D. Moss, Turbulent dynamos in spherical shell segments of varying geometrical extent. Astrophys. J. (2008b, submitted). arXiv:0812.3106
Å. Nordlund, A. Brandenburg, R.L. Jennings, M. Rieutord, J. Ruokolainen, R.F. Stein, I. Tuominen, Dynamo action in stratified convection with overshoot. Astrophys. J. **392**, 647–652 (1992)
E.N. Parker, Hydromagnetic dynamo models. Astrophys. J. **122**, 293–314 (1955)
E.N. Parker, A solar dynamo surface wave at the interface between convection and nonuniform rotation. Astrophys. J. **408**, 707–719 (1993)
J.H. Piddington, Turbulent diffusion of magnetic fields in astrophysical plasmas. Astrophys. J. **247**, 293–299 (1981)
Y. Ponty, H. Politano, J.-F. Pinton, Simulation of induction at low magnetic Prandtl number. Phys. Rev. Lett. **92**, 144503 (2004)
Y. Ponty, P.D. Mininni, D.C. Montgomery, J.-F. Pinton, H. Politano, A. Pouquet, Numerical study of dynamo action at low magnetic Prandtl numbers. Phys. Rev. Lett. **94**, 164502 (2005)
K.-H. Rädler, A. Brandenburg, Mean-field effects in the Galloway-Proctor flow. Mon. Not. R. Astron. Soc. **393**, 113–125 (2009)
K.-H. Rädler, M. Rheinhardt, Mean-field electrodynamics: critical analysis of various analytical approaches to the mean electromotive force. Geophys. Astrophys. Fluid Dyn. **101**, 11–48 (2007)
K.-H. Rädler, M. Rheinhardt, E. Apstein, H. Fuchs, On the mean-field theory of the Karlsruhe dynamo experiment. Nonlinear Processes Geophys. **38**, 171–187 (2002)

G.O. Roberts, Dynamo action of fluid motions with two-dimensional periodicity. Phil. Trans. R. Soc. **A271**, 411–454 (1972)
I. Rogachevskii, N. Kleeorin, Electromotive force and large-scale magnetic dynamo in a turbulent flow with a mean shear. Phys. Rev. **68**, 036301 (2003)
I. Rogachevskii, N. Kleeorin, Nonlinear theory of a 'shear–current' effect and mean-field magnetic dynamos. Phys. Rev. **70**, 046310 (2004)
G. Rüdiger, A. Brandenburg, A solar dynamo in the overshoot layer: cycle period and butterfly diagram. Astron. Astrophys. **296**, 557–566 (1995)
G. Rüdiger, V.V. Pipin, Viscosity-alpha and dynamo-alpha for magnetically driven compressible turbulence in Kepler disks. Astron. Astrophys. **362**, 756–761 (2000)
G. Rüdiger, V.V. Pipin, G. Belvedère, Alpha-effect, helicity and angular momentum transport for a magnetically driven turbulence in the solar convection zone. Sol. Phys. **198**, 241–251 (2001)
A.A. Schekochihin, N.E.L. Haugen, A. Brandenburg, S.C. Cowley, J.L. Maron, J.C. McWilliams, Onset of small-scale dynamo at small magnetic Prandtl numbers. Astrophys. J. **625**, L115–L118 (2005)
M. Schrinner, K.-H. Rädler, D. Schmitt, M. Rheinhardt, U. Christensen, Mean-field view on rotating magnetoconvection and a geodynamo model. Astron. Nachr. **326**, 245–249 (2005)
M. Schrinner, K.-H. Rädler, D. Schmitt, M. Rheinhardt, U.R. Christensen, Mean-field concept and direct numerical simulations of rotating magnetoconvection and the geodynamo. Geophys. Astrophys. Fluid Dyn. **101**, 81–116 (2007)
M. Schüssler, Is there a phase constraint for solar dynamo models? Astron. Astrophys. **439**, 749–750 (2005)
E.A. Spiegel, N.O. Weiss, Magnetic activity and variation in the solar luminosity. Nature **287**, 616–617 (1980)
M. Steenbeck, F. Krause, Zur Dynamotheorie stellarer und planetarer Magnetfelder I. Berechnung sonnenähnlicher Wechselfeldgeneratoren. Astron. Nachr. **291**, 49–84 (1969)
M. Steenbeck, F. Krause, K.-H. Rädler, Berechnung der mittleren Lorentz-Feldstärke $\overline{\boldsymbol{v} \times \boldsymbol{B}}$ für ein elektrisch leitendendes Medium in turbulenter, durch Coriolis-Kräfte beeinflußter Bewegung. Z. Naturforsch. **21a**, 369–376 (1966). See also the translation in Roberts & Stix, The turbulent dynamo, Tech. Note 60, NCAR, Boulder, Colorado (1971)
M. Stix, Differential rotation and the solar dynamo. Astron. Astrophys. **47**, 243–254 (1976)
K. Subramanian, A. Brandenburg, Nonlinear current helicity fluxes in turbulent dynamos and alpha quenching. Phys. Rev. Lett. **93**, 205001 (2004)
K. Subramanian, A. Brandenburg, Magnetic helicity density and its flux in weakly inhomogeneous turbulence. Astrophys. J. **648**, L71–L74 (2006)
S. Sur, K. Subramanian, A. Brandenburg, Kinetic and magnetic alpha effects in nonlinear dynamo theory. Mon. Not. R. Astron. Soc. **376**, 1238–1250 (2007)
S. Sur, A. Brandenburg, K. Subramanian, Kinematic alpha effect in isotropic turbulence simulations. Mon. Not. R. Astron. Soc. **385**, L15–L19 (2008)
H. Svensmark, Imprint of Galactic dynamics on Earth's climate. Astron. Nachr. **327**, 866–870 (2007a)
H. Svensmark, Cosmic rays and the biosphere over 4 billion years. Astron. Nachr. **327**, 871–875 (2007b)
J.-C. Thelen, A mean electromotive force induced by magnetic buoyancy instabilities. Mon. Not. R. Astron. Soc. **315**, 155–164 (2000)
A. Tilgner, Magnetohydrodynamic flow in precessing spherical shells. J. Fluid Mech. **379**, 303–318 (1999)
A. Tilgner, A. Brandenburg, A growing dynamo from a saturated Roberts flow dynamo. Mon. Not. R. Astron. Soc. **391**, 1477–1481 (2008)
S. Tobias, Relating stellar cycle periods to dynamo calculations. Mon. Not. R. Astron. Soc. **296**, 653–661 (1998)
S.M. Tobias, F. Cattaneo, N.H. Brummell, Convective dynamos with penetration, rotation, and shear. Astrophys. J. **685**, 596–605 (2008)
S.I. Vainshtein, F. Cattaneo, Nonlinear restrictions on dynamo action. Astrophys. J. **393**, 165–171 (1992)
E.T. Vishniac, J. Cho, Magnetic helicity conservation and astrophysical dynamos. Astrophys. J. **550**, 752–760 (2001)
N.O. Weiss, Linear and nonlinear dynamos. Astron. Nachr. **326**, 157–165 (2005)
H. Yoshimura, Phase relation between the poloidal and toroidal solar-cycle general magnetic fields and location of the origin of the surface magnetic fields. Sol. Phys. **50**, 3–23 (1976)
T.A. Yousef, A. Brandenburg, G. Rüdiger, Turbulent magnetic Prandtl number and magnetic diffusivity quenching from simulations. Astron. Astrophys. **411**, 321–327 (2003)

Planetary Dynamos from a Solar Perspective

U.R. Christensen · D. Schmitt · M. Rempel

Originally published in the journal Space Science Reviews, Volume 144, Nos 1–4, 105–126.
DOI: 10.1007/s11214-008-9449-6

Abstract Direct numerical simulations of the geodynamo and other planetary dynamos have been successful in reproducing the observed magnetic fields. We first give an overview on the fundamental properties of planetary magnetism. We review the concepts and main results of planetary dynamo modeling, contrasting them with the solar dynamo. In planetary dynamos the density stratification plays no major role and the magnetic Reynolds number is low enough to allow a direct simulation of the magnetic induction process using microscopic values of the magnetic diffusivity. The small-scale turbulence of the flow cannot be resolved and is suppressed by assuming a viscosity far in excess of the microscopic value. Systematic parameter studies lead to scaling laws for the magnetic field strength or the flow velocity that are independent of viscosity, indicating that the models are in the same dynamical regime as the flow in planetary cores. Helical flow in convection columns that are aligned with the rotation axis play an important role for magnetic field generation and forms the basis for a macroscopic α-effect. Depending on the importance of inertial forces relative to rotational forces, either dynamos with a dominant axial dipole or with a small-scale multipolar magnetic field are found. Earth is predicted to lie close to the transition point between both classes, which may explain why the dipole undergoes reversals. Some models fit the properties of the geomagnetic field in terms of spatial power spectra, magnetic field morphology and details of the reversal behavior remarkably well. Magnetic field strength in the dipolar dynamo regime is controlled by the available power and found to be independent of rotation rate. Predictions for the dipole moment agree well with the observed field strength of Earth and Jupiter and moderately well for other planets. Dedicated dynamo models for Mercury, Saturn, Uranus and Neptune, which assume stably stratified layers above or below the dynamo region, can explain some of the unusual field properties of these planets.

Keywords Planetary magnetic fields · Geodynamo · Dynamo models

U.R. Christensen (✉) · D. Schmitt
Max-Planck-Institut für Sonnensystemforschung, 37191 Katlenburg-Lindau, Germany
e-mail: christensen@mps.mpg.de

M. Rempel
High Altitude Observatory, National Center for Atmospheric Research, Boulder, CO 80307-3000, USA

M.J. Thompson et al. (eds.), *The Origin and Dynamics of Solar Magnetism*.
DOI: 10.1007/978-1-4419-0239-9_8

1 Introduction

Starting in 1995 numerical modeling of the Earth's dynamo has flourished with remarkable success. Direct numerical simulation of convection-driven MHD-flow in a rotating spherical shell show magnetic fields that resemble the geomagnetic field in many respects: they are dominated by the axial dipole of approximately the right strength, they show spatial power spectra similar to that of Earth, and the magnetic field morphology and the temporal variation of the field resembles that of the geomagnetic field (Christensen and Wicht 2007). Some models show stochastic dipole reversals whose details agree with what has been inferred from paleomagnetic data (Glatzmaier and Roberts 1995; Kutzner and Christensen 2002; Wicht 2005). While these models represent direct numerical simulations of the fundamental MHD equations without parameterized induction effects, they do not match actual planetary conditions in a number of respects. Specifically, they rotate too slowly, are much less turbulent, and use a viscosity and thermal diffusivity that is far too large in comparison to magnetic diffusivity. Because of these discrepancies, the success of geodynamo models may seem surprising.

In order to better understand the extent to which the models are applicable to planetary dynamos, scaling laws that relate basic properties of the dynamo to the fundamental control parameters play an important role. In recent years first attempts have been made to derive such scaling laws from a set of numerical simulations that span the accessible parameter space (Christensen and Tilgner 2004; Christensen and Aubert 2006). The extrapolation of these laws to planetary parameters gives reasonable results, which suggests that despite their shortcomings the dynamo models are already in the appropriate dynamical regime.

Most planets in the solar system have internal magnetic fields or once had such fields (Stevenson 2003). In many, but not all, cases the axial dipole dominates the field at the planetary surface. In fact, a surprising diversity is found in magnetic field strength and field morphology. A comparative dynamo theory, that explains the common features and the differences between planetary magnetic fields, is still in its infancy.

This paper intends to give a brief overview on the progress made in understanding planetary dynamos, mostly achieved with the aid of numerical simulations. We will first review the salient properties of the geomagnetic field and then proceed to what is known on magnetism of other planets. Next we present the conceptual assumptions used in modeling planetary dynamos, and we discuss some essential differences to the solar dynamo. We then describe successes and failures in modeling the geodynamo and proceed to discuss attempts to derive scaling laws that put the understanding on a more general level. We close by discussing some hypotheses and models to explain the observed magnetic properties of specific planets other than Earth.

2 Planetary Magnetic Fields

2.1 Geomagnetic Field

The properties of the recent geomagnetic field have been mapped with high spatial resolution by dedicated satellite missions in a low Earth orbit, such as MAGSAT, ØRSTED, and CHAMP (Olsen et al. 2007). The variation of the field during the last four centuries (the so-called historical field) has been recorded with somewhat lower spatial resolution by a network of magnetic observatories and by individual measurements, most of which have been performed routinely by mariners (Jackson et al. 2000). Going further back in time, information on the geomagnetic field has been retrieved from the remanent magnetization in

man-made artifacts, covering the last couple of thousand years (Korte and Constable 2005), or in natural rocks, covering geological time as far back as 3 billion years (Tarduno et al. 2007). The spatial and temporal resolution is moderate for archeomagnetic data and poor for paleomagnetic data.

Historically, a variety of hypotheses have been proposed for the origin of the geomagnetic field. The only viable model that survived the test of time is that of a dynamo process operating in the Earth's core. Seismology and high pressure research have shown that its outer part is an iron-rich metallic liquid, which contains approximately 10% of light alloying elements (such as sulphur, oxygen or silicon in unknown proportions). There is a small solid inner core that is depleted in the light elements relative to the outer core. Secular cooling of the core implies that the inner core grows by freezing iron on its surface, enriching the overlying residual liquid in the light elements. Both thermal and compositional convection drive a circulation in the outer core. Hence the most basic requirement for a dynamo, flow of an electrically conducting fluid, seems to be satisfied inside the Earth.

While the geomagnetic field is observed at or above the Earth's surface, maps of the field at the top of the iron core are more useful for interpreting field structure in terms of the underlying dynamo process. For the historical geomagnetic field such maps have been constructed under the assumption of a potential field outside the core (Gubbins and Bloxham 1985; Jackson et al. 2000). This assumption seems to be satisfied approximately for the long-wavelength part of the magnetic field up to spherical harmonic degree 13. At smaller scales the field at the Earth's surface is dominated by contributions arising from the inhomogeneous remanent magnetization of rocks in the Earth's crust. Therefore, it is not possible to determine the structure of the core field at wavelengths smaller than 2000 km.

Figure 1a shows the radial component of Earth's field at the core-mantle boundary. The axial dipole component is dominant, but higher multipoles contribute to the overall field morphology more strongly than they do at the Earth's surface. In particular, the magnetic flux contributing most prominently to the axial dipole field is concentrated in two lobes in each hemisphere at approximately 60°–70° latitude. The flux close to the rotation poles is weak or slightly inverse with respect to the dipole polarity. At low latitudes flux spots of both polarities exist.

The rms field strength at the top of Earth's core is approximately 0.4 mT (4 Gauss) in degrees up to 13. It is plausible to assume that the magnetic field B inside the core is larger by a factor between three and ten, which puts the characteristic field strength in the dynamo region into the range between one and a few mT. This means that the Elsasser number Λ

$$\Lambda = \frac{B^2}{\mu_o \eta \rho \Omega} \tag{1}$$

is of order one, where μ_o is magnetic permeability, η magnetic diffusivity, ρ density and Ω rotation rate. The Elsasser number is often taken as measure for the ratio of Lorentz forces to Coriolis forces acting on the flow. The finding that $\Lambda \approx 1$ in the Earth's core has been taken as support for the notion that planetary dynamos are in a magnetostrophic regime, where Coriolis force and Lorentz force balance to first order, and that this balance controls the strength of the magnetic field (e.g., Stevenson 2003). As we will see in Sect. 8, dynamo simulations put some doubt on this hypothesis.

The Earth's magnetic field changes significantly on time scales of one hundred years. This secular variation concerns mainly the higher multipole components. The axial dipole component is much more stable, at least in terms of polarity. Polarity changes occur stochastically in time intervals of some hundred thousand years. The duration of a reversal

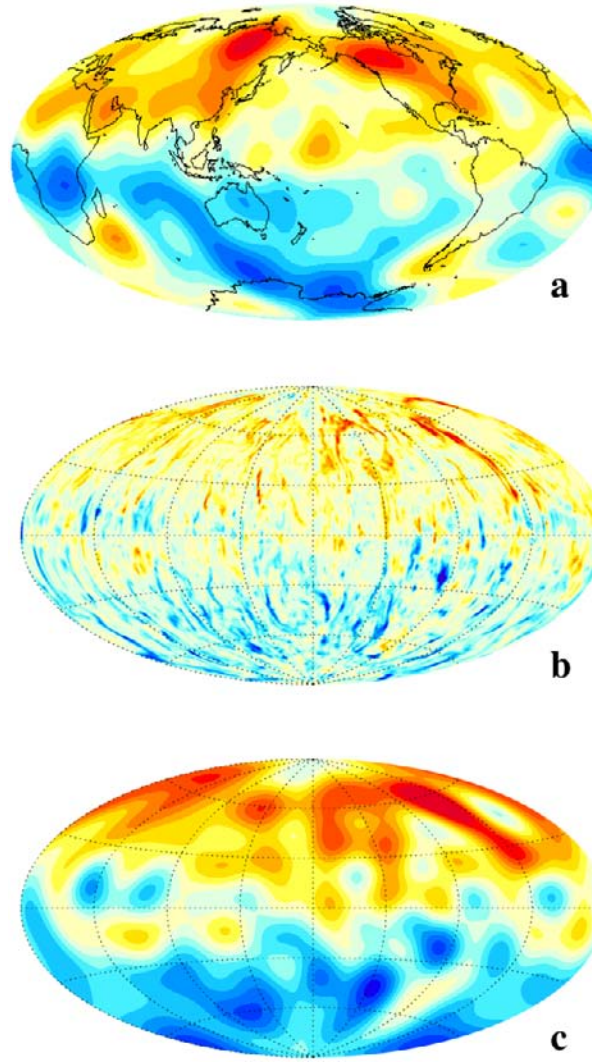


Fig. 1 **a** Radial component of the geomagnetic field at the core-mantle boundary in 1990, *red colors* for inward field, *blue colors* for outward field. **b** Snapshot of radial field from a dynamo model with parameters $E = 10^{-5}$, $Ra^* = 0.12$, $Pm = 0.8$, $Pr = 1$. **c** Same snapshot for the field low-pass filtered to spherical harmonic degrees <14

is comparatively short, on the order of a few thousand years. The frequency of reversals itself varies on time scales of 100 million years, which is the typical time scale for mantle convection. This is interpreted as reflecting the coupling of the dynamo to heterogeneities (for example in temperature) in the Earth's lower mantle (Glatzmaier et al. 1999; Kutzner and Christensen 2004; Courtillot and Olson 2007).

2.2 Other Planets

The magnetic fields of all major planets in the solar system have been characterized by space missions, although often this provided only a crude snapshot compared to our knowledge of the geomagnetic field. In particular, next to nothing is known about the time variability of other planetary fields. Table 1 gives an overview on the field properties.

Table 1 Planetary magnetic fields

Planet	Active dynamo	Past dynamo	r_c/r_p	B_s[nT]	Structure
Mercury	Yes ?		0.75	400	dipolar
Venus	No	?	0.55		
Earth	Yes	Yes	0.55	50, 000	dipolar
Moon	No	Yes ?	0.2 ?		
Mars	No	Yes	0.5		
Jupiter	Yes		0.85	500, 000	dipolar
Ganymede	Yes		0.3 ?	1, 000	dipolar
Saturn	Yes		0.5	30, 000	dipolar
Uranus	Yes		0.7	45, 000	multipolar
Neptune	Yes		0.7	45, 000	multipolar

Probably all large planetary bodies have an electrically conducting fluid core (or, at least had a fluid core at some time in their history) that can potentially sustain a dynamo. However, the nature of this core region is quite different between the various groups of planets. In the terrestrial planets it consists of a molten iron-rich alloy. In Jupiter and Saturn it is a metallic high-pressure form of hydrogen and the deep interior of Uranus and Neptune consists of a water-rich fluid with ionic conductivity. In Jupiter's satellite Ganymede, which has a magnetic field of its own, the dynamo probably resides in a deep liquid iron core, although dynamo action in a salty water ocean at shallower depth can perhaps not be excluded entirely. The values of the core radius r_c relative to the planetary radius r_p are given in Table 1.

Not all large planetary bodies have a magnetic field at present, although in some cases (Moon, Mars) local magnetic fields arise from remanent magnetization of minerals in the planet's crust (akin to so-called magnetic anomalies at Earth). The most likely explanation of how the magnetization was acquired is that an internal dynamo generated a strong magnetic field early in the planet's history.

In most cases the magnetic field at the planetary surface is dominated by a dipole that is nearly aligned with the rotation axis. As in case of the Earth higher multipoles contribute on the order of 10% to the fields at the surface of Jupiter and Saturn. Mercury's field morphology has not yet been characterized sufficiently to clearly quantify the non-dipole contributions. A particular property of Saturn's magnetic field not found at other planets is an extremely high degree of axisymmetry including a dipole tilt relative to the rotation axis that is indistinguishable from zero. In the case of Uranus and Neptune, the quadrupole and octupole contribute to the surface field at a similar level as the dipole and the latter is strongly tilted against the rotation axis. When continued downward to radius r_c the dipole still remains the dominant field component in most planets, but becomes relatively weaker than higher multipoles in the cases of Uranus and Neptune. Therefore, the field of the latter two planets is characterized as 'multipolar', in Table 1.

The magnetic field strength at the planetary surface, B_s, covers a wide range for the different planets, from ten times the strength of the geomagnetic field at Jupiter to 1% of the Earth's field strength in the case of Mercury. When downward continuing the magnetic field to r_c the differences in field strength are slightly reduced but remain large between the end members.

In summary, despite some similarities, large differences in field strength and structure exist. Theory must strive at explaining them either in terms of particular conditions in the

planet, on a case-by-case basis, or by establishing laws describing a systematic dependence of the dynamo properties on the essential controlling parameters. Which these parameters are is not entirely clear a-priori, but size, rotation rate, electrical conductivity and vigor of convection are some obvious candidates. To explain the diversity of magnetism in the planetary system, a combination of these two approaches seems to be necessary.

3 Solar Versus Planetary Dynamos

In contrast to the dipole-dominated geomagnetic field, which undergoes stochastic reversals, the solar magnetic field is dominated by small scales, yet it shows a high degree of regularity. This is expressed in the eleven-year activity cycle, the systematic latitudinal migration of the emergence region of sunspots and the east-west alignment of bipolar active regions that follow Hale's polarity rules. These differences in magnetic field behavior point at profound differences in the underlying dynamo process. In this section we discuss some essential differences between the solar dynamo and planetary dynamos and their implications for the modeling of planetary dynamos.

3.1 Energetics

While strong convection is a matter of fact in the outer parts of the Sun, the occurrence of sufficient flow that can drive a dynamo is less clear for planetary cores, in particular in the case of the terrestrial planets. The energy flux in planets is obviously much weaker and results from the slow secular cooling that releases internal energy (of gravitational origin) acquired during accretion. This may be augmented by heat due to the decay of radioactive trace elements, although it is uncertain if the iron cores contain significant amounts of them (e.g. Rama Murthy et al. 2003). While radiative heat transport plays no significant role in planetary dynamo regions, a substantial fraction of the heat flow in the iron cores of terrestrial planets can be transported by conduction along an adiabatic temperature gradient. It is important to note that the heat flow from the core into the overlying mantle made of solid silicate rock is controlled by the very sluggish convective circulation in the mantle (the core delivers as much heat as the mantle is able to carry away). Most models that consider the energy budget of the Earth's interior conclude that the heat flow at the top of the core is larger than the 'adiabatic heat flow' (Nimmo 2007), although a slightly subadiabatic heat flow cannot be excluded. The latter case implies a stable thermal stratification in the upper part of the liquid core, which may be overcome in the Earth by compositional convection arising from inner core growth. Also, the latent heat of inner core solidification is an important effective heat source that ensures a superadiabatic heat flow at greater depth in the core. The lack of plate tectonics on planets other than Earth means less efficient heat loss, implying slower cooling of the core. The heat flow at the core-mantle boundary is probably sub-adiabatic in these planets. Furthermore, if the planet has failed to nucleate a solid inner core, compositional buoyancy is unavailable, and the fluid core would not convect at all. This may be the case in Mars and Venus and would explain the absence of a present dynamo. Early in the planet's history cooling rates have probably been higher. In the case of Mars, a thermally driven dynamo that has operated in the first 500 million years would explain the strong magnetization of parts of the old Martian highland crust which has been formed at this time.

3.2 Magnetic Turbulence

The magnetic Reynolds number

$$Rm = \frac{UD}{\eta}, \tag{2}$$

where U is the characteristic flow velocity, D the thickness of the convecting layer and η the magnetic diffusivity, is very different for stellar and planetary dynamos. For the solar dynamo it is of the order 10^{10} using the microscopic value of magnetic diffusivity deep in the convection zone. The induction effects of the highly turbulent flow and magnetic field cannot be captured in direct global simulations of the solar dynamo, and is often treated in parameterized form in the framework of the mean-field dynamo concept (Krause and Rädler 1980); for recent reviews see Ossendrijver (2003) and Charbonneau (2005).

For the Earth's core, Rm can be estimated to be of the order 1000 (Christensen and Tilgner 2004) and in the dynamo regions of the hydrogen planets it is probably an order of magnitude larger (Stevenson 2003). At least for the geodynamo and for dynamos of other terrestrial planets the value of the magnetic Reynolds number is sufficiently small to allow a direct numerical solution of the magnetic induction equation in global models, without the need to introduce parameterizations for the magnetic induction, such as an α-effect or an effective magnetic diffusivity. The ability of running the simulations at the correct value of the magnetic Reynolds number is perhaps the essential reason for the success of geodynamo models. The hydrodynamic Reynolds number is of order 10^8 in planetary dynamos and therefore far too large to capture the small-scale turbulence of the flow in a direct simulation. These flow scales may be too small to have a direct effect on the magnetic field, however, by their back-reaction on the large scale-flow they could play a role for the dynamo.

3.3 Stratification

The density in the Sun varies by many orders of magnitude and the convection region spans many density scale heights. The dynamo region in planets covers roughly one scale height in Jupiter and substantially less in all other planets. It is not possible to simulate the entire solar convection zone in one comprehensive model. Large scale simulations typically neglect the upper most 5–10% of the solar convection zone where the pressure scale height is small and where the sound velocity is not much larger than the flow velocity and the anelastic approximation breaks down (Miesch 2005). The strong density changes are thought to play an important role for the solar dynamo. Coherent flow helicity, which is an essential ingredient for the dynamo process, arises in the Sun because of the action of the Coriolis force on rising expanding and sinking contracting parcels of plasma. Strong magnetic flux tubes have their own dynamics, because the reduction of fluid pressure that compensates magnetic pressure makes them buoyant. Both effects probably do not play a significant role in planetary dynamos; therefore most present geodynamo models neglect the small density variation and assume incompressible flow in the Boussinesq approximation. In the incompressible case, other mechanisms generate helicity that is preferentially negative in the Northern hemisphere and positive in the southern hemisphere (see Olson et al. 1999, for a discussion).

Of particular importance for the solar dynamo is the tachocline between the convection zone and the radiative interior of the Sun (Hughes et al. 2007), where differential rotation generates a strong toroidal magnetic field and the subadiabatic stratification ensures its storage for a sufficiently long time, until magnetic buoyancy leads to the rise of magnetic flux through the convection zone and emergence at the photosphere (Caligari et al. 1995; Fan 2004). For planetary dynamos it is usually assumed that the process of magnetic field generation occurs in the bulk of the convecting layer.

3.4 Inertial Forces

Another difference between the solar dynamo and planetary dynamos is the potential role of inertial forces. Their importance relative to the Coriolis force is measured by the Rossby number

$$Ro = \frac{U}{\Omega L}, \tag{3}$$

where U and L are characteristic velocity and length scale, respectively, and Ω is the rotation frequency. In the solar convection zone $Ro \approx 1$. Using estimates for the flow velocity at the top of the Earth's core of the order 1 mm/sec from the secular variation of geomagnetic field, the Rossby number is of order 10^{-6} when a global scale such as the core radius or D is used for L. Therefore, fluid motion in the Earth is often considered to be largely unaffected by inertial forces (except for special modes of motion termed 'torsional oscillations', representing collective motion on cylinders that are coaxial to the rotation axis). The general force balance is believed to be that between Coriolis force, pressure gradient force, Lorentz forces and buoyancy forces (magnetostrophic balance; Roberts 1987). However, inertial forces may become important at small length scales and can potentially feed back on the large scale flow (see Sect. 5).

4 Setup and Parameters for Geodynamo Models

For the large-scale solar dynamo direct numerical simulations of the anelastic magnetohydodynamic equations have been successful in demonstrating self-excited dynamo action and in reproducing the internal differential rotation of the Sun (e.g. Miesch 2005). But so far they fail to reproduce the cyclic behavior and the latitudinal propagation of the solar magnetic activity (Brun et al. 2004), despite the fact that most of these solar models are comparable to geodynamo models in terms of their resolution and the degree of turbulence reached. This may indicate that in the case of the Sun important induction effects occur at scales below the currently achievable resolution or in boundary layers, such as the tachocline. Currently these effects can only be modeled by local 3-D simulations or in global mean-field models. In contrast, self-consistent global simulations are able to reproduce the observed first-order properties of the geomagnetic field and much of our understanding of planetary dynamos is now based on the results of such simulations.

Most geodynamo models use basically the same setup. The equations for convective flow in the Boussinesq limit and the magnetic induction equation are solved for a rotating and electrically conducting spherical shell. Usually non-dimensional variables and dimensionless control parameters are employed. As an example, we give the equations in the form used by Christensen and Aubert (2006):

$$\frac{\partial \mathbf{u}}{\partial t} + \mathbf{u}\cdot\nabla\mathbf{u} + 2\hat{\mathbf{z}}\times\mathbf{u} + \nabla\Pi = E\nabla^2\mathbf{u} + Ra^*\frac{\mathbf{r}}{r_o}T + (\nabla\times\mathbf{B})\times\mathbf{B}, \tag{4}$$

$$\frac{\partial \mathbf{B}}{\partial t} - \nabla\times(\mathbf{u}\times\mathbf{B}) = \frac{E}{Pm}\nabla^2\mathbf{B}, \tag{5}$$

$$\frac{\partial T}{\partial t} + \mathbf{u}\cdot\nabla T = \frac{E}{Pr}\nabla^2 T, \tag{6}$$

$$\nabla\cdot\mathbf{u} = 0, \qquad \nabla\cdot\mathbf{B} = 0. \tag{7}$$

u is velocity, Π dynamical pressure, T temperature, **B** magnetic induction, and the unit vector $\hat{\mathbf{z}}$ indicates the direction of the rotation axis. The four non-dimensional control parameters are the Ekman number

$$E = \frac{\nu}{\Omega D^2}, \tag{8}$$

a modified Rayleigh number

$$Ra^* = \frac{\alpha g_o \Delta T}{\Omega^2 D}, \tag{9}$$

the Prandtl number

$$Pr = \frac{\nu}{\kappa}, \tag{10}$$

and the magnetic Prandtl number

$$Pm = \frac{\nu}{\eta}, \tag{11}$$

where ν is viscosity, α the thermal expansion coefficient, g_o gravity at the outer boundary, ΔT the (superadiabatic) temperature contrast, and κ the thermal diffusivity. We note that $(Ra^*)^{1/2}$ is often called the convected Rossby number in the astrophysical literature. Ra^* is related to the conventional Rayleigh number Ra by $Ra^* = RaE^2Pr^{-1}$.

In Table 2 we compare control parameter values used in geodynamo models with those for the Earth's core. The Rayleigh number has been normalized with its critical value Ra_c^* for the onset of convection in the absence of a magnetic field. We also list several other non-dimensional numbers that characterize the dynamo and that are a result of the model solution. The magnetic Reynolds number Rm agrees with Earth values at least in the more advanced models, whereas the hydrodynamic Reynolds number $Re = UD/\nu$ is far too small and the Rossby number Ro is too large. The Elsasser number Λ can be taken as a non-dimensional measure for the magnetic field strength. The claim that the models reproduce the geomagnetic field strength actually means that they give an Elsasser number of order one.

While the Prandtl number in the models is of the right order, the values of the other control parameters are far off. The Ekman number and the magnetic Prandtl number are too large by factors of 10^{10} and 10^6, respectively. The modified Rayleigh number is too small with respect to supercriticality, but, as we will later see, its absolute value is larger than the core value. In terms of physical parameters, the viscosity (and thermal diffusivity)

Table 2 Order of magnitude of parameters in the core and in dynamo models

	Control Parameters			
	Ra^*/Ra_c^*	E	Pm	Pr
Earth's core	$\approx$5000	10^{-15}–10^{-14}	10^{-6}–10^{-5}	0.1–1
Models	1–100	10^{-3}–10^{-6}	0.1–10^3	0.1–10^3
	Diagnostic numbers			
	Rm	Re	Ro	Λ
Earth's core	$\approx 10^3$	10^8–10^9	$\approx 10^{-6}$	0.1–10
Models	50–10^3	$<$2000	3×10^{-4}–10^{-2}	0.1–100

is too large by a factor of order 10^6 (compared to the magnetic diffusivity, which is about right). In addition, the rotation rate is too small by a factor of $\approx 10^4$ in most models. An exception are models by Glatzmaier and Roberts (1996, 1997), who use the right rotation rate at the expense of an even larger viscosity and of values for the two Prandtl numbers much in excess of one. Because the commonly employed codes work most efficiently when the Prandtl numbers are of order one, most modelers preferred the former choice. Perhaps somewhat surprisingly, the two types of models do not show fundamental differences in their magnetic field structure.

Most planetary dynamo models use a spectral transform technique, in which all variables are expanded in spherical harmonic functions concerning the angular dependence. In the radial direction different schemes are used, for example an expansion in Chebychev polynomials combined with a collocation method or a finite difference representation. Linear terms in the governing equations, for which the spherical harmonic modes decouple, are treated implicitly. Non-linear terms (often also the Coriolis force term) are treated explicitly and are evaluated on a spatial grid, which requires back-and-forth transformations at each time step. A detailed account of the standard method is given in Christensen and Wicht (2007).

5 Classes of Dynamo Solutions

Many published geodynamo models show a strong dipole field (an example is shown in Fig. 1b, c). Often the dipole in such models shows no tendency to ever reverse, although the model run time may not have been long enough to capture one of these rare events. Some dynamos show a quadrupole-dominated field, or an exotic geometry with a strong field in only one hemisphere (Grote et al. 2000; Grote and Busse 2000; Simitev and Busse 2005). In many of the non-dipolar dynamos the magnetic field structure is spatially complex (Fig. 2) and changes rapidly with time. The spatial power spectrum at the outer boundary of the dynamo models is typically white, but the dipole stands above the higher multipoles in one class of solutions and falls below the multipole level in the other class (Fig. 3). In the fully developed multipolar regime the weak dipole component changes its polarity continuously in an erratic way.

Kutzner and Christensen (2002) found that a transition from dipolar to multipolar dynamos occurs when the convective driving is enhanced. Comparing a set of models in which the relative importance of the inertial force was varied Sreenivasan and Jones (2006) found that inertia plays an important role for the selection of the dynamo regime. Christensen and Aubert (2006) analyzed a large number of model cases, varying all four control parameters. They determined that a transition from dipolar to multipolar magnetic field occurs when a local Rossby number Ro_ℓ, in which a characteristic flow length scale ℓ determined from the kinetic energy spectrum is used for L in (3), exceeds a critical value of approximately 0.12. Arguably Ro_ℓ is a more accurate measure for the ratio between inertial force and Coriolis force than the Rossby number Ro formed with the global length scale. Olson and Christensen (2006) confirmed the rule of the local Rossby number as selection criterion for the dynamo regime by including additional dynamo solutions from the literature into the analysis. They also found that dipolar dynamos that do show occasional reversals have a local Rossby number that is near the transitional value. Hence a reversal may represent an accidental brief lapse of the basically dipolar dynamo into the multipolar regime. In fact, during a model reversal the magnetic power spectrum (blue band in Fig. 3) resembles the spectrum of a (permanently) multipolar dynamo. When the dipole recovers it may then take either polarity.

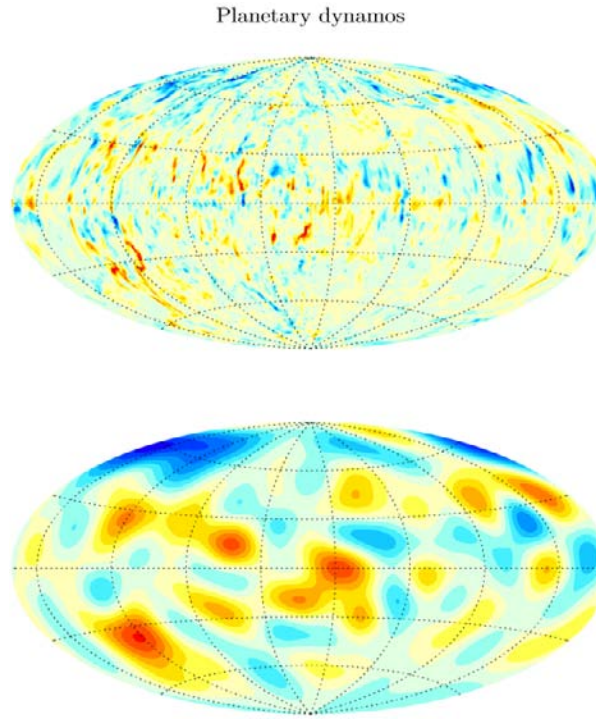


Fig. 2 Snapshot of radial field from a dynamo model with a multipolar surface field. Parameters are $E = 10^{-5}$, $Ra^* = 0.17$, $Pm = 0.5$, $Pr = 1$. *Top panel* at full resolution and in bottom panel at degrees 1–14

Olson and Christensen (2006) derived an empirical rule based on the model data set of Christensen and Aubert (2006) for relating the local Rossby number to the fundamental control parameters of the dynamo. It involves powers of all four control parameters and obviously requires an extrapolation over a large range to apply it to the planets. Nonetheless, using appropriate parameter values for the Earth, a value of $Ro_\ell \approx 0.1$ is predicted for the geodynamo, which puts it close to the transition point between the dipolar and the multipolar class (see also Sect. 8 for the application to other planets). A problem is that the predicted

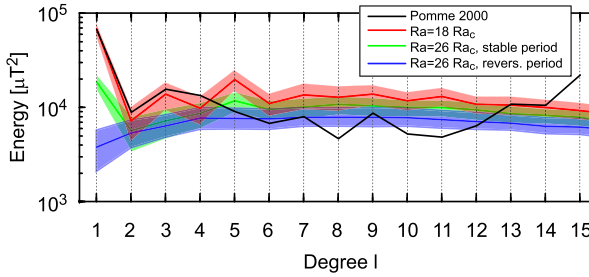


Fig. 3 Spatial power spectra at the outer boundary of the dynamo. *Black*: Spectrum at Earth's core-mantle boundary from surface observations (Maus et al. 2006), *red*: spectrum of a dynamo model in the dipolar regime. Spectra of a model in the reversing regime are shown during periods of stable dipole polarity (*green*) and during reversals (*blue*). In the model cases the lines represent time-averages and the bands indicate standard deviation. Model parameters are $E = 3 \times 10^{-4}$, $Pr = 1$, $Pm = 3$, $Ra^*/Ra_c^* = 18$ (*red*) and 26 (*green* and *blue*)

characteristic flow length scale associated with this value of Ro_ℓ is only of the order 100 m in the Earth's core. At these scales the magnetic field is diffusion-dominated and cannot be affected directly by the flow. However, in rotational flow an inverse cascade can transport energy through the action of Reynolds stresses to larger scales (e.g. Christensen 2002) which are relevant for the magnetic induction process. Hence the inertial effects at small flow scales may affect the magnetic field structure in an indirect way.

6 Flow Structure and Field Generation Mechanism

For the solar dynamo the stretching of magnetic field lines by differential rotation, particularly at the tachocline, is thought to be of major importance for the generation of a toroidal magnetic fields that is much stronger than the poloidal field. The regeneration of the poloidal component of the magnetic field is thought to be by helical fluid motions (α-effect), but the details of the physical mechanism and its precise location have not yet been identified with confidence (Charbonneau 2005). In most geodynamo models differential rotation is not a dominant part of the flow field and the toroidal and poloidal magnetic field components have similar strength (an exceptions are the models by Kuang and Bloxham 1997, 1999). The flow is strongly organized by rotational forces and is fundamentally different in the regions inside and outside the tangent cylinder (an imaginary cylinder that is co-axial with the direction of rotation $\hat{z}$ and touches the inner core at the equator). The most prominent flow pattern are convection columns outside the tangent cylinder that align with the rotation axis. The main circulation is nearly geostrophic, i.e. nearly perpendicular to and independent of z. Superimposed is an ageostrophic flow along the column axis that converges at the equator in columns with a cyclonic sense of rotation (same as the planetary rotation) and diverges away from the equator in anticyclonic vortices. The superposition of columnar and ageostrophic circulation implies a coherent negative flow helicity in the northern hemisphere and positive helicity in the southern hemisphere. This simple picture was first derived by Busse (1975) and has been used as basis for a conceptual dynamo model. Although the flow pattern in numerical dynamos is complex and time-dependent, basically it still conforms with this generic picture in most geodynamo models.

Several authors have analyzed their numerical solutions in order to understand the basic mechanism by which the magnetic field is maintained. In the tradition of mean-field dynamo theory it is studied how large-scale (e.g. axisymmetric) poloidal field is generated from large-scale toroidal field and vice versa. There is general agreement that the axial dipole field is generated from the axisymmetric toroidal field by an α-effect associated with the helical flow in the convection columns. In contrast to the classical α-effect by small helical turbulent eddies, this is a 'macroscopic', α-effect because the convection columns are not very much thinner than the width of convecting shell, at least in the models. Kageyama and Sato (1997) have been the first to describe this effect in detail.

The mechanism for generating the axisymmetric toroidal field, in which often two bundles of opposite polarity near the equatorial plane are prominent, is less clear. Olson et al. (1999) demonstrate that it is generated from the axisymmetric poloidal field by a similar macroscopic α-affect in their models, i.e. the dynamo is of the α^2-type. Also Ishihara and Kida (2002) find that field line stretching associated with the helical flow in the convection columns is the source for intense magnetic bundles that contribute both to the axisymmetric poloidal and toroidal field. Schrinner et al. (2005, 2007) studied in more detail the induction process in simple dynamo models in the framework of the mean-field concept. Their results confirm basically an α^2-mechanism. However, a description by a simple isotropic α-term is

not adequate, and an α-tensor with significant off-diagonal components and complex spatial variation is required.

Other authors present evidence for a significant role of the Ω-effect in their models. Simitev and Busse (2005) find that the toroidal field in their models is mainly generated by the interaction of the axisymmetric toroidal flow field with the axisymmetric poloidal magnetic field, even though differential rotation is not a prominent flow component. Buffett and Bloxham (2002) show that in the model of Kuang and Bloxham (1999) the strong toroidal field energy is replenished by conversion of the kinetic energy of the axisymmetric toroidal flow. While in weakly driven numerical dynamo models the regions inside the tangent cylinder (north and south of the inner core) are quiescent, vigorous flow is found in more strongly driven models (Glatzmaier and Roberts 1995; Christensen et al. 1999). In these cases a strong axisymmetric toroidal field is often generated inside the tangent cylinder region by the shearing of poloidal field lines in a polar vortex whose sense of rotation changes from clockwise near the outer boundary to anticlockwise near the inner core. In conclusion, the overall role of the Ω-effect in planetary dynamo models is not entirely clear. There may be genuine differences between models, some being of the α^2-type and others of the $\alpha\Omega$-type.

7 Comparison of Geodynamo Models with Earth's Field

The following criteria can be applied to judge the similarity between the magnetic field of a dynamo model and the geomagnetic field: (1) agreement in dipole moment or generally in field strength, (2) agreement in the shape of the spatial power spectrum, (3) qualitative agreement in the magnetic field morphology at the core-mantle boundary, (4) agreement in the time scales of secular variation, (5) agreement in the frequency and characteristic properties of dipole reversals. Many published models satisfy some of these criteria and a few satisfy all of them to a fair degree. A good guide for a dynamo model to generate a closely Earth-like magnetic field is probably that the magnetic Reynolds number and the local Rossby number must assume the appropriate value. Other parameters may be less critical. We defer the discussion of the field strength of dynamo models to Sect. 8 and discuss the other criteria below.

The spectral power distribution as function of multipole degree n is nearly white at the Earth's core-mantle boundary for $n > 2$, at a level that is about a factor of ten below the dipole power (Fig. 3). Many geodynamo models in the dipolar regime reproduce it closely (see Christensen and Wicht 2007, for a more detailed discussion), although often the dipole is somewhat stronger or weaker relative to higher multipoles than in the observed spectrum (which represents basically a single snapshot in time). In addition, characteristic features in the morphology of the geomagnetic field at the core-mantle boundary are captured in a number of models. In Fig. 1 we compare the present geomagnetic field with a snapshot from a geodynamo model. In some models that have a large Ekman number and a fairly moderate value of the magnetic Reynolds number the model field can be compared directly. In a more advanced model, like the one shown here, the field must be low-pass filtered to the resolution that is available for the geomagnetic core field to reveal the similarity. The model reproduces the flux lobes at high latitudes, weak flux at the poles and it shows flux spots of both polarities at low latitudes.

The cause for these various magnetic structures in the core field has tentatively been inferred from the associated flow structures seen in the dynamo models (e.g. Christensen et al. 1998). The high-latitude flux concentrations are related to helical convection columns outside of the inner core tangent cylinder. Cyclonic vortices are associated with downwelling

near the surface that concentrates magnetic flux. Low flux at the poles can be related to upwelling plumes near the rotation axis (Sreenivasan and Jones 2005). The plumes are accompanied by an anticyclonic vortex motion arising from a thermal wind effect. The variation of the geomagnetic field in the north polar region of the core-mantle boundary over the past hundred years, assuming that is to first order frozen into the fluid, indeed suggests that there is an anticlockwise motion inside the tangent cylinder (Olson and Aurnou 1999). Finally, bipolar pairs of flux spots at low latitudes are found in many dynamo models. They have been associated with the emergence of toroidal magnetic field tubes through the core-mantle boundary, analogous to the mechanism for the formation of sunspots (Christensen et al. 1998; Christensen and Olson 2003). Sometimes the pairs are north-south rather than east-west aligned and they show a polarity that is opposite to the general dipole polarity in their respective hemisphere, in particular in simpler dynamo models. Such configuration can arise because strong toroidal fields of opposite polarity are found at close distance near the equator and because the convective flow is strongly north-south aligned and acts on both toroidal tubes in a similar way. Comparable structures exist in the Earth's field at the core-mantle boundary (see Fig. 1a, below Africa and the Atlantic ocean) and have been explained by flux expulsion (Bloxham 1989). However, in the geomagnetic field they are more strongly offset from the equator than they are in dynamo models and other interpretations have been given for these field structures (Finlay and Jackson 2003).

Models that match Earth's magnetic Reynolds number show secular variations of the magnetic field of the right time scales. In fact, the fluid velocity in the core and hence the magnetic Reynolds number Rm can only be inferred from the secular variation of the geomagnetic field. Usually this is done assuming that the magnetic flux is frozen into the flow. Christensen and Tilgner (2004) have used results from numerical dynamo models to establish that the characteristic secular variation time scale, measured in terms of magnetic diffusion time, varies as Rm^{-1}. With this scaling law they obtain a value around $Rm = 1000$ for the Earth, in fair agreement with estimates based on the frozen-flux assumption.

Geodynamo models that are in the right regime for dipole reversals (e.g. Glatzmaier and Roberts 1995; Glatzmaier et al. 1999; Kutzner and Christensen 2002; Takahashi et al. 2005; Wicht 2005) often show a degree of agreement with the paleomagnetic record that goes beyond the simple occurrence of reversals, even in cases with very modest parameters such as a relatively large values of the Ekman number. Figure 4 shows time series of the dipole tilt, dipole moment and relative dipole field strength in such a model. Traits that are similarly found in the geomagnetic field are: (1) The directional change of the dipole field is a relatively brief event, compared to the length of the period in which the dipole is nearly aligned with the rotation axis. (2) The dipole moment starts to drop before the directional change occurs. During the reversal the magnetic field is multipolar. (3) Aside from complete reversals, strong changes in the dipole direction that are brief and non-persistent also occur (geomagnetic excursions). The actual frequency of reversals in geodynamo models seems to depend on the fine tuning of parameters and is also controlled by second order effects, such as the non-uniform pattern of heat flow at the core-mantle boundary that is imposed by a heterogeneous structure of Earth's lower mantle (Glatzmaier et al. 1999).

8 Scaling of Dynamo Properties

An important question is what controls the vigor of convection and the strength of the magnetic field in a planetary dynamo. Some heuristic scaling laws based on mixing length theory or an assumed magnetostrophic force balance have been proposed for relating the velocity

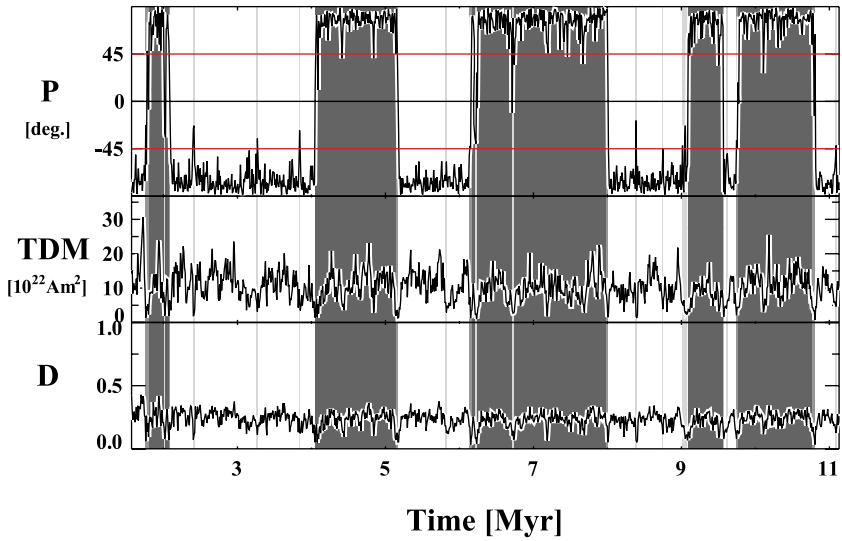


Fig. 4 Time series of the dipole tilt with respect to the equator (P), the true dipole moment (TDM), and the dipole strength relative to the total field strength at the core mantle boundary (D) for a dynamo model with $E = 10^{-3}$, $Ra^* = 0.5$, $Pm = 10$, $Pr = 1$. The present TDM of the geomagnetic field is 8×10^{22} Am^2. Dark and light bands indicate polarity intervals (unpublished, courtesy of Johannes Wicht)

to the heat flux, or more generally, to the buoyancy flux, which also comprises compositionally driven convection (Stevenson 1979, 2003; Starchenko and Jones 2002). These theories predict that the velocity should scale with the buoyancy flux to the $1/3$ or $1/2$ power, respectively. The usual assumption on the magnetic field strength for a dynamo in a magnetostrophic force balance is that the Elsasser number (1), representing the ratio of Lorentz forces to Coriolis forces, must be of order one inside the dynamo region, which sets the value of B.

The availability of a sufficiently large set of dynamo solutions which cover a decent range of control parameter space has led Christensen and Tilgner (2004) and Christensen and Aubert (2006) to derive scaling laws in a partly empirical way from numerical model results. Since the model values of several control parameters are far removed from planetary values, the question arises if the models are in the same dynamical regime. In particular, viscous friction which is thought to be negligible in planetary cores, has been suspected to play a major role in the models.

Christensen and Aubert (2006) tested if the values of viscous, thermal and magnetic diffusion constants influence the first-order properties of the numerical dynamos. In terms of non-dimensional parameters, the role of diffusion is controlled by the Ekman number and the two Prandtl numbers. Taking the modified form of the Rayleigh number (9) rather than the conventional Rayleigh number as the fourth control parameter has the advantage that it is independent of any of the diffusion constants. Because for planetary dynamos the driving heat flux or buoyancy flux is more readily estimated than the superadiabatic temperature contrast, Christensen and Aubert (2006) use a flux-based modified Rayleigh number

$$Ra_Q^* = \frac{r_i}{r_c} \frac{q_{buoy}}{\Omega^3 D^2}, \tag{12}$$

where q_{buoy} is the convective buoyancy flux (per unit area on the outer boundary) and r_i the radius of the inner core. The total flux Q_{buoy} as a function of radius is constant in their models. For purely thermal convection $q_{buoy} = \alpha g_o q_{conv}/(\rho c_p)$, where q_{conv} is the convected heat flux and c_p the heat capacity. Ra_Q^* is a non-dimensional measure for the power generated by buoyancy forces. The characteric flow velocity, magnetic field strength and heat transport properties of the dynamo solutions are expressed by non-dimensional numbers that are defined such that they do not relate to any of the diffusion parameters. These are the Rossby number

$$Ro = \frac{U}{\Omega D}, \tag{13}$$

the Lorentz number

$$Lo = \frac{B}{\sqrt{\mu\rho}\Omega D}, \tag{14}$$

and a modified Nusselt number

$$Nu^* = \frac{r_i}{r_c}\frac{q_{conv}}{\rho c_p \Delta T \Omega D}, \tag{15}$$

where U and B refer to the mean (rms) values of velocity and magnetic field inside the dynamo. Covering a range of at least two orders of magnitude in each of the parameters describing diffusive processes and six orders of magnitude for Ra_Q^*, Christensen and Aubert (2006) found that the results depend at most weakly on E, Pr and Pm. The three characteristic numbers Nu^*, Ro and Lo relate by simple power-laws to the modified Rayleigh number Ra_Q^*, with exponents of approximately 1/2, 2/5 and 1/3, respectively. The exponent for the Rossby number is intermediate between the value derived from mixing length theory and that from magnetostrophic balance. It can be rationalized by assuming a triple balance of buoyancy, inertial forces and Coriolis forces (Aubert et al. 2001). The exponent for the Lorentz number can be understood on the basis that ohmic dissipation must be balanced by the power generated by buoyancy and that the ratio of magnetic energy to ohmic dissipation (the ohmic dissipation time constant) is inversely related to the magnetic Reynolds number (Christensen and Tilgner 2004). The fit to the model data could actually be improved by correcting for the energy lost by viscous dissipation. Figure 5 shows an updated plot of the corrected Lorentz number versus Ra_Q^*. It is encouraging that all results can be collapsed fairly well on a simple dependence on the modified Rayleigh number, irrespective of the values of the three control parameters E, Pm, Pr that refer to diffusive processes.

The Elsasser number, based on the rms field strength in the dynamo, was found to cover values between 0.1 and 100 in the various numerical models. Unless one is willing to accept that an 'order-one value', can actually lie within a range of three orders of magnitude, the Elsasser number rule does not seem to apply for the numerical dynamos, and may not be a generally valid guide to planetary magnetic field strength.

Casting the scaling law for the magnetic field into dimensional form and setting the exponent to 1/3, the magnetic field strength is given by

$$B \propto \mu_o^{1/2} \rho^{1/6} q_{buoy}^{1/3}, \tag{16}$$

and is independent of the rotation rate and of the conductivity. In that respect it differs from all previously suggested scaling laws. Of course, the conductivity must be high enough so that the critical magnetic Reynolds number for a dynamo is exceeded. Furthermore, rotation must be fast (the local Rossby number small) to obtain a dipolar dynamo. However, once

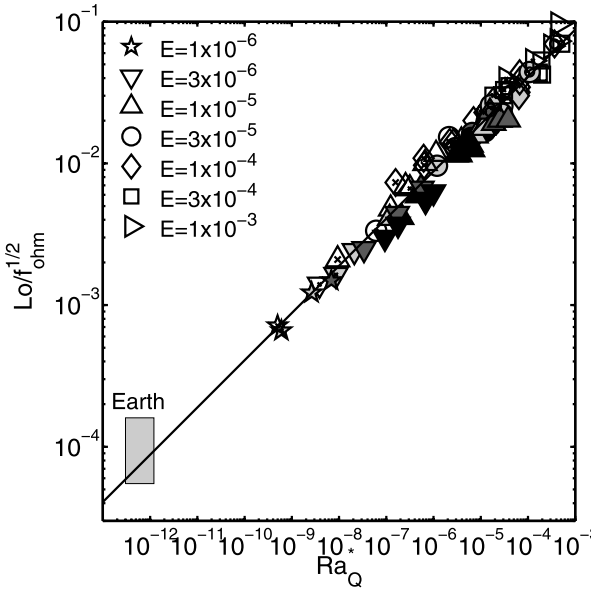


Fig. 5 Lorentz number corrected for dissipation versus modified flux-based Rayleigh number. Model data are mostly from Christensen and Aubert (2006) with some additions. Shading of the symbols indicates the value of the magnetic Prandtl number, where darker means a lower value. *Crosses* inside the main symbol indicate $Pr > 1$ and *circles* indicate $Pr < 1$. The *line* represents the best fit for a forced slope of 1/3. f_{ohm} is the contribution of ohmic dissipation to total dissipation. The *shaded region* for the geodynamo is based on estimates for the core field strength and buoyancy flux assuming $f_{ohm} \approx 1$ in the core

these conditions are met, (16) implies that the precise values of the conductivity and of the rotation rate have no influence on the magnetic field strength.

For testing whether the scaling laws are compliant with the geodynamo, the modified Rayleigh number and the Lorentz number of the Earth's core must be determined. Estimates for the buoyancy flux in the Earth's core suffer from substantial uncertainties and rely on indirect arguments. Using their scaling law that relates velocity to the buoyancy flux, Christensen and Aubert (2006) derived a value $Q_{buoy} = 4\pi r_c^2 q_{buoy} \approx 3 \times 10^4$ kg s^{-1} for a typical flow velocity of 1 mm/s obtained from secular variation. This value for the buoyancy flux is rather low and implies that the heat flow at the core-mantle boundary is close to the conductive heat flow along an adiabatic gradient. Most other estimates for the core heat flow are higher (e.g. Nimmo 2007). Considering a range of 3–12 $\times 10^4$ kg s^{-1}, the modified Rayleigh number for Earth's core is of the order 10^{-12}. To estimate the mean field strength inside the core requires an assumption of how this relates to the 'observed', strength at the top of the core, which is 0.26 mT for the dipole part of the field. Taking their dynamo models with an Earth-like field morphology as a guide, in which the toroidal field and the poloidal field are of similar strength, Christensen and Aubert (2006) estimated a ratio of 6–7 between the mean internal field strength and the dipole strength at the core surface. Considering values between 1 and 3 mT for the core field, the Lorentz number is in the range 0.6–1.6 $\times 10^{-4}$.

Even though the scaling law must be extrapolated a long way from the model parameter values to Earth values, the estimates for the geodynamo fall on the line defined by the various model results (Fig. 5). This strongly suggests that despite viscosity is far too large, the model dynamos operate in the same dynamical regime as the geodynamo does and that the agreement, for example in magnetic field morphology and reversal behavior, is not only a coincidence.

Christensen and Aubert (2006) applied their scaling law also to Jupiter, where the buoyancy flux can be estimated from the planet's excess luminosity. They predict a field strength of 8 mT inside the dynamo. This is in good agreement with the observed dipole moment, assuming a similar ratio between mean internal field strength and dipole strength as in case of

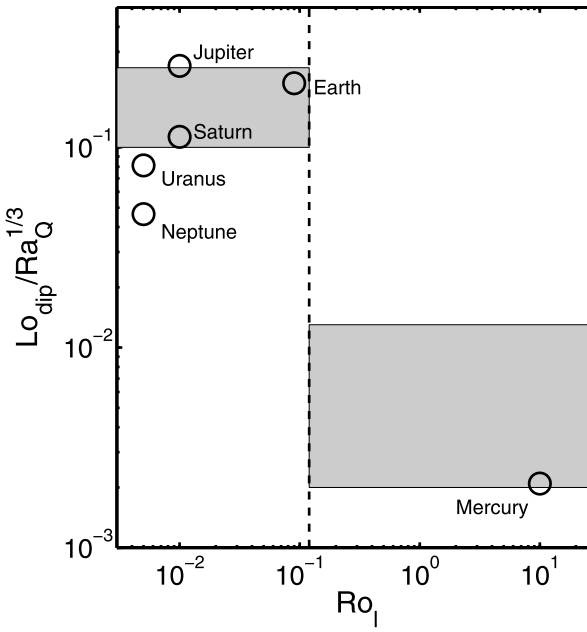


Fig. 6 Lorentz number based on the observed dipole moments normalized by estimated values of the modified flux-based Rayleigh number plotted against the local Rossby number for the magnetic planets in the solar system. *Shaded ranges* indicate predicted values based on numerical model results in the dipolar regime (*left*) and the multipolar regime (*right*)

the Earth. Despite a much larger buoyancy flux, the bigger size and the more rapid rotation put Jupiter's dynamo at lower values of Ra_Q^* and Lo compared to the geodynamo.

Olson and Christensen (2006) used an even larger set of numerical dynamos, many of them taken from the literature, to derive a scaling law for the dipole moment as the most fundamental observable property of planetary magnetic fields. Expressing the dipole field strength again by a Lorentz number Lo_{dip}, they confirm that it depends on the cubic root of the power driving the dynamo (expressed by Ra_Q^*) as long as the magnetic field is dominated by the dipole. The scatter is somewhat larger than in the case of scaling the mean internal field, partly because a more heterogeneous set of dynamo models with different boundary conditions has been considered, and partly because the dipole is just one component of the field that does not need to keep a constant ratio to the total field when parameters are changed. The normalized dipole Lorentz number $Lo_{dip}/Ra_Q^{*1/3}$ is nearly constant in the dipolar regime, but drops by more than an order of magnitude upon transition to the non-dipolar regime at values of the local Rossby number around 0.12.

Applying their scaling law that relates the local Rossby number to the control parameters, Olson and Christensen (2006) find that all planets except Mercury should fall into the dipolar regime ($Ro_\ell < 0.12$). The scaling law for Lo_{dip} then predicts dipole moments in fair agreement with their observed values for most planets, despite large uncertainties in some cases on their internal properties and particularly the buoyancy flux (Fig. 6).

9 Specific Models for Various Planets Other than Earth

In order to explain idiosyncrasies in the structure or strength of the magnetic field of various planets dedicated dynamo models have been presented in recent years. Several of them rely on the existence of stably stratified layers in the fluid core of the planet.

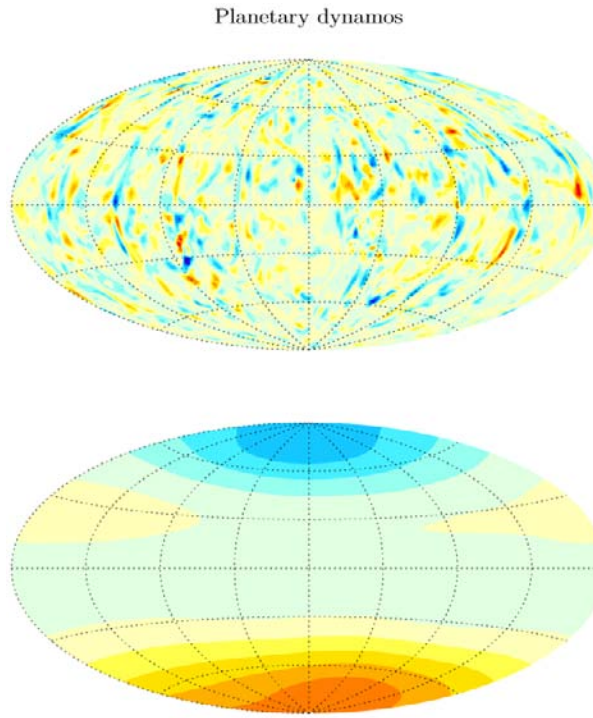


Fig. 7 Radial magnetic field in a model for Mercury's dynamo with a partly stable core. Parameters are $E = 10^{-4}$, $Ra_Q^* = 6 \times 10^{-4}$, $Pm = 3$, $Pr = 1$, inner core radius half the core radius and unstable layer thickness 44% of fluid core thickness. *Top panel*: at the top of the dynamo region (deep inside the core) with color contour step 100,000 nT; *bottom panel*: at Mercury's surface with step 100 nT

9.1 Mercury

The main problem with Mercury's magnetic field is to reconcile its relative weakness with the assumption of a hydromagnetic dynamo operating in the large iron core of the planet, whose outer boundary is at approximately 0.75 planetary radii. Observations of Mercury's forced libration (Margot et al. 2007) strongly indicate that Mercury's core is at least partially liquid. The existence of a solid inner core is likely, but its size is unconstrained. Dynamo models with a very large inner core (Stanley et al. 2005; Takahashi and Matsushima 2006) or with a very small inner core (Heimpel et al. 2005) succeeded in producing relatively weak magnetic fields in the exterior. However, the field in these models is either still too strong by a factor of ten or more, or it contains strong higher multipole components. Magnetometer data from the recent Messenger flyby have reinforced the preliminary conclusion from Mariner 10 data that the internal field is large-scaled and dominated by a slightly tilted dipole (Anderson et al. 2008). This, however, is in conflict with the prediction that Mercury should be in the multipolar dynamo regime based on a large value of the local Rossby number caused by the planet's very slow rotation (Fig. 6).

Christensen (2006) and Christensen and Wicht (2008) present dynamo models in which only a deep sublayer of the fluid core is convecting, whereas the upper region is stably stratified. This is based on thermal evolution models that predict a heat flow at Mercury's core-mantle boundary substantially less than the heat that can be conducted along an adiabatic temperature gradient (Breuer et al. 2007). Compositional buoyancy and the latent heat of inner core growth make the deep core region convectively unstable. Here a strong magnetic field is generated, which is small-scaled in models where the local Rossby number exceeds the threshold value for the multipolar regime (Fig. 7). The small-scale field varies

rapidly with time. Therefore, it is strongly attenuated by a skin effect in the conducting stable layer and is virtually unobservable outside the core. The dipole component makes only a small contribution inside the dynamo, but varies more slowly. Hence it can penetrate through the stable layer and dominates the structure of the very weak field at the planetary surface (Fig. 7, bottom; note the factor 1000 difference in the color scheme).

9.2 Saturn

Stevenson (1980, 1982) suggested that stable stratification at the top of Saturn's metallic hydrogen layer could be the cause for the extremely high degree of axisymmetry in the planet's magnetic field. In Saturn the stratification can arise because helium may be partly immiscible with metallic hydrogen near the top of the metallic layer. While the density stratification suppresses convection, it allows for toroidal flow, in particular differential rotation. Let us assume for simplicity that the whole stable layer rotates like a uniform shell with respect to the underlying dynamo region and that the dynamo field is stationary. Seen from a reference frame that is fixed to the rotating shell, the non-axisymmetric field components will become time-dependent, whereas the axisymmetric components remain stationary. If the magnetic Reynolds number characterizing the shell motion is large enough, a skin effect will eliminate the non-axisymmetric parts of the field, leaving the axisymmetric components unaffected.

Christensen and Wicht (2008) find in their models (originally intended for Mercury's dynamo) that latitudinal differences in the heat flow from the dynamo region into the stable shell drive strong differential rotation as a thermal wind circulation. In their models the magnetic field inside the dynamo region has strong non-axisymmetric contributions, whereas the field outside the core has a high degree of axisymmetry. The latter disappears when in a control experiment the differential rotation in the stable layers is suppressed. In model cases where the local Rossby number is below the threshold value for the dipole-multipole transition, which should be the case in Saturn (see Fig. 6), the axial dipole is a strong and slowly time-varying component of the magnetic field inside the dynamo. Consequently the field outside the core is much stronger than it is in the case of Mercury. Although the dipolar models in Christensen and Wicht (2008) have not been tuned to Saturn parameters, they produce axisymmetric magnetic fields of a similar strength as Saturn's field and basically support Stevenson's hypothesis for the cause of axisymmetry.

9.3 Uranus and Neptune

The observed fields of Uranus and Neptune are multipolar. The rule of the local Rossby number would put them into the dipolar regime (Fig. 6), and may fail in these cases. Stanley and Bloxham (2004, 2006) present a dynamo model with a thin convecting shell that surrounds a fluid conducting but stable core region. Such a structure had been proposed to explain the relatively low excess luminosity of the planets. Some of their dynamo models generate magnetic fields that agree well with the observed distribution of power in the dipole, quadrupole and octupole components.

The conductivity of the ionic liquid in Uranus and Neptune is lower than that of metallic liquids by a factor of 10^2–10^3. Gómez-Pérez and Heimpel (2007) showed in dynamo models that the magnetic field becomes less dipolar when the magnetic diffusivity is increased relative to viscosity.

10 Discussion

Numerical dynamo models based on the direct numerical simulation of the fundamental MHD equations are remarkably successful in matching the main properties of the geomagnetic field and to some extent those of other planetary magnetic fields. In this respect modeling of planetary dynamos seems to be more advanced than modeling the solar dynamo. The reasons for the success of planetary dynamo models are a matter of speculation, but the following points may be important:

(1) Density stratification (compressibility) plays a small role in planetary dynamos, which at least eases the task. (2) It is possible to fully resolve the magnetic field structure and therefore the details of the magnetic induction process. Put differently, direct numerical simulations at the correct value of the magnetic Reynolds number are possible. (3) Although the model viscosity and thermal diffusivity are far larger than realistic microscopic values, the scaling laws obtained from systematic parameter studies suggest that they are low enough to not play a first-order role. (4) It seems that the large-scale flow structure, which is responsible for magnetic induction, is modeled realistically. Small flow scales may be important in planetary cores through their feedback on the large-scale flow. These small scales are missing in the models, but their effect is perhaps similar to that obtained by the interaction of the smallest resolved scales and the large scales in the models when the local Rossby number has the appropriate value. As in the case of the magnetic Reynolds number, simulations at Earth-like values of Ro_ℓ are numerically feasible.

Acknowledgements We thank Johannes Wicht for providing Fig. 4. This research was supported in part by National Science Foundation (NSF) under Grant No. NSF PHY05-51164. KITP preprint number NSF-KITP-08-104. NSF also sponsors the National Center for Atmospheric Research.

References

B.J. Anderson, M.H. Acuña, H. Korth et al., Science **321**, 82 (2008)
J. Aubert, D. Brito, H.C. Nataf, P. Cardin, J.P. Masson, Phys. Earth Planet. Int. **128**, 51 (2001)
J. Bloxham, Philos. Trans. R. Soc. Lond. **87**, 669 (1989)
D. Breuer, S.A. Hauck II, M. Buske, M. Pauer, T. Spohn, Space Sci. Rev. **229–260**, 132 (2007)
A.S. Brun, M.S. Miesch, J. Toomre, Astrophys. J. **614**, 1073 (2004)
B.A. Buffett, J. Bloxham, Geophys. J. Int. **149**, 211 (2002)
F.H. Busse, Geophys. J. R. Astron. Soc. **42**, 437 (1975)
P. Caligari, F. Moreno-Insertis, M. Schüssler, Astrophys. J. **441**, 886 (1995)
P. Charbonneau, Living Rev. Sol. Phys. (2005). http://www.livingreviews.org/lrsp-2005-2
U.R. Christensen, J. Fluid Mech. **470**, 115 (2002)
U.R. Christensen, Nature **444**, 1056 (2006)
U.R. Christensen, J. Aubert, Geophys. J. Int. **166**, 97 (2006)
U.R. Christensen, P. Olson, Phys. Earth Planet. Int. **138**, 39 (2003)
U.R. Christensen, P. Olson, G.A. Glatzmaier, Geophys. Res. Lett. **25**, 1565 (1998)
U.R. Christensen, P. Olson, G.A. Glatzmaier, Geophys. J. Int. **138**, 393 (1999)
U.R. Christensen, A. Tilgner, Nature **429**, 169 (2004)
U.R. Christensen, J. Wicht, in *Core Dynamics*, ed. by G. Schubert. Treatise of Geophysics, vol. 8 (Elsevier, Amsterdam, 2007), pp. 245–282
U.R. Christensen, J. Wicht, Icarus **196**, 16 (2008)
V. Courtillot, P. Olson, Earth Planet. Sci. Lett. **260**, 495 (2007)
Y. Fan, Living Rev. Sol. Phys. (2004). http://www.livingreviews.org/lrsp-2004-1
C.C. Finlay, A. Jackson, Science **300**, 2084 (2003)
G.A. Glatzmaier, R.S. Coe, L. Hongre, P.H. Roberts, Nature **401**, 885 (1999)
G.A. Glatzmaier, P.H. Roberts, Nature **337**, 203 (1995)
G.A. Glatzmaier, P.H. Roberts, Physica D **97**, 81 (1996)
G.A. Glatzmaier, P.H. Roberts, Contemp. Phys. **38**, 269 (1997)

N. Gómez-Pérez, M. Heimpel, Geophys. Astrophys. Fluid Dyn. **101**, 371 (2007)
E. Grote, F.H. Busse, Phys. Rev. **62**, 4457 (2000)
E. Grote, F.H. Busse, A. Tilgner, Phys. Earth Planet. Int. **117**, 259 (2000)
D. Gubbins, J. Bloxham, Geophys. J. R. Astron. Soc. **80**, 695 (1985)
M. Heimpel, J.M. Aurnou, F.M. Al-Shamali, N. Gómez-Pérez, Earth Planet. Sci. Lett. **236**, 542 (2005)
D.W. Hughes, R. Rosner, N.O. Weiss, *The Solar Tachocline* (Cambridge University Press, Cambridge, 2007)
N. Ishihara, S. Kida, J. Fluid Mech. **465**, 1 (2002)
A. Jackson, A.R.T. Jonkers, M.R. Walker, Philos. Trans. R. Soc. Lond. A **358**, 957 (2000)
A. Kageyama, T. Sato, Phys. Rev. E **55**, 4617 (1997)
M. Korte, C.G. Constable, Geochem. Geophys. Geosyst. **6**, Q02H16 (2005)
F. Krause, K.-H. Rädler, *Mean-Field Magnetohydrodynamics and Dynamo Theory* (Pergamon Press, Oxford, 1980)
W. Kuang, J. Bloxham, Nature **389**, 371 (1997)
W. Kuang, J. Bloxham, J. Comput. Phys. **153**, 51 (1999)
C. Kutzner, U.R. Christensen, Phys. Earth Planet. Int. **131**, 29 (2002)
C. Kutzner, U.R. Christensen, Geophys. J. Int. **157**, 1105 (2004)
J.L. Margot, S.J. Peale, M.A. Slade, I.V. Holin, Science **316**, 710 (2007)
S. Maus, M. Rother, C. Stolle et al., Geochem. Geophys. Geosyst. **7**, Q07008 (2006)
M.S. Miesch, Living Rev. Sol. Phys. (2005). http://www.livingreviews.org/lrsp-2005-1
F. Nimmo, in *Core Dynamics*, ed. by G. Schubert. Treatise of Geophysics, vol. 8 (Elsevier, Amsterdam, 2007), pp. 31–65
N. Olsen, G. Hulot, T.J. Sabaja, in *Geomagnetism*, ed. by G. Schubert. Treatise of Geophysics, vol. 5 (Elsevier, Amsterdam, 2007), pp. 33–75
P. Olson, J. Aurnou, Nature **402**, 170 (1999)
P. Olson, U.R. Christensen, Earth Planet. Sci. Lett. **250**, 561 (2006)
P. Olson, U.R. Christensen, G.A. Glatzmaier, J. Geophys. Res. **104**, 10,383 (1999)
M. Ossendrijver, Astron. Astrophys. Rev. **11**, 287 (2003)
V. Rama Murthy, W. van Westrenen, Y. Fei, Nature **423**, 163 (2003)
P.H. Roberts, in *Geomagnetism*, vol. 2, ed. by J. Jacobs (Academic Press, London, 1987), pp. 251–306
M. Schrinner, K.-H. Rädler, D. Schmitt, M. Rheinhardt, U. Christensen, Astron. Nachr. **326**, 245 (2005)
M. Schrinner, K.-H. Rädler, D. Schmitt, M. Rheinhardt, U. Christensen, Geophys. Astrophys. Fluid Dyn. **101**, 81 (2007)
R. Simitev, F.H. Busse, J. Fluid Mech. **532**, 365 (2005)
B. Sreenivasan, C.A. Jones, Geophys. Res. Lett. **32**, L20301 (2005)
B. Sreenivasan, C.A. Jones, Geophys. J. Int. **164**, 467 (2006)
S. Stanley, J. Bloxham, Nature **428**, 151 (2004)
S. Stanley, J. Bloxham, Icarus **184**, 556 (2006)
S. Stanley, J. Bloxham, W.E. Hutchison, Earth Planet. Sci. Lett. **234**, 341 (2005)
S.V. Starchenko, C.A. Jones, Icarus **157**, 426 (2002)
D.J. Stevenson, Geophys. Astrophys. Fluid Dyn. **12**, 139 (1979)
D.J. Stevenson, Science **208**, 746 (1980)
D.J. Stevenson, Geophys. Astrophys. Fluid Dyn. **21**, 113 (1982)
D.J. Stevenson, Earth Planet. Sci. Lett. **208**, 1 (2003)
F. Takahashi, M. Matsushima, Geophys. Res. Lett. **33**, L10202 (2006)
F. Takahashi, M. Matsushima, Y. Honkura, Science **309**, 459 (2005)
J.A. Tarduno, R.D. Cottrell, M.K. Watkeys, D. Bauch, Nature **466**, 657 (2007)
J. Wicht, Geophys. J. Int. **162**, 371 (2005)

Observations of Photospheric Dynamics and Magnetic Fields: From Large-Scale to Small-Scale Flows

N. Meunier · J. Zhao

Originally published in the journal Space Science Reviews, Volume 144, Nos 1–4, 127–149.
DOI: 10.1007/s11214-008-9472-7

Abstract This paper reviews solar flows and magnetic fields observed at the photospheric level. We first present the context in which these observations are performed. We describe the various temporal and spatial scales involved, and the coupling between them. Then we present small-scale flows, mainly supergranulation and flows around active regions. Flows at the global scale are then reviewed, again with emphasis on the flows, i.e. differential rotation, torsional oscillation and meridional circulation. In both small- and global-scale we discuss the coupling between flow fields and magnetic field and give an overview of observational techniques. Finally, the possible connection between studies of solar activity and stellar activity is briefly discussed.

Keywords Sun · Magnetic fields · Dynamics · Helioseismology · Meridional flows

1 Introduction

This paper presents an overview of photospheric flows and magnetic fields. The photospheric magnetic fields have been observed and studied at different scales, from small-scale structures (such as sunspots or plages) to the cyclic behavior at the 11-year scale or more. The generation of solar magnetism and the periodicity of the solar magnetism are closely related to the solar large-scale and global-scale flows, such as differential rotation and meridional flows. The dynamo action is thought to take place at the bottom of the convective zone in the tachocline (e.g. Ossendrijver 2003 for a review, and Thompson and Weiss 2008; Dikpati and Gilman 2006 in this issue). However, apart from the recent results obtained using helioseismology, we do not have direct access to that region. This is particularly true for

N. Meunier (✉)
Laboratoire d'Astrophysique de l'Observatoire de Grenoble, CNRS, Université Joseph Fourier, BP 43, 38041 Grenoble Cedex, France
e-mail: nadege.meunier@obs.ujf-grenoble.fr

J. Zhao
W.W. Hansen Experimental Physics Laboratory, Stanford University, Stanford, CA 94305-4085, USA

M.J. Thompson et al. (eds.), *The Origin and Dynamics of Solar Magnetism*.
DOI: 10.1007/978-1-4419-0239-9_9

the magnetic field, which is not well characterized below the photosphere. Therefore surface observations are very important to constrain the dynamo models. On the other hand, because they are indirect manifestations of the internal magnetic fields, they must be taken with caution as surface effects or processes occurring during the rising of the magnetic flux through the convective zone (Schüssler 2005) may alter the original properties. Also, it is important to keep in mind the coupling of the photosphere with the upper layers of the solar atmosphere (chromosphere and corona), e.g., heating of the corona, or more energetic events such as flares and CMEs (for more details see Wedemeyer-Böhm et al. 2008 and van Driel-Gesztelyi and Culhane 2008 in this issue).

In this review, we will focus on selected topics since specific magnetic structures such as sunspots and small-scale magnetic fields are presented elsewhere in this issue. The outline of the paper is as follows. In Sect. 2, we describe the different spatial and temporal scales involved and how they are coupled together. In the two following sections, we describe in more details some of the photospheric flows and magnetic fields on small scales (Sect. 3) and on large scales (Sect. 4), with emphasis on the flows and on the coupling between flows and magnetic fields. Then in Sect. 5 we illustrate a few questions concerning stellar magnetic fields which are of interest for the study of solar activity or related flows. We conclude in Sect. 6.

2 The Spatial and Temporal Scales

2.1 Overview of Different Scales

Many velocity and magnetic structures, with sizes and lifetimes covering a wide range, are observed on the Sun (see Fig. 1). It is beyond the scope of this paper to illustrate all of them, as there are other papers in this issue devoted to some of them specifically. We therefore focus here on the flows and on some more general considerations, as well as on the coupling among these structures of different spatial and temporal scales.

Known velocity structures cover scales from granulation (typically 1 Mm but including some fine structures at smaller scales) to supergranulation (of a typical scale of 30 Mm) and to the global dynamics such as angular rotation, meridional circulation and torsional oscillations. Magnetic structures also cover a wide range of scales in the photosphere, from

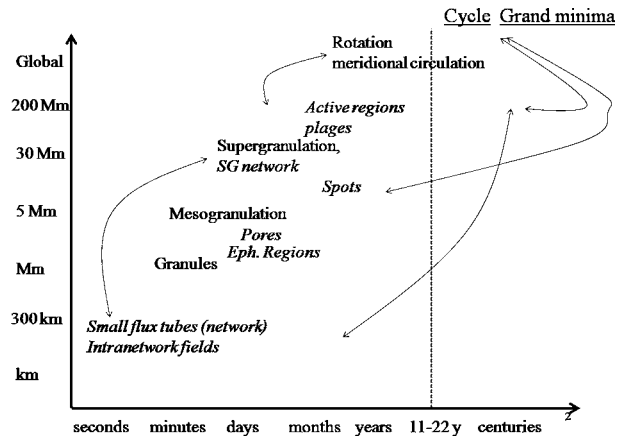


Fig. 1 Qualitative overview of the spatial and temporal scales of various magnetic (*italic*) and velocity structures in the solar photosphere. *Arrows* indicate some of the connections between structures at different scales

100 km strong-field structures (flux tubes) to sunspots and active regions (up to several 100 Mm). Thus typical studies of the size distribution and fractal analyses usually cover 2 or 3 orders of magnitude in area. A general correlation between the spatial scale and the temporal scale is observed, as large structures tend to live longer, from a few minutes to weeks or months. In case of magnetic structures, there is also a correlation between their size and the amount of magnetic flux they contain. An anti-correlation between the size and the number of structures is also observed, as there are more small structures than large ones.

2.2 Coupling between Different Scales

A given category of structures usually covers a limited range of sizes and lifetimes. For example, sunspots have sizes of ∼5–20 Mm and lifetimes from a few days to a few weeks. However, they are strongly related to other scales, for example the 11-year solar cycle, as their number varies in phase with it. They also include small-scale features, such as penumbral filaments, umbral dots and light bridges (e.g. Bharti et al. 2007; Langhans et al. 2007; Rimmele 2008 for recent works). Another example is the magnetic network: it is spatially organized at the supergranular scale (∼30 Mm), but it is also constituted of small-scale flux tubes (∼100 km) in the lower part of that broad spectrum. A presentation of all these structures by separating different scales has therefore some limitations because they are all interconnected. Furthermore, the two topics of this paper, dynamics and magnetic fields, are strongly connected, and this will be illustrated in Sects. 3 and 4. Again, the strong relationship between supergranulation and the magnetic network is a good example of this coupling.

A simple illustration of the coupling between different scales is presented in Fig. 2, which shows an ordinary image of the photosphere (upper panel). The complexity of sunspots and their uneven distribution on the surface is obvious. Higher spatial resolution observations show an even larger complexity, which is difficult to model. However, with only a few images, it is already possible to observe that they are not randomly distributed in the solar photosphere, but appear in two bands of latitude. This is even more apparent on a magnetogram (lower panel of Fig. 2), which shows the Hale's law and the absence of active regions close to the poles. So even if the individual structures are complicated (either because of their general shape or because of their multi-component structure), they follow a large-scale organization that is crucial to model different aspects of solar magnetic activity. The result is quite well-known when one considers long time series: it leads for example to the classical butterfly diagram, and to the variation of sunspot number over time with the 11-year solar cycle. Such a variation includes the existence of grand minima, such as the Maunder minimum, and more generally a variability in the amplitude of solar cycles. This constitutes of course only a subset of the possible constraints, as other well-known properties are observed, for example: the Joy's law describing the tilt of active region polarities with respect to the equator (e.g. Howard 1992), the existence of active longitudes (Gaizauskas et al. 1983; Berdyugina and Usoskin 2003), and irradiance variations (e.g. Fröhlich 2006).

2.3 Multi-Scale Analysis

Another way to look at the various scales involved in the description of solar flows and activity is the following. Most studies fall into either one of the two categories that are presented in more details in Sects. 3 and 4: studies focusing on small scales (with a relatively broad definition, i.e. including individual active regions), and those focusing on the global scale. On the small-scale side, for example, a very large number of studies have been devoted

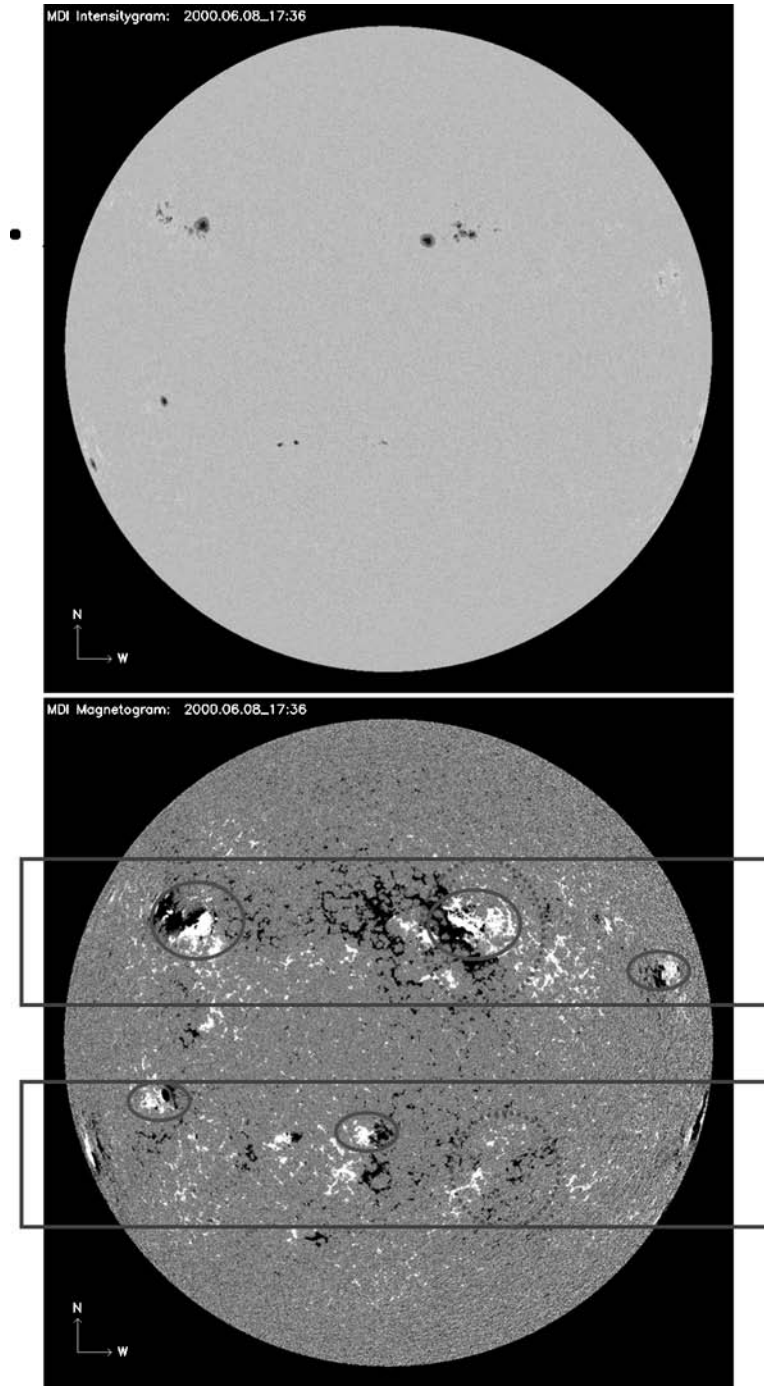


Fig. 2 *Upper panel*: solar photosphere in the continuum close to 6768 Å Ni line observed by MDI/SOHO. *Lower panel*: Magnetogram in that line. The *two rectangular boxes* focus on the two active bands of latitude. The *ovals* emphasize a few bipolar regions, showing opposite polarities of the leading (respectively following) polarities in the two hemispheres

to the detailed observation of individual structures (such as sunspots or active regions). On the other hand, as we see for sunspots, these structures are organized on a larger scale. They are doing so on long time scales, leading to studies of time series, for example, the sunspot number series. We will focus on some of these in the following sections. These two approaches naturally have counterparts on the theoretical side, for example, the study of individual flux tubes for the small scales, or the mean-field dynamo theory for the global scale. Several solutions have been investigated to take into account the couplings among these scales.

A first possibility is the study of structure properties (for example the size or the dynamics) over the solar cycle, therefore relating small-scale properties to the solar cycle. Such analyses provide some constraints on the physical processes involved in the formation of these structures. An example is the study of sunspot intensity variations over the solar cycle (Albregtsen and Maltby 1981; Albregtsen et al. 1984; Penn and Livingston 2006; Penn and MacDonald 2007). In this latter work, they showed that umbrae are brighter at cycle minimum compared to cycle maximum, which might imply that the toroidal field strength produced by the dynamo could vary during the cycle. It also has an impact on sunspot models. Along the same line, Lawrence (1987) has studied various properties of active regions over the solar cycle, such as average sunspot area per region, average plage to sunspot area ratio, and average plage intensity. They all vary during the cycle, suggesting the existence of some strong connection between active region sunspot areas and plage intensities, with some influence on the models relating sunspot and plage luminosities.

Other classical and well-studied approaches are the scaling laws that have been derived from the study of magnetic features and from time series of activity. This approach is natural given their turbulent nature. The simplest example is probably the size distribution of structures, which have been used for a long time for active regions (Harvey 1993). It is well known that this distribution is changing during the solar cycle. We also know that the size distribution of the magnetic network features is also varying in phase with the solar cycle (Meunier 2003). It remains however an open question for the weak-field intranetwork component (see Meunier et al. 2008 for a partial answer), and it is of course of great interest to determine its origin, either in the global dynamo or in a local dynamo action.

The fractal analysis of magnetic features is another example of a constraint that can be derived using properties over more than two orders of magnitude in size. Many techniques have been used, from the well-known area-perimeter relation (e.g. Meunier 1999a) to structure functions (Abramenko and Longcope 2005) and multi-fractal analysis (Abramenko 2005). Among the results, it has been shown that a single fractal dimension of active regions could not be defined (e.g. Meunier 1999a). However the fractal dimension as a tool can still be used and provide a complementary diagnostic to the size distribution: Meunier (1999a) has shown that the use of the size distribution alone was not sufficient to constrain models of active regions. Note also that fractal analyses have also been performed on velocity structures. Roudier and Muller (1987) have performed a fractal analysis of granulation and shown that below a critical size (which is also seen on the size distribution of granules), the fractal dimension was quite different and compatible with turbulence. Studies of supergranulation (Paniveni et al. 2005) have also provided a comparison with turbulence properties, as they were found to be compatible with isobars in the Kolmogorov turbulence, although with a large uncertainty (see Sect. 3.3). Other multi-scale approaches have been used to characterize both velocity and magnetic structures and their spatial distribution (see for example Cadavid et al. 1998 for the use of space–time spectrum of full disk velocity and magnetogram, and Lawrence et al. 1999 who determined the wavelet flatness spectrum on similar data). All these constraints are complementary. These tools are crucial to characterize the properties of magnetic structures statistically and can be used to test active region

formation models, such as those of Wentzel and Seiden (1992), Seiden and Wentwel (1996) and Schrijver (2001).

3 Small-Scale Flows and Magnetic Fields

3.1 Introduction

In this section we present some recent results concerning small-scale flows, with a focus on supergranulation and flows around active regions. Flows in the photosphere, despite their turbulent nature, are spatially organized at three different scales, granulation (∼1 Mm), mesogranulation (a few Mm), and supergranulation (30 Mm). Granulation has been extensively studied over the past decades and its convective origin associated with a strong radiative cooling is well established. Its main characteristics are: a scale of an order of 1 Mm, a lifetime of an order of 10 minutes and velocity field in the km/s range (e.g. Title et al. 1989). There is of course a large dispersion around these values. Granulation models are now quite sophisticated (e.g. Stein and Nordlund 1998) and also include the effects of magnetic fields (e.g. Carlsson et al. 2004). Two other scales have been observed, mesogranulation and supergranulation, and both are much less understood. Mesogranulation has been detected for the first time by November et al. (1981). Its origin has been questioned, for example by Rieutord et al. (2000) who showed that it is probably not a true convective scale but might be a combination of effects of highly energetic granulation and averaging effects of data processing. The mesogranular scale does not appear either in the power spectrum of flows derived from Dopplergrams (Hathaway et al. 2000) or horizontal velocity fields (Rieutord et al. 2008). Finally, the origin of supergranulation, which was discovered by Hart (1954), has also been questioned, although its existence as a specific scale is much clearer (Hathaway et al. 2000; Rieutord et al. 2008).

In this paper we will focus on one of these non-axisymmetric velocity features, supergranulation. Supergranules are strongly related with the magnetic network, which is located at their boundary, although in an intermittent manner. They play an important role in the diffusion of the magnetic field across the surface and therefore in the dynamo process. The advection of the magnetic features by the photospheric flows such as supergranulation, from active regions to the magnetic network is quite well known but the complete understanding of what happens to the whole magnetic flux from active regions over the solar cycle remains to be resolved. The nature of supergranulation therefore remains an open question, between a convective origin as suggested in the 60's (Simon and Weiss 1968) and large-scale instabilities as proposed more recently by a number of authors (Rieutord et al. 2000; Rast 2003; Rieutord et al. 2008). Recent results on supergranulation will be described in Sect. 3.3.

The coupling between these flows at small scale and the magnetic field is naturally crucial. These flows may play a role in the generation of small-scale fields (e.g. Cattaneo et al. 2003), and in their transport across the surface (e.g. Dikpati et al. 2004). On the other hand, the magnetic field is also expected to have an influence on these flows. This feedback is not always well established, but is very interesting to understand the origin of these flows (supergranulation for example). In a more general context it can also provide information on the influence of magnetic fields on turbulent flows and therefore be of interest for turbulence theories, as such influences have not entirely been investigated: Iroshnikov (1964) and Kraichnan (1965) found that for MHD turbulence in a strong guide magnetic field, the 1/3 exponent of the Kolmogorov law for the velocity field became 1/4, although Goldreich and Sridhar (1995) recently showed that in these anisotropic conditions, the Kolmogorov

exponent remained 1/3 for a velocity field perpendicular to the magnetic field. On the other hand, numerical simulations (e.g. Schekochihin et al. 2004) studying the case of small-scale magnetic field (closer to the small-scale solar magnetic fields) but at much larger Prandtl number than in the photosphere tend to show a very significant increase of the Kolmogorov exponent (up to 1.5). Given the current theoretical works, it is therefore difficult to predict the behavior of supergranulation in presence of a realistic magnetic field.

It is beyond the scope of this paper to describe in detail the properties of small-scale magnetic field in active regions and in the quiet sun, as there are specific papers in this issue, on sunspots (Schlichenmaier 2008; Scharmer 2008) and small-scale magnetic fields (de Wijn et al. 2008; Wedemeyer-Böhm et al. 2008). We will therefore focus on the coupling between supergranulation and the magnetic field (network and intranetwork) in Sect. 3.3 and with active regions in Sect. 3.4 (see also Wedemeyer-Böhm et al. 2008, this issue).

3.2 Observational Techniques

As we are focusing on observations of the solar photosphere, it is interesting to provide a general overview of the various techniques involved, especially in view of the comparison with other approaches (other layers in the atmosphere of the Sun or other stars, see Sect. 5). A huge advantage of the Sun compared to other stars is of course its proximity. This allows us to perform highly detailed studies, in particular of the photosphere, thanks to the high spatial resolution and high photon flux, and therefore high spectral resolution. Two basic techniques have been used intensively to observe the photosphere, i.e., imaging and spectroscopy (and more specifically spectropolarimetry).

High resolution observations have either taken advantage of the natural quality of specific ground-based sites such as La Palma (e.g. Scharmer and Löfdahl 1991; Rouppe van der Vort 2002) and the Pic du Midi Observatory (e.g. Roudier and Muller 1987) and of space facilities such as Hinode (Kosugi et al. 2007), or developed specific techniques to improve the spatial resolution. These include the implementation of tip-tilt or adaptive optics systems (Rimmele 2004; Denker et al. 2005; Puschmann and Sailer 2006; Denker et al. 2007; Wöger et al. 2008) or image processing techniques combined with speckle observations (Paxman et al. 1996; van Noort et al. 2005; Mikurda and von der Lühe 2006).

Recent observations tend to combine as much as possible a large field-of-view with a high spatial resolution and a high temporal cadence, although not all criteria can usually be met at the same time. The purpose is to be able to connect events that are occurring in different active regions, to avoid missing any events for statistical purposes, or simply to study large-scale flows or structures (such as supergranulation or quiescent filaments) in detail. The largest field-of-view allowing to routinely map a large amount of supergranules are, e.g., MDI/SOHO observations (0.6 arcsec/pix, Scherrer et al. 1995) and TRACE (0.5 arcsec/pix, Tarbell et al. 1994), i.e. with a resolution about 5 times worse than the current best observations.

Another approach is the use of spectropolarimetry (e.g. Domínguez Cerdeña et al. 2003 in the visible, Khomenko et al. 2003 in the infrared, to study small-scale intranetwork magnetic fields). Precise spectropolarimetry is usually used on a small field-of-view, as it requests a scan of the studied region. However, magnetograms and Dopplergrams, because of the use of filter-type instruments, are obtained with a spatial resolution similar to the large field-of-view mentioned above. In addition, local helioseismology techniques such as time–distance (Duvall et al. 1993) or ring diagrams (Hill 1990) also allow the study of subsurface structures such as supergranules or active regions over a large field-of-view.

3.3 Supergranulation

As pointed above, the origin of the supergranular scale remains an open question. A number of recent results may give some clues on it and answer to some long-standing questions about the velocity field:

– The observation of a decreasing intensity toward the boundaries of the cells (Meunier et al. 2007b, 2008), as illustrated in Fig. 3, which corresponds to a temperature difference of the order of 0.5–1 K. Previous results were contaminated by the presence of the magnetic network. This trend therefore favors the convective origin of supergranules, as for granules. However, there are also some significant differences with the behavior of granules, and a precise comparison with theoretical works remains to be performed.
– The characterization of the power law relating to the typical velocity fields in supergranules and their size (Krishan et al. 2002, Meunier et al. 2007a, 2008), in relation with the turbulent properties of these flows. These latter results show that it is not consistent with a Kolmogorov turbulence nor with a Bolgiano-Obukhov turbulence corresponding to a stratified medium, as the exponent is larger than expected in both cases. Here again, a comparison with theoretical work must be done.
– The possible interpretation of supergranules as a wavelike feature (Gizon et al. 2003; Schou 2003).

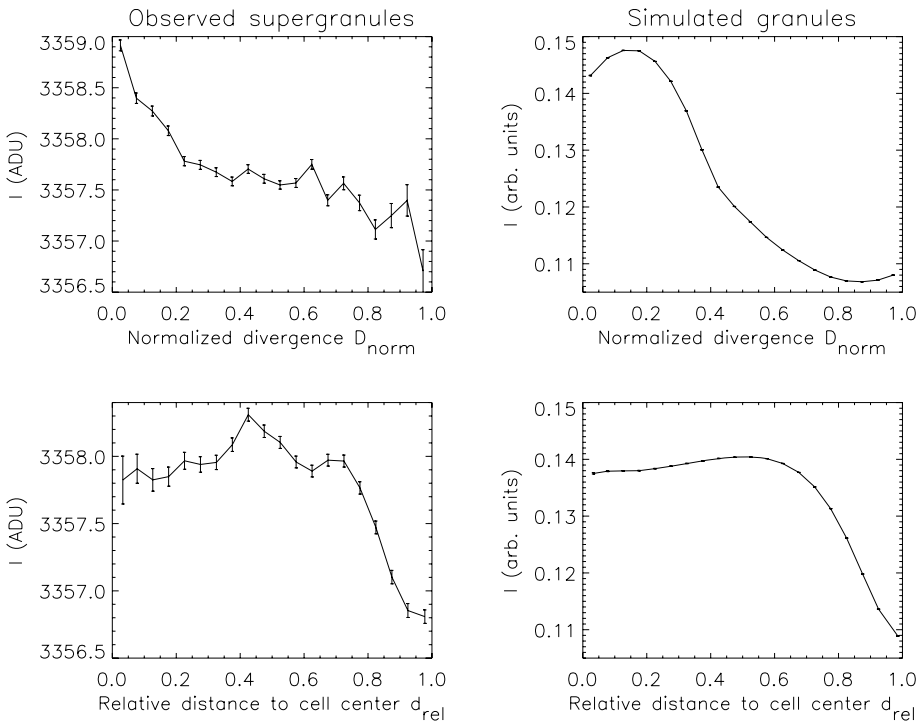


Fig. 3 *Left column*: Intensity variation in supergranules (for pixels such that 80% of the surrounding pixels at the granular scale have a magnetic flux below 3 G) versus the normalized divergence of the horizontal velocity field (*upper panel*) and versus the relative distance to cell centre (*lower panel*), in arbitrary units. *Right panel*: Same for simulated granules. From Meunier et al. (2008)

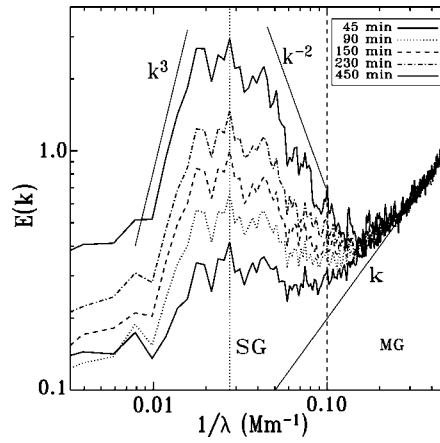


Fig. 4 Kinetic energy spectra obtained for various time windows. The *vertical dotted line* indicates the position of the peak at 36.4 Mm. Two power laws are shown on each side of the peak to give an idea of the slopes. The velocity fields have been derived from granule tracking on intensity images obtained with the CALAS instrument at the Lunette Jean Rösch (Pic du Midi). From Rieutord et al. (2008)

A new instrument has been implemented at the Pic du Midi (CALAS), whose objective is to observe the Sun on a large field of view simultaneously with a high spatial resolution for the first time. Given the turbulent nature of these flows, there is indeed much to gain by considering multi-scale analysis. The horizontal velocity field has been derived from granule tracking on a very large field-of-view for the first time (allowing to free oneself from the projection effects present in Dopplergrams), which led to a typical scale for supergranules of 36 Mm but with a broad peak in the kinetic energy spectrum, as illustrated in Fig. 4. This spectrum has been computed for different temporal windows, which shows that the peak is still small at the 45 minute scale, but significantly larger at the scale of 7.5 hours (maximum for these observations). The decrease for small scales is close to a k^{-2} power law, i.e. steeper than the equipartition Kolmogorov one.

As pointed above, a comparison with numerical simulation experiments is now necessary in order to be able to discriminate among different origins of supergranulation. There are however very few simulations at this scale (Rincon et al. 2005; Benson et al. 2006). If they produce a spatial organization of the flows at a scale larger than granulation, it is not yet clear whether such scales are similar to supergranulation.

Concerning the coupling with magnetic field, some progresses have also been made recently by the study of the magnetic field inside the cells (Meunier et al. 2007c). Previous studies found very different results from each other, from a correlation to anti-correlation or no variation at all of the cell size (Sýkora 1970; Wang 1988; Wang et al. 1996, Hagenaar et al. 1997; Raju and Singh 2002). Most of these studies have determined supergranules using Ca II K images or magnetograms, therefore this determination depended on the magnetic field or proxies of the magnetic field. DeRosa and Toomre (2004) have determined them independently, using diverging flows, but considered only two short observations during the solar cycle. Meunier et al. (2007c) have found that the variation of the cell size with the local magnetic field depends on the magnetic component one is considering (typically the network or the intranetwork fields), as illustrated in Fig. 5. Strong magnetic flux inside the cells associated with smaller cells, while the strong magnetic network is associated with large cells, probably due to their longer lifetime. Different sensitivity thresholds could then explain the variety of previous results. Meunier et al. (2008), in a study covering a complete cycle, have also found supergranules to be smaller at cycle minimum (Fig. 6). This confirms previous results obtained by DeRosa and Toomre (2004) based on two short time series. Such a trend can be expected from the influence of the Lorentz force.

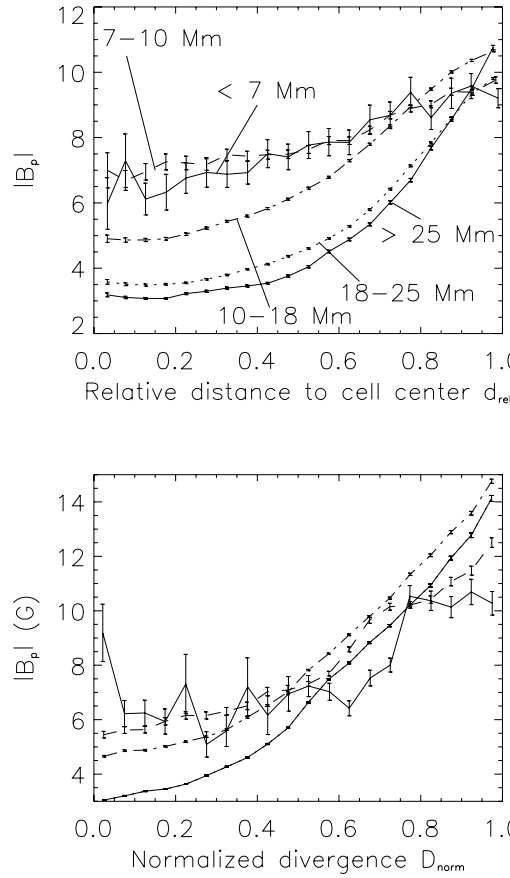


Fig. 5 *Upper panel:* absolute value of magnetic field at each pixel $|B_p|$ versus the relative distance to cell center d_{rel}, for various size ranges: R lower than 7 Mm (*solid line*), in the range 7–10 Mm (*dashed line*), in the range 10–18 Mm (*dotted-dashed line*), in the range 18–25 Mm (*dotted line*), and larger than 25 Mm (*dot-dot-dot-dashed line*). *Lower panel:* same versus the normalized smoothed divergence of the horizontal velocity field D_{norm}. From Meunier et al. (2007c)

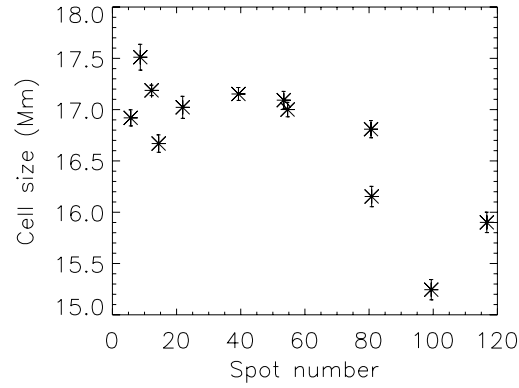


Fig. 6 Average cell size R (in Mm) as a function of the monthly sunspot number. From Meunier et al. (2008)

It is also in agreement with the simulation of the magnetic network made by Crouch et al. (2007). There is however a lack of theoretical work on the behavior of turbulence in the presence of magnetic field, especially in cases that would be comparable to the solar case, which is neither a uniform magnetic field (due to the internetwork component) nor a completely turbulent isotropic magnetic field (due to the intermittent network, constituted

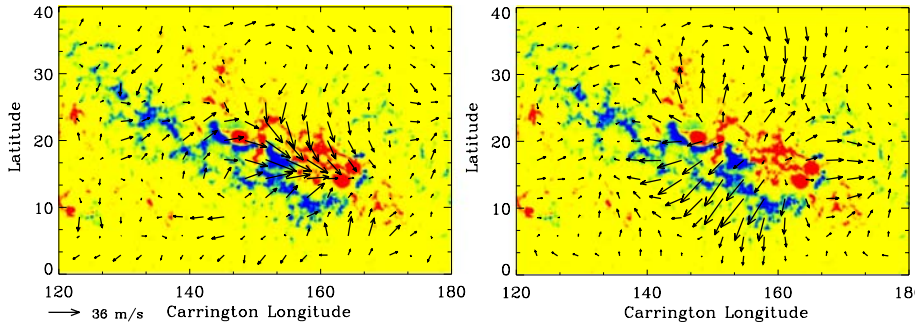


Fig. 7 Large scale average flow maps for a large active region, AR9433, at two different depth intervals: 0–3 Mm (*left*) and 9–12 Mm (*right*)

of strong magnetic fields field with mostly vertical field lines). Numerical simulations allowing to produce scales larger than granules are not advanced enough to be able to include the magnetic field. By comparison, it is interesting to note that the size variation of granules with the activity level, although quite well-known at a given time between quiet sun and plages (Title et al. 1992), is very difficult to estimate over the solar cycle (Muller et al. 2007) and seems to be very small. In addition to the supergranule size, the slope mentioned above relating flows with cell sizes is also found to be dependent on the local magnetic field, which shows that supergranules are not a purely hydrodynamical velocity field.

3.4 Flows around Active Regions

Large scale flows around large active regions are important for understanding the formation, evolution, and sometimes decay of those regions. It may also be relevant to the flux transport as mentioned above (supergranulation). By use of time–distance helioseismology, Zhao and Kosovichev (2004) derived the large scale flow maps around one large active region, AR9433, during the region's passage of the central meridian in April of 2001. As shown in Fig. 7, near the solar surface at a depth of 0–3 Mm, converging flows can be found toward the neutral line of this huge active region, with a typical speed of approximately 40 m/s. This is generally consistent with the finding of converging flow pattern toward sunspot center (Zhao et al. 2001), but with much smaller speed. Large scale divergent flow patterns are only found beneath 9 Mm, much deeper than the depth of where converging flows are seen underneath sunspots. Haber et al. (2004) analyzed the same active region during approximately the same period using ring-diagram analysis. Similar flow patterns were also reported for this region (see also Gizon et al. 2008 in this issue).

4 Global-Scale Flows and Magnetic Fields

4.1 Introduction

In this section we focus on the global flows, i.e. axisymmetric flows, and mainly torsional oscillations and meridional circulation. If differential rotation has been extensively studied, the knowledge of torsional oscillations and meridional circulation has been much improved

during the last decade, either concerning their characterization in the photosphere and below the surface, and their variation during the solar cycle. We will emphasize a few recent observationnal results and refer to Rempel and Brun (2008) in this issue for more theoretical considerations. Among the open questions we will detail are: why do structures of different types (i.e. either magnetic or velocity structures) rotates at different speeds (and more generally exhibit a different dynamics)? What is the depth extent of torsionnal oscillations and meridional circulation? What are their temporal variations and their spatio-temporal organization? What is the impact of these properties on dynamo models?

Although we focus mostly on flows in this section, we also give a brief review on solar irradiance studies, as most works attempts to reconstruct irradiance variations using photospheric magnetic structures.

4.2 Observational Techniques

Most of the large-scale analyses have been performed on full-disk data, such as provided by MDI/SOHO with 2 arcsec/pixels. The flows have been derived using various techniques, e.g., feature tracking (e.g. Meunier 2005a) correlation tracking (e.g. Komm et al. 1993; Meunier 1999a), Doppler measurements (e.g. Snodgrass and Ulrich 1990) and helioseismology (e.g. Deubner and Gough 1984; Thompson et al. 1996). Different techniques can be sensitive to different components of the photosphere (non-magnetic plasma, spots, etc.). They are also sensitive to different scales: for example, cross-correlation of the magnetic network at the supergranular scale will probe a structure that is different from the tracking of small-scale features (e.g. Meunier 2005a).

4.3 Differential Rotation

Global scale flows, especially the rotation, have been studied using many magnetic or velocity tracers, and using different techniques (see above). There is a large variability in the results depending on many factors, such as method and type of structures, age, lifetime and size of structures, polarities, and cycle phase (see Beck 2000 for a review). Spots are probably the structures that have been the most extensively studied in that respect (e.g. Zappala and Zuccarello 1991; Zuccarello 1993; Ruždjak et al. 2004). Magnetic structures such as sunspots are typically rotating 2% faster than the surrounding plasma, while supergranules rotate 1 or 2% faster than the magnetic structures.

A classical interpretation of this dispersion (with variations within a few percent at the equator for the rotation) is that different structures could be anchored at different depths in the convective zone: being coupled to layers rotating at a different velocity (we know from helioseismology that there is a gradient just below the surface), they would themselves exhibit a dynamics different from the photospheric one. It may work well for young sunspots, as shown by Javaraiah and Gokhale (1997). However, it may not apply to other structures, for example the magnetic network: Meunier (2005a) has shown that in the case of network, the combined determination of differential rotation and meridional circulation was not compatible with that explanation. Some mechanism must therefore be found in order to accelerate the magnetic features with respect to the surrounding plasma. The nature of such a mechanism remains an open question. For this kind of study the coupling with the internal dynamics is of course crucial.

This coupling is probably very important for supergranulation as well. Indeed we can relate some properties of supergranules with the global dynamics. Some results obtained by different techniques showing some variability of the differential rotation with the polarity of magnetic features, namely leading and following polarities (Zhao et al. 2004;

Meunier 2005b), or the super-rotation of supergranules with respect to the magnetic network in particular, remain to be understood. The super-rotation of supergranules in particular, detected for the first time by Duvall (1980) and later confirmed by Snodgrass and Ulrich (1990), has been challenged by Hathaway et al. (2006) because these determinations were made using Dopplergrams. The analysis of such data is however subject to very strong projection effects, which can bias the determination of the velocity fields. However, Meunier and Roudier (2007d) have recently confirmed this super-rotation of supergranules using a projection-free technique. The interpretation of this super-rotation remains to be understood. The possible interpretation of supergranules as a wave (Gizon et al. 2003; Schou 2003) could be a solution, but this has been questioned by Rast et al. (2004).

4.4 Torsional Oscillation

In addition to this smooth differential rotation, there are bands of plasma at particular latitudes rotating either slightly faster or slower, as first reported by Howard and LaBonte (1980) by analyzing photospheric rotation patterns. These bands also shift positions as the solar cycle progresses, migrating towards the equator together with the solar activity zones in both hemispheres. Later helioseismological analyses have demonstrated that the torsional oscillation do not just stay in surface, but penetrates into the deep interior of the solar convection zone (e.g. Kosovichev and Schou 1997; Howe et al. 2000; Vorontsov et al. 2002). Furthermore, an additional strong, poleward branch was also revealed; this branch appears to penetrate the entire convection zone. The amplitudes and phases of these migrating bands show a systematic variation with position in the convection zone (Howe et al. 2005). To interpret the observed torsional oscillation, Schüssler (1981) and Yoshimura (1981) proposed a Lorentz force feedback, and Spruit (2003) tried to explain them by thermal driving. However, Rempel (2007) found by numerical simulation that although the poleward high latitude branch could be explained by Lorentz force feedback or thermal driving, the low latitude equatorward branch likely had a thermal origin. For a more detailed review on this subject, please see Thompson and Weiss (2008) of this issue.

4.5 Poleward Meridional Flow

Compared with the solar differential rotation, solar meridional flows have relatively smaller speed (2 orders of magnitude weaker), hence they are a bit more difficult to detect. Despite some confusion of earlier years, poleward meridional flow of an order of 20 m/s was finally observed by analyzing Doppler velocities at the solar photosphere obtained from Stanford Solar Observatory (Duvall 1979). Similar but more robust results were later found by analyzing Doppler velocities obtained using GONG observations (Hathaway et al. 1996). The poleward meridional flow result was also confirmed by tracking specific features on the solar surface, such as sunspots (e.g. Howard and Gilman 1986), small magnetic features (Komm et al. 1993), and by doing correlation tracking of MDI magnetograms (Meunier 1999b) and MDI Dopplergrams (Švanda et al. 2007) Therefore, meridional flow with a speed of 20 m/s or so and with a poleward direction has been well established at the solar photospheric level.

The development of helioseismology has given solar physicists a chance to detect flow fields inside the solar interior. By use of a local helioseismology technique, namely time–distance helioseismology, Duvall et al. (1993) and Giles et al. (1997) reported that the poleward meridional flow, as well, did not just stay at the solar photospheric level, but also penetrated into the solar convection zone to a depth of at least 0.04 $R_{\odot}$, with a peak speed of around 10–20 m/s. The meridional flows in the deep interior of this kind would help to redistribute angular momentum within the Sun.

That the poleward meridional flow extends to at least 0.04 $R_\odot$ into the solar interior was also well established by use of different helioseismological analysis techniques, e.g., ring-diagram analysis (González Hernández et al. 1999; Haber et al. 2002; Basu et al. 2004), time–distance helioseismology (Gizon 2003; Zhao and Kosovichev 2004), and analysis of acoustic frequency shifts (Braun and Fan 1998; Krieger et al. 2007; Mitra-Kraev and Thompson 2007). All these results were essentially in agreement with the results reported by Giles et al. (1997). Now, the question facing helioseismologists is how deep these poleward meridional flows would penetrate, and where and how large the return meridional flows would be (see Sect. 4.7).

The poleward meridional flows in the outer solar convection zone are crucially important to the flux transport theory, hence the interpretation of solar activity cycles (Wang et al. 1991). The accurate determination of the meridional flow profile through the solar convection zone is also essential to the solar dynamo simulations (e.g. Dikpati and Gilman 2006).

4.6 Solar Cycle Variations of the Flows

The solar differential rotation profile does not stay the same over the course of one whole solar cycle, and mixed faster and slower bands relative to a smooth differential background move towards the solar equator as the solar cycle progresses, which is known as torsional oscillation.

There is also a general agreement that the Sun rotates more rigidly at cycle maximum (Balthasar and Wöhl 1980; Lustig 1983; Nesme-Ribes et al. 1993; Brajša et al. 2006). Solar meridional flows do not stay the same during the solar cycle course, either. This has been already realized in earlier years (Snodgrass 1987; Komm et al. 1993; Hathaway et al. 1996) for the photospheric level flow patterns, and has also been systematically studied by, e.g., Meunier (2005c), and Švanda et al. (2008). Chou and Dai (2001) studied the subsurface meridional flows as a function of latitude and depth for the period of 1994 to 2000 using time–distance helioseismology. They found that a new component of meridional flow, centered at about 20° latitude, was created in each hemisphere as the solar activity increased from 1997 to 2000. Beck et al. (2002) did similar studies using the same technique, but with more continuous data coverage. They also found one extra time-varying component that had a banded structure matching the torsional oscillations with an equatorward migration over the solar cycle. The time-varying component of meridional flow consists of a flow diverging from the dominant latitude of magnetic activity. Both studies targeted at the deeper interior of the Sun.

The near surface helioseismological studies confirmed the time variations of meridional flows. Using ring-diagram analysis, Haber et al. (2002) found out that the gradient of the near-equator meridional flows steepened with the development of the solar cycle toward the solar maximum. Employing time–distance technique, Zhao and Kosovichev (2004) subtracted the mean meridional flow profile of 1996, a solar minimum year, as a reference, and found out that the residual meridional flows actually converged toward the solar activity belts (see Fig. 8). Using meridional flow profiles derived from photospheric supergranulation patterns, Gizon and Duvall (2004) found similar converging residual meridional flow patterns. The converging meridional flow toward the activity belts was believed caused by thermal driving, as suggested by Spruit (2003) and further modeled by Rempel (2006) taking into consideration of Lorentz force feedback. As discussed by Meunier (2005c), all results are not yet entirely in agreement with each other, especially at high latitude, and meridional circulation seems to exhibit a behavior that is more complex than differential rotation. This may be due to the fact the meridional circulation is driven by small differences between

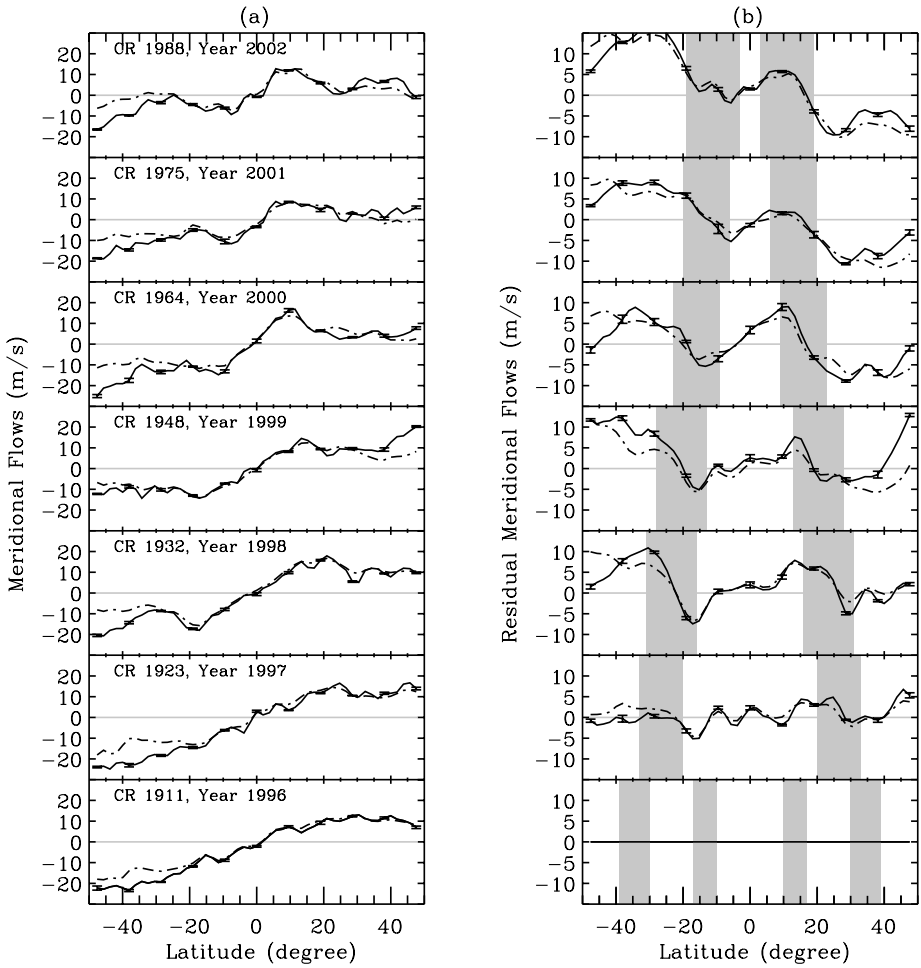


Fig. 8 (**a**) Meridional flows obtained from 3–4.5 Mm (*solid curves*) and 6–9 Mm (*dash-dotted curves*) for different Carrington rotations. (**b**) Residual meridional flows after the flows of CR1911 have been subtracted from each rotation. *Shaded regions* indicate the location of activity belts

large forces that are nearly in balance (Miesch 2005). This is crucial as the meridional circulation is also playing a key role in flux transport models, for example in the well-known equatorward motion of the dynamo wave (e.g. Dikpati et al. 2004). The temporal variation should therefore influences the amplitude of the cycles (see Dikpati and Gilman 2008 in this issue) and it is therefore necessary to have a reliable determination of these variations.

4.7 Search for Return Meridional Flows

It is quite clear that the poleward meridional flow exists at the solar photosphere and extends to some depth into the convection zone, yet it is still not quite clear what the meridional flow profiles are inside the deeper interior. It is even not clear how many circulation cells exist in one meridional plane in the direction of depth, or how many cells there are in the latitudinal

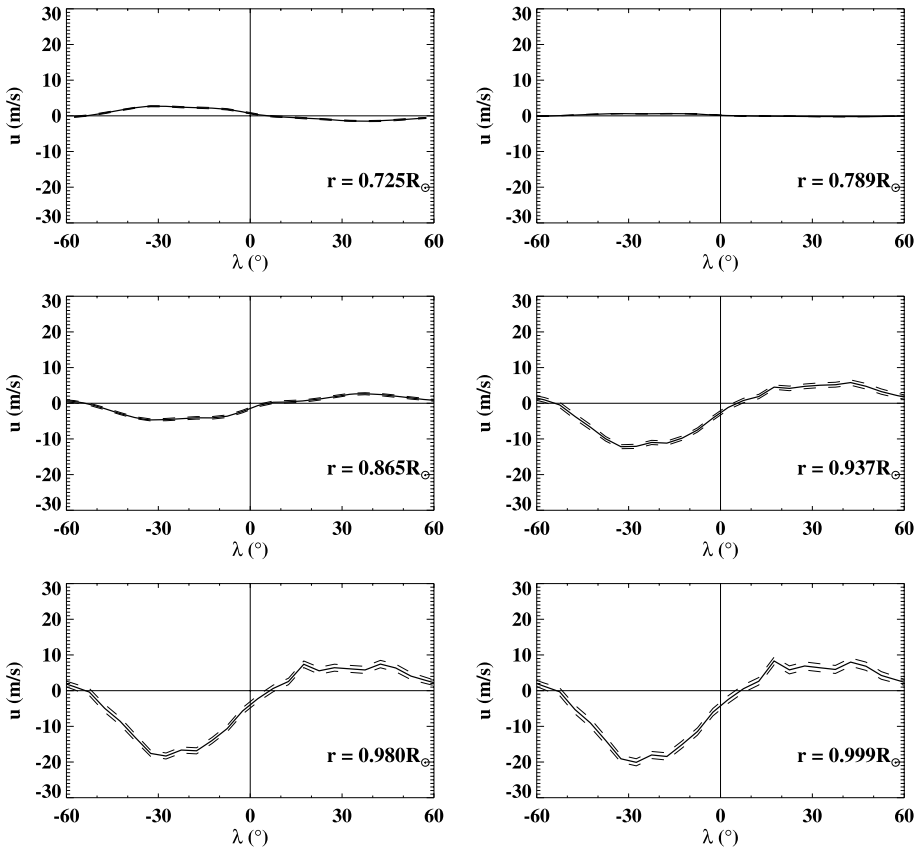


Fig. 9 The meridional circulation inversion results, as a function of latitude λ, for 6 different depths. The inversion was performed with a constraint of mass conservation inside the solar convection zone, and an assumption of that velocity is 0 below 0.71 $R_\odot$. Positive velocities are northward. This figure is adopted from Giles (1999)

direction. Numerical simulations have shown some interesting results, for example, multi-cell meridional flows can hardly persist a long time inside the Sun without changing the eventual butterfly diagram that should match the observed one (Jouve and Brun 2007).

Giles (1999) has measured acoustic travel times between two photospheric locations very far apart, hence being able to derive the meridional flow profiles in the very deep interior of the Sun. He found that if inverting the flow profiles only using acoustic travel times without further constraints, the poleward meridional flow profile would extend to the tachocline area, which is located at the bottom of the solar convection zone. However, if a mass conservation and a zero velocity below the convection zone were forced in the inversion procedure, a return meridional flow, i.e., equatorward flow, would be found with a magnitude of 2 m/s or so, as shown in Fig. 9.

More recently, by analyzing acoustic frequency shifts observed by MDI, Mitra-Kraev and Thompson (2007) derived meridional flow as a function of depth through nearly the whole convection zone, without a determination of latitudinal dependence. They found an equatorward return meridional flow at approximately 0.95 $R_\odot$. However, this is not in agreement with the results of Giles (1999), and is also at odds with previous re-

sults that were obtained using the similar analysis technique (Braun and Fan 1998; Krieger et al. 2007).

In addition to the direct measurements using helioseismic analyses, researchers also tried to search evidence of return meridional flows using indirect approaches. Hathaway et al. (2003) examined the drift of centroid of the sunspot area toward the equator in each hemisphere from 1874 to 2002, and they found that these observations were consistent with a meridional counterflow deep within the Sun. That equatorward flow had approximately an amplitude of 1.2 m/s at the base of the convection zone, which was in good agreement with results derived by Giles (1999).

4.8 Solar Irradiance Variations

We now come back to another aspect of the global properties of the magnetic field in the photosphere, namely their contribution to the temporal variation of solar irradiance. Most of the variations of solar irradiance are indeed due to the presence of dark sunspots and bright plages at the surface, therefore are related to magnetic activity (see Jones et al. 2008 for a classification of the different structures and their influence on irradiance). Different approaches can be used to study these variations and various contributions (see e.g. Solanki and Krivova 2004 for a review).

A first category of works consider reconstructions of the irradiance for the 2–3 last cycles, i.e. when direct measurements of this irradiance are available (e.g. Fröhlich 2006). Such studies allow to test the quality of the reconstruction. They can be based on regressions of various proxies (e.g. Foukal and Lean 1988; Chapman et al. 1996; Fröhlich and Lean 1998; Fligge et al. 1998). More sophisticated models takes into account various components using maps of the different contributions (spots, plages, ...) such as in Solanki and Fligge (1999). Wenzler et al. (2005) and Krivova et al. (2006) have obtained an unbiased reconstruction, which shows that magnetic features account for all observed variations. Minarovjech et al. (2007) have tested the use of the coronal index and the Mg II index and studied in detail the correlation with the observed irradiance. This kind of approach allows to test the contribution of various components, for example the quiet sun (Withbroe 2006). It is also of great interest to consider information in the UV part of the spectrum, such as done by Dudok de Wit et al. (2008) or Woods (2008), because the observed variations are much larger and the impact on the earth expected to be very significant.

A second category of works uses models tested on the short time scale in order to reconstruct the solar irradiance over a longer period, typically back to $\sim$1700, during which direct measurements were not available. This allows to test the respective influence of the Sun and human activities on the earth climate over the last century. An example is given by the work of Solanki and Krivova (2003): with the assumption that the Sun explains the observed earth temperature variation between 1856 and 1970, they found an upper limit for the solar contribution since 1970 of 50%. Based on active region decay, Crouch et al. (2008) have reconstructed the solar irradiance back to 1874. The most sophisticated models take into account various components and non-linearities between them, including time delays (e.g. Preminger and Walton 2006a, 2006b) or flux transport (Wang et al. 2005). These models can either reproduce cyclic variations or include a secular trend (e.g. Foster 2004 for such a model). Another period that has been extensively studied is the Maunder minimum, during which there was a very low activity level, with a very strong North–South asymmetry (Sokoloff and Nesme-Ribes 1994). There are strong indications that during these few decades the cycle was still present, as obtained for example by Beer et al. (1998) using ^{10}Be measurements. The rotation has also been observed to be weaker (Ribes and Nesme-Ribes

1993). Several works have attempted to build a reconstruction of solar irradiance since the Maunder minimum (e.g. Tapping et al. 2007; Krivova et al. 2007) and they obtained a difference of the order of 1 W/m^2 between the level during the Maunder minimum and the current level. The organization of the grand minima over a longer time scale remains an open question and is of course of great interest for dynamo models.

5 Stellar Activity

As already pointed out, thanks to its proximity, it is possible to observe the Sun with a very high spatial resolution associated to a large field of view, a high flux and a high temporal cadence. Multi-wavelength observations are also possible in good conditions (thanks to the large flux). And, very important as well, there is also a long history of observations as far as the photosphere is concerned, several centuries for systematic observations of sunspots for example, but even longer when using proxies of solar activity on earth.

However, it is only one realization, or, to a lesser extent, only one point of view (mainly from the equator, as the only observations from the polar perspective have been made by Ulysses). The detailed study of many other stars would permit the exploration of the space of parameters, and study the influence of various parameters such as their age, rotation and mass on the activity level. Due to the limited flux and spatial resolution, only a limited number of constraints can be determined, and most of them are 1D observations, such as the Ca II K emission over time (Baliunas et al. 1995). Attempts to derive 2D information from such time series are however promising (Baliunas et al. 2006). Detailed butterfly diagrams, which provide an excellent constraint for dynamo models, can not be derived, but it is possible to get some information on the direction of the dynamo wave for example.

Some very interesting results have also been obtained using spectropolarimetric observations and Zeeman-Doppler imaging techniques in the more recent years applied to a variety of stars (e.g. Donati et al. 1997; Brown et al. 1998). Again, because of the lack of resolution, it is crucial to use complementary techniques, which is reminiscent of the filling factor issue in case of solar weak fields (see de Wijn et al. 2008, this issue), as features of opposite polarities observed with the Zeeman effect cancel out if the resolution is too low.

Some (not exhaustive) interesting issues (but open questions) in stellar physics that are of high interest for solar activity are, for example:

- How many stars are in a “Maunder minimum” state (or more generally speaking, a grand minimum)?
- How the relative importance of plages and sunspots vary with age, as observed by Lockwood et al. (2007); this may be of interest for solar irradiance variation as well.
- What are the spot properties on other stars? The spot temperature contrast as a function of the photospheric temperature has been compared to the solar case (e.g., Berdyugina 2005).

Finally, Donati et al. (2003) have observed the star AB Dor, and found a temporal variation of the differential rotation (over a few years), and also a difference between the rotation rate of intensity features determined by Stokes I observations and the magnetic features determined by stokes V observations. This is reminiscent of what we have described about the solar rotation in Sect. 4.3, as the precise result depends on the technique, type of structure and cycle phase. These observations therefore remain to be understood as the anchoring depth hypothesis may not be valid in the solar case.

6 Conclusion

In this review we have shown new observational results concerning either small-scale flows or large-scale flows and magnetic fields.

Concerning small-scale flows, new results on supergranulation have answered some long-standing open questions, such as the intensity variation, and the cycle-size relationship. They have also confirmed the super-rotation of supergranules with respect to the magnetic network. However, these results do not yet provide a completely coherent picture, and there is not yet a complete understanding of supergranulation, either pointing toward a convective origin or a large-scale instabilities related to explosive granules. What is missing is a realistic numerical simulation with enough spatial resolution to resolve granules but also with a size large enough to allow the supergranular scale to naturally emerge. The addition of a realistic magnetic field in such simulations would be a further step. With such simulations, it would then be easy to compare the intensity profile in supergranules with the observed one, as well as the various power laws that have been observed. On the other hand, 3D global simulations (see Rempel and Brun 2008 in this issue), are another way to approach the understanding of supergranulation as well in the future.

On the observational point of view, we see two ways which should be investigated. The first one is to move toward large-scale observations while keeping a high spatial resolution in order to keep as much as possible of the granular information, because firstly, this allows a reliable determination of flow fields, and secondly their behavior, in particular explosive granules, could be at the origin of supergranules. Large-scale observations are however necessary if a large-scale instability must be shown. The second solution will be to explore the relationship between magnetic field and supergranules using high resolution observations. With Hinode, the complexity of small-scale fields have been investigated in detail (Centeno et al. 2007; Orozco Suárez et al. 2007; Lites et al. 2008), and their organization at the scale of supergranule should now be studied (see Meunier et al. 2007c, 2008 for such a study at low spatial resolution).

The knowledge of global flows has also made a huge step during the cycle-long SOHO observations, using various complementary techniques. The characterization of flows such as the torsional oscillation or the meridional circulation has been much improved. More specifically, in addition to a better determination of the surface flows, the coupling with the solar interior has been investigated thanks to time–distance helioseismology and other helioseismology techniques. For example, we now know that torsional oscillations extend deep into the convective zone. Another striking result is the strong variability of the meridional circulation. The correlation of this variation with the solar cycle is not as clear as for the rotation, and this has a strong impact on dynamo models. The future monitoring of these flows on a longer time scale (in particular with SDO) will therefore be crucial. Dynamo models such as the flux transport model will also have to take this variability into account in order to make realistic predictions of cycle amplitudes. The fact that they are already giving good predictions without taking it into account is however puzzling. Along the same line, some results are yet to be understood, for example, the variation of the rotation depending on the polarity of the magnetic network (leading or following polarities), and the fact that this variability, among others, can not be merely explained by the anchoring of the corresponding flux tube deeper in the convective zone, as previously thought. The association of surface flows determination and techniques such as helioseismology (following different approaches) has made such advances possible and will continue in the future, mostly with SDO.

Finally, going beyond the original scope of this paper, we have also emphasized the inputs to gain from stellar observations, and especially from the new results obtained through

spectropolarimetry. It is very interesting that the connection works both ways. For example, the observations of dynamos acting in different conditions may help to finalize a model of the solar dynamo. Vice versa, knowing the details on the solar surface at any given time may allow to test techniques aiming at retrieving spatial information from integrated signals on other stars.

References

V.I. Abramenko, Sol. Phys. **228**, 29 (2005)
V.I. Abramenko, D.W. Longcope, Astrophys. J. **619**, 1160 (2005)
F. Albregtsen, P. Maltby, Sol. Phys. **71**, 269 (1981)
F. Albregtsen, P.B. Joras, P. Maltby, Sol. Phys. **90**, 17 (1984)
S.L. Baliunas, R.A. Donahue, W.H. Soon et al., Astrophys. J. **438**, 269269 (1995)
S. Baliunas, P. Frick, D. Moss, E. Popova, D. Sokoloff, W. Soon, Mon. Not. R. Astron. Soc. **365**, 181 (2006)
H. Balthasar, H. Wöhl, Astron. Astrophys. **92**, 111 (1980)
S. Basu, H.M. Antia, R.S. Bogart, Astrophys. J. **610**, 1157 (2004)
J. Beck, Sol. Phys. **191**, 47 (2000)
J.G. Beck, L. Gizon, T.L. Duvall, Astrophys. J. Lett. **575**, L47 (2002)
J. Beer, S. Tobias, N. Weiss, Sol. Phys. **181**, 237 (1998)
D. Benson, R. Stein, Å. Nordlund, in *Solar MHD Theory and Observations: A High Spatial Resolution Perspective*, ed. by J. Leibacher, R.F. Stein, H. Uitenbroek. ASP Conference Series, vol. 354, Proceedings of the Conference Held 18–22 July, 2005, at the National Solar Observatory, Sacramento Peak, Sunspot, New Mexico, USA (Astronomical Society of the Pacific, San Francisco, 2006), p. 92
S.V. Berdyugina, Living Rev. Sol. Phys. **2**, 8 (2005)
S.V. Berdyugina, I.G. Usoskin, Astron. Astrophys. **405**, 1121 (2003)
L. Bharti, C. Joshi, S.N.A. Jaaffrey, Astrophys. J. **669**, 57 (2007)
R. Brajša, D. Ruždjak, H. Wöhl, Sol. Phys. **237**, 365 (2006)
D.C. Braun, Y. Fan, Astrophys. J. Lett. **508**, L105 (1998)
S.F. Brown, A. Collier Cameron, Y.C. Unruh, J.-F. Donati, G.A.J. Hussain, Mon. Not. R. Astron. Soc. **299**, 904 (1998)
A.C. Cadavid, J.K. Lawrence, A.A. Ruzmaikin, S.R. Walton, T. Tarbell, Astrophys. J. **509**, 918 (1998)
M. Carlsson, R.F. Stein, Å. Nordlund, G. Scharmer, Astrophys. J. **610**, 137 (2004)
F. Cattaneo, T. Emonet, N. Weiss, Astrophys. J. **588**, 1183 (2003)
R. Centeno, H. Socas-Navarro, B. Lites, M. Kubo, Z. Frank, R. Shine, T. Tarbell, A. Title, K. Ichimoto, S. Tsuneta, Y. Katsukawa, Y. Suematsu, T. Shimizu, S. Nagata, Astrophys. J. **666**, 137 (2007)
G.A. Chapman, A.M. Cookson, J.J. Dobias, J. Geophys. Res. **101**, 13541 (1996)
D.-Y. Chou, D.-C. Dai, Astrophys. J. Letters **559**, L175 (2001)
A.D. Crouch, P. Charbonneau, K. Thibault, Astrophys. J. **662**, 715 (2007)
A.D. Crouch, P. Charbonneau, G. Beaubien, D. Paquin-Ricard, Astrophys. J. **677**, 723 (2008)
C. Denker, D. Mascarinas, Y. Xu, W. Cao, G. Yang, H. Wang, P.R. Goode, T. Rimmele, Sol. Phys. **227**, 217 (2005)
C. Denker, A. Tritschler, T. Rimmele, K. Richards, S.L. Hegwer, F. Wöger, Publ. Astron. Soc. Pac. **119**, 170 (2007)
M.L. DeRosa, J. Toomre, Astrophys. J. **616**, 1242 (2004)
F.-L. Deubner, D. Gough, Annu. Rev. Astron. Astrophys. **22**, 593 (1984)
M. Dikpati, P.A. Gilman, Astrophys. J. **649**, 498 (2006)
M. Dikpati, P. Gilman, Space Sci. Rev. (2008, this issue)
M. Dikpati, G. de Toma, P.A. Gilman, N. Arge, O.R. White, Astrophys. J. **601**, 1136 (2004)
I. Domínguez Cerdeña, J. Sánchez Almeida, F. Kneer, Astron. Astrophys. **407**, 741 (2003)
J.-F. Donati, M. Semel, B.D. Carter, D.E. Rees, A.C. Cameron, Mon. Not. R. Astron. Soc. **291**, 658 (1997)
J.-F. Donati, A. Collier-Cameron, P. Petit, Mon. Not. R. Astron. Soc. **345**, 1187 (2003)
T.L. Jr. Duvall, Sol. Phys. **63**, 3 (1979)
T. Dudok de Wit, M. Kretzschmar, J. Aboudarham, P.-O. Amblard, F. Auchère, J. Lilensten, Adv. Space Res. **42**, 903 (2008)
T.L. Duvall, Sol. Phys. **66**, 213 (1980)
T.L. Jr. Duvall, S.M. Jefferies, J.W. Harvey, M.A. Pomerantz, Nature **362**, 430 (1993)
M. Fligge, S.K. Solanski, Y.C. Unruh, C. Fröhlich, C. Wehrli, Astron. Astrophys. **335**, 709 (1998)

S. Foster, Reconstruction of solar irradiance variations for use in studies of global climate change: application of recent SOHO observations with historic date from Greenwich Observatory. Ph.D. Thesis (2004)
P. Foukal, J. Lean, Astrophys. J. **328**, 347 (1988)
C. Fröhlich, J. Lean, Geophys. Res. Lett. **25**, 4377 (1998)
C. Fröhlich, Space Sci. Rev. **125**, 53 (2006)
V. Gaizauskas, K.L. Harvey, J.W. Harvey, C. Zwaan, Astrophys. J. **265**, 1056 (1983)
P.M. Giles, Ph.D. Thesis Stanford Univer. (1999)
P.M. Giles, T.L. Jr. Duvall, P.H. Scherrer, R.S. Bogart, Nature **390**, 52 (1997)
L. Gizon, Ph.D. Thesis, Stanford Univer. (2003)
L. Gizon, T.L. Duvall, J. Schou, Nature **421**, 43 (2003)
L. Gizon, T.L. Jr. Duvall, in Multi-wavelength Investigations of Solar Activity, ed. by A.V. Stepanov, E.E. Benevolenskaya, A.G. Kosovichev (Cambridge University Press), IAU Symp. **223**, 41 (2004)
L. Gizon et al., Space Sci. Rev. (2008, this issue)
P. Goldreich, S. Sridhar, Astrophys. J. **438**, 763 (1995)
I. González Hernández, J. Patrón, R.S. Bogart, SOI Ring-Diagrams Team. Astrophys. J. Lett. **510**, L153 (1999)
D.A. Haber, B.W. Hindman, J. Toomre, R.S. Bogart, R.M. Larsen, F. Hill, Astrophys. J. **570**, 855 (2002)
D.A. Haber, B.W. Hindman, J. Toomre, M.J. Thompson, Sol. Phys. **220**, 371 (2004)
H.J. Hagenaar, C.J. Schrijver, A.M. Title, Astrophys. J. **481**, 988 (1997)
A.B. Hart, Mon. Not. R. Astron. Soc. **114**, 17 (1954)
K.L. Harvey, PhD Thesis (1993)
D.H. Hathaway et al., Science **272**, 1284 (1996)
D.H. Hathaway, J.G. Beck, R.S. Bogart, K.T. Bachmann, G. Khatri, J.M. Petitto, S. Han, J. Raymond, Sol. Phys. **193**, 299 (2000)
D.H. Hathaway, D. Nandy, R.M. Wilson, E.J. Reichmann, Astrophys. J. **589**, 665 (2003)
D.H. Hathaway, P.E. Williams, M. Cuntz, Astrophys. J. **644**, 598 (2006)
F. Hill, Sol. Phys. **128**, 321 (1990)
R.F. Howard, Sol. Phys. **137**, 205 (1992)
R.F. Howard, B.J. LaBonte, Astrophys. J. Lett. **239**, L33 (1980)
R.F. Howard, P.A. Gilman, Astrophys. J. **307**, 389 (1986)
R. Howe, J. Christensen-Dalsgaard, F. Hill, R.W. Komm, R.M. Larsen, J. Schou, M.J. Thompson, J. Toomre, Astrophys. J. Lett. **533**, L163 (2000)
R. Howe, J. Christensen-Dalsgaard, F. Hill, R.W. Komm, J. Schou, M.J. Thompson, Astrophys. J. **634**, 1405 (2005)
P.S. Iroshnikov, Sov. Astron. **7**, 566 (1964)
J. Javaraiah, M.H. Gokhale, Astron. Astrophys. **327**, 795 (1997)
H.F. Jones, G.A. Chapman, K.L. Harvey, J.M. Pap, D.G. Preminger, M.J. Turmon, S.R. Walton, Sol. Phys. **248**, 323 (2008)
L. Jouve, A.S. Brun, Astron. Astrophys. **474**, 239 (2007)
E.V. Khomenko, M. Collados, S.K. Solanki, A. Lagg, J. Trujillo Bueno, Astron. Astrophys. **4408**, 1115 (2003)
T. Kosugi, K. Matsuzaki, T. Sakao, T. Shimizu, Y. Sone, S. Tachikawa, T. Hashimoto, K. Minesugi, A. Ohnishi, T. Yamada et al., Sol. Phys. **243**, 3 (2007)
L. Krieger, M. Roth, O. von der Lühe, Astronomische Nachrichten **328**, 252 (2007)
N.A. Krivova, S.K. Solanki, L. Floyd, Astron. Astrophys. **452**, 631 (2006)
R.W. Komm, R.F. Howard, J.W. Harvey, Sol. Phys. **147**, 207 (1993)
A.G. Kosovichev, J. Schou, Astrophys. J. Lett. **482**, L207 (1997)
R.H. Kraichnan, J. Fluid Mech. **8**, 1385 (1965)
V. Krishan, U. Paniveni, J. Singh, R. Srikanth, Mon. Not. R. Astron. Soc. **334**, 230 (2002)
N.A. Krivova, L. Balmaceda, S.K. Solanki, Astron. Astrophys. **467**, 335 (2007)
K. Langhans, G.B. Scharmer, D. Kiselman, M.G. Löfdahl, Astron. Astrophys. **464**, 763 (2007)
J.K. Lawrence, Sol. Phys. **110**, 73 (1987)
J.K. Lawrence, A.C. Cadavid, A.A. Ruzmaikin, Astrophys. J. **513**, 506 (1999)
B.W. Lites, M. Kubo, H. Socas-Navarro, T. Berger, Z. Frank, R. Shine, T. Tarbell, A. Title, K. Ichimoto, Y. Katsukawa, S. Tsuneta, Y. Suematsu, T. Shimizu, S. Nagata, Astrophys. J. **672**, 1237 (2008)
G.W. Lockwood, B.A. Skiff, G.W. Henry, S. Henry, R.R. Radick, S.L. Baliunas, R.A. Donahue, W. Soon, Astrophys. J. Suppl. Ser. **171**, 260 (2007)
H. Lustig, Astron. Astrophys. **125**, 355 (1983)
N. Meunier, Astrophys. J. **515**, 801 (1999a)
N. Meunier, Astrophys. J. **527**, 967 (1999b)
N. Meunier, Astron. Astrophys. **405**, 1107 (2003)

N. Meunier, Astron. Astrophys. **436**, 1075 (2005a)
N. Meunier, Astron. Astrophys. **437**, 303 (2005b)
N. Meunier, Astron. Astrophys. **442**, 693 (2005c)
N. Meunier, R. Tkaczuk, T. Roudier, M. Rieutord, Astron. Astrophys. **461**, 1141 (2007a)
N. Meunier, R. Tkaczuk, T. Roudier, Astron. Astrophys. **463**, 745 (2007b)
N. Meunier, T. Roudier, R. Tkaczuk, Astron. Astrophys. **466**, 1123 (2007c)
N. Meunier, T. Roudier, Astron. Astrophys. **466**, 691 (2007d)
N. Meunier, T. Roudier, M. Rieutord, Astron. Astrophys. **488**, 1109 (2008)
M.S. Miesch, Living Rev. Sol. Phys. **2**, 1 (2005)
K. Mikurda, O. von der Lühe, Sol. Phys. **235**, 31 (2006)
M. Minarovjech, V. Rušin, M. Saniga, Sol. Phys. **241**, 269 (2007)
U. Mitra-Kraev, M.J. Thompson, Astronomische Nachrichten **328**, 1009 (2007)
R. Muller, A. Hanslmeier, M. Saldaña-Muñoz, Astron. Astrophys. **475**, 717 (2007)
E. Nesme-Ribes, E. Ferreira, P. Mein, Astron. Astrophys. **274**, 563 (1993)
L.J. November, J. Toomre, K.B. Gebbie, G.W. Simon, Astrophys. J. **245**, 123 (1981)
D. Orozco Suárez, L.R. Bellot Rubio, J.C. del Toro Iniesta, S. Tsuneta, B.W. Lites, K. Ichimoto, Y. Katsukawa, S. Nagata, T. Shimizu, R.A. Shine, Y. Suematsu, T.D. Tarbell, A.M. Title, Astrophys. J. **670**, 61 (2007)
M. Ossendrijver, Astron. Astrophys. Rev. **11**, 287 (2003)
U. Paniveni, V. Krishan, J. Singh, R. Srikanth, Sol. Phys. **231**, 1 (2005)
R.G. Paxman, J.H. Seldin, M.G. Löfdahl, G.B. Scharmer, C.U. Keller, Astrophys. J. **466**, 1087 (1996)
M.J. Penn, W. Livingston, Astrophys. J. **649**, 45 (2006)
M.J. Penn, R.K.D. MacDonald, Astrophys. J. **662**, L123 (2007)
D.G. Preminger, S.R. Walton, Sol. Phys. **235**, 387 (2006a)
D.G. Preminger, S.R. Walton, Geophys. Res. Lett. **33**, L23108 (2006b)
K.G. Puschmann, M. Sailer, Astron. Astrophys. **454**, 1011 (2006)
K.P. Raju, J. Singh, Sol. Phys. **207**, 11 (2002)
M.P. Rast, ESA SP-517: GONG + 2002. Local and Global Helioseismology: the Present and Future, ed. by H. Sawaya-Lacoste (2003), p. 163
M.P. Rast, J.P. Lisle, J. Toomre, Astrophys. J. **608**, 1156 (2004)
M. Rempel, Astrophys. J. **647**, 662 (2006)
M. Rempel, Astrophys. J. **655**, 651 (2007)
M. Rempel, S. Brun, Space Sci. Rev. (2008, this issue)
J.-C. Ribes, E. Nesme-Ribes, Astron. Astrophys. **276**, 549 (1993)
M. Rieutord, T. Roudier, J.-M. Malherbe, F. Rincon, Astron. Astrophys. **357**, 1063 (2000)
M. Rieutord, N. Meunier, T. Roudier, S. Rondi, F. Beigbeder, L. Parès, Astron. Astrophys. **479**, L17 (2008)
F. Rincon, F. Lignières, M. Rieutord, Astron. Astrophys. **430**, 57 (2005)
T. Rimmele, SPIE **5490**, 34 (2004)
T. Rimmele, Astrophys. J. **672**, 684 (2008)
T. Roudier, R. Muller, Sol. Phys. **107**, 11 (1987)
L.H.M. Rouppe van der Vort, Astron. Astrophys. **389**, 1020 (2002)
D. Ruždjak, V. Ruždjak, R. Brajša, H. Wöhl, Sol. Phys. **221**, 225 (2004)
G.B. Scharmer, M. Löfdahl, Advances in Space Research **11**, 129 (1991)
G. Scharmer, Space Sci. Rev. (2008, this issue)
A.A. Schekochihin, S.C. Cowley, S.F. Taylor, J.L. Maron, J.C. McWilliams, Astrophys. J. **612**, 276 (2004)
P.H. Scherrer, R.S. Bogart, R.I. Bush, J.T. Hoeksema, A.G. Kosovichev, J. Schou, W. Rosenberg, L. Springer, T.D. Tarbell, A. Title, C.J. Wolfson, I. Zayer, MDI Engineering Team, Sol. Phys. **162**, 129 (1995)
R. Schlichenmaier, Space Sci. Rev. (2008, this issue)
J. Schou, Astrophys. J. Lett. **596**, L259 (2003)
C.J. Schrijver, Astrophys. J. **547**, 475 (2001)
M. Schüssler, Astron. Astrophys. **94**, L17 (1981)
M. Schüssler, Astron. Nachr. **326**, 194 (2005)
P.E. Seiden, D.G. Wentwel, Astrophys. J. **460**, 522 (1996)
G.W. Simon, N.O. Weiss, Zeit. Astrophys. **69**, 435 (1968)
H.B. Snodgrass, Astrophys. J. Lett. **316**, L91 (1987)
H.B. Snodgrass, R.K. Ulrich, Astrophys. J. **351**, 309 (1990)
D. Sokoloff, E. Nesme-Ribes, Astron. Astrophys. **2888**, 293 (1994)
S.K. Solanki, M. Fligge, Geophys. Res. Lett. **26**, 2465 (1999)
S.K. Solanki, N.A. Krivova, J. Geophys. Res. **108**, 1200 (2003)
S.K. Solanki, N.A. Krivova, Sol. Phys. **224**, 197 (2004)
H.C. Spruit, Sol. Phys. **213**, 1 (2003)

R.F. Stein, Å. Nordlund, Astrophys. J. **499**, 914 (1998)
M. Švanda, M. Klvaňa, M. Sobotka, V. Bumba, Astron. Astrophys. **477**, 285 (2008)
M. Švanda, J. Zhao, A.G. Kosovichev, Sol. Phys. **241**, 27 (2007)
J. Sýkora, Sol. Phys. **13**, 292 (1970)
K.F. Tapping, D. Boteler, P. Charbonneau, A. Crouch, A. Manson, H. Paquette, Sol. Phys. **246**, 309 (2007)
T.D. Tarbell, M. Bruner, B. Jurcevich, J. Lemen, K. Strong, A. Title, J. Wolfson, Solar dynamic phenomena and solar wind consequences, in *Proceedings of the Third SOHO Workshop*, Estes Park, Colorado, ed. by J.J. Hunt. ESA SP-373 (1994), p. 375
M.J. Thompson, N. Weiss, Space Sci. Rev. (2008, this issue)
M.J. Thompson, J. Toomre, E. Anderson, H.M. Antia, G. Berthomieu, D. Burtonclay, S.M. Chitre, J. Christensen-Dalsgaard, T. Corbard, M. Derosa et al., Science **272**, 1300 (1996)
A.M. Title, T.D. Tarbell, K.P. Topka, S.H. Ferguson, R.A. Shine, SOUP Team, Astrophys. J. **336**, 475 (1989)
A.M. Title, K.P. Topka, W. Schmidt, C. Balke, G. Scharmer, Astrophys. J. **393**, 782 (1992)
L. van Driel-Gesztelyi, L. Culhane, Space Sci. Rev. (2008, this issue)
M. van Noort, L. Rouppe van der Voort, M.G. Löfdahl, Sol. Phys. **228**, 191 (2005)
S.V. Vorontsov, J. Christensen-Dalsaard, J. Schou, V.N. Strakhov, M.J. Thompson, Science **296**, 101 (2002)
H. Wang, Sol. Phys. **117**, 343 (1988)
Y.-M. Wang, N.R. Jr. Sheeley, A.G. Nash, Astrophys. J. **383**, 431 (1991)
H. Wang, F. Tang, H. Zirin, J. Wang, Sol. Phys. **165**, 223 (1996)
Y.-M. Wang, J.L. Lean, N.R. Jr. Sheeley, Astrophys. J. **625**, 522 (2005)
S. Wedemeyer-Böhm, Å. Nordlund, A. Lagg, Space Sci. Rev. (2008, this issue)
D.G. Wentzel, P.E. Seiden, Astrophys. J. **390**, 280 (1992)
T. Wenzler, S.K. Solanki, N.A. Krivova, Astron. Astrophys. **432**, 1057 (2005)
A.G. de Wijn, S. Solanki, J. Stenflo, S. Tsuneta, Space Sci. Rev. (2008, this issue)
G. Withbroe, Sol. Phys. **265**, 369 (2006)
F. Wöger, O. von der Lühe, K. Reardon, Astron. Astrophys. **488**, 375 (2008)
T.N. Woods, Adv. Space Res. **42**, 895 (2008)
H. Yoshimura, Astrophys. J. **247**, 1102 (1981)
R.A. Zappala, F. Zuccarello, Astron. Astrophys. **242**, 480 (1991)
J. Zhao, A.G. Kosovichev, Astrophys. J. **603**, 776 (2004)
J. Zhao, A.G. Kosovichev, T.L. Jr Duvall, Astrophys. J. Lett. **607**, L135 (2004)
J. Zhao, A.G. Kosovichev, T.L. Jr. Duvall, Astrophys. J. **557**, 384 (2001)
F. Zuccarello, Astron. Astrophys. **272**, 587 (1993)

Large Scale Flows in the Solar Convection Zone

Allan Sacha Brun · Matthias Rempel

Originally published in the journal Space Science Reviews, Volume 144, Nos 1–4, 151–173.
DOI: 10.1007/s11214-008-9454-9

Abstract We discuss the current theoretical understanding of the large scale flows observed in the solar convection zone, namely the differential rotation and meridional circulation. Based on multi-D numerical simulations we describe which physical processes are at the origin of these large scale flows, how they are maintained and what sets their unique profiles. We also discuss how dynamo generated magnetic field may influence such a delicate dynamical balance and lead to a temporal modulation of the amplitude and profiles of the solar large scale flows.

Keywords Sun: convection, rotation, mean flows, magnetism, torsional oscillations

1 Introduction

Understanding the origin of large scale flows (differential rotation and meridional flow) is crucial for a comprehensive understanding of the solar dynamo. While differential rotation is the dominant process for producing toroidal field and therefore the primary energy supplier in an $\alpha\Omega$-dynamo, the meridional flow is a transport process for poloidal and toroidal fields. The latter is of particular interest for flux-transport dynamo models, in which the butterfly diagram is the result of an equatorward directed flow at the base of the convection zone advecting magnetic field. While the differential rotation profile is known in great detail from helioseismology (Thompson et al. 2003, see also Meunier and Zhao, these proceedings), the meridional flow is only directly accessible from observations in upper most layers of the solar convection zone; the structure of this flow in the deeper layers of the convection zone is currently only constrained through extrapolations based on continuity and models of large

A.S. Brun (✉)
DSM/IRFU/SAp, CEA-Saclay & UMR AIM, CEA-CNRS-Université Paris 7, 91191 Gif-sur-Yvette, France
e-mail: sacha.brun@cea.fr

M. Rempel
HAO/NCAR, P.O. Box 3000, Boulder, CO 80307, USA
e-mail: rempel@hao.ucar.edu

M.J. Thompson et al. (eds.), *The Origin and Dynamics of Solar Magnetism*.
DOI: 10.1007/978-1-4419-0239-9_10

scale flows. The magnetic field produced by the dynamo is acting back on large scale flows via Maxwell stresses, leading to changes of the mean as well as cyclic variations, which are known in the case of the differential rotation as torsional oscillations.

Due to the vast range of length and time scales encountered in the highly stratified convection zone and tachocline beneath, a direct numerical simulation of the solar convection zone poses a significant numerical challenge. As a result large scale flows in the convection zone of the Sun and solar like stars have been modelled using two different but complementary approaches: mean field models that focus on the large scale and rely on parametrisation of turbulent transport processes and 3-D MHD simulations that resolve self consistently the turbulent transport processes, but are restricted in the degree of turbulence they can reach. Despite these differences both approaches have equally contributed to our current understanding of large scale flows in the solar convection zone. In this article we want to point out the common denominator between mean field approaches and 3-D simulations, discuss their restrictions and point out aspects in which they differ.

2 Modelling Approach

2.1 2-D Axisymmetric Mean Field Models

The mean field approach is based on a decomposition of properties into their large scale mean values (e.g. differential rotation, meridional flow, large scale magnetic field) and small scale fluctuating parts, typically associated with unresolved turbulence. Non-linear terms in the momentum, energy and induction equations lead to non-vanishing second order correlation terms of small scale quantities that act as drivers for large scale flows or as turbulent induction effects for the large scale magnetic field. The decomposition into large and small scale properties and the arising correlation terms driving large scale flows are the strength and the weakness of this approach at the same time. On the one hand the computational expense is decreased by orders of magnitude allowing for simulations covering long time scales as well as exploring wide parameter ranges, on the other hand the results are heavily dependent on parametrisation of the second order correlation terms. A key ingredient for mean field differential rotation models is the parametrisation of the non-diffusive turbulent angular momentum transport. Expressions for these transport terms were derived by Durney and Spruit (1979), Hathaway (1984) and more recently by Kitchatinov and Rüdiger (1993) using a quasi-linear approach (see also Kitchatinov and Rüdiger 2005 for an improved representation). Comparisons between mean field parametrisation and local numerical simulations (f-plane approximation) lead in general to a qualitative agreement (Rüdiger et al. 2005b); however, larger differences are found when compared to results from global simulations of the entire convection zone. The latter results primarily from the fact that a significant fraction of energy and angular momentum is transported by large scale coherent structures ‘banana cells’ (see also Sect. 3) that are very difficult to capture in mean field approaches as well as local 3-D simulations. For a comprehensive description of mean field theory we refer to Rüdiger and Hollerbach (2004).

2.2 3-D Global MHD Simulations

Contrary to the mean field approach described briefly above, three dimensional numerical simulations of the solar convection and its associated mean flows do not assume scale separation and the MHD equations, i.e., Navier-Stokes, energy, mass conservation and the magnetic induction equations are solved self-consistently, non linearly and coupled. Solar

convection can be studied either in Cartesian or in spherical geometry. Both approaches have their advantages (high turbulence level or global geometry) and disadvantages (local geometry or lower degree of turbulence). Understanding the origin of large scale flows in the solar convection, implies that a global approach must be employed. We will thus describe in the sections below scientific results obtained with a 3-D MHD spherical code, i.e. the Anelastic Spherical Harmonic (ASH) code. ASH solves the three-dimensional anelastic MHD equations of motion in a rotating frame using a pseudo spectral approach (Clune et al. 1999; Miesch et al. 2000; Brun et al. 2004). The anelastic approximation is assumed in order to retain compressibility effects (important in the solar convective envelope) without having to follow the sound waves generated by convective motions. The resulting equations are fully nonlinear in velocity variables and linearised in thermodynamic variables with respect to a spherically symmetric mean state. However given the very high degree of turbulence characterising the solar outer layers, with a Reynolds number $Re = VL/\nu \sim 10^{12}$ or greater (with V and L being representative velocity and length of the system and ν the kinematic viscosity), it is currently not possible to model in 3-D the whole range of dynamical scales present in stars even with the most powerful parallel supercomputers and some simplifying assumptions must be considered. The ASH code thus relies for describing the large scale nonlinear dynamics occurring in the Sun on the so-called large eddy simulation/sub grid scale (LES-SGS) approach (Lesieur 1997). As a result ASH uses effective diffusivities for momentum, heat and magnetic field rather than their microscopic values and models mainly large scale convection (down to supergranules for today's most resolved runs) and mean flows. Of course more refined the model will be, better the nonlinear effects and the turbulence will be described resulting in more realistic simulations. As of today ASH has been able to attain the equivalent of a resolution of $\sim 1500^3$ grid points (Miesch et al. 2008) and one can expect that in the near future global simulations on petaflop supercomputers will reach a resolution of $10\,000^3$, which will enable the modelling of granular size convection patterns ($\sim$1000 km) on the sphere thus getting closer and closer to reality.

3 Solar Convection

Convection plays a central role for the dynamical behaviour and the evolution of our star. Simply put convection is the process transporting energy via bulk motions: a parcel of fluid (eddy) that is locally hot rises toward the surface because it is lighter than the surrounding media,and it cools by releasing its heat content through radiation and sinks because it became locally heavier. By reaching the bottom it is heated again, rises, and so on and so forth. This continuous process tends to reduce the temperature gradient (homogenise the temperature throughout the layer). In the case of our star it transports outward the heat generated deep inside its nuclear core when radiation due to a too large opacity fails at doing so (this occurs around 71% of the solar radius). In the case of heavily stratified convective layer such as in the solar envelope the entropy gradient is the key quantity used to characterise convective motions. When convection is very efficient, the stratification is very close to adiabatic. This is indeed the case deep inside the Sun but clearly not so near its surface. Convective motions are triggered near the system boundaries in so called thermal (cold or warm) boundary layers from which plumes or intense vorticity tubes detach and reach the opposite boundary (Castaing et al. 1989). In the Sun the large stratification leads to an asymmetry between fast cold concentrated plumes and slow warm broad upflows (see below and Fig. 1). A key parameter to characterise convection is the Rayleigh number *Ra*, e.g. the ratio between the physical process at the origin of the bulk motion (here the Archimedes' force)

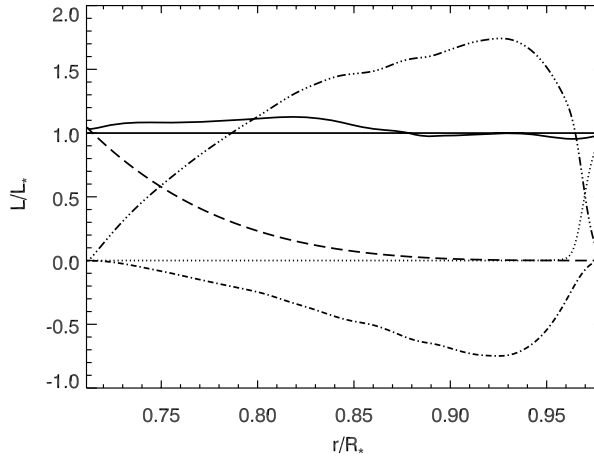


Fig. 1 Energy flux balance realised in a simulation of solar turbulent convection (Miesch et al. 2008). Shown are respectively the enthalpy flux (*dash-three dots*), radiation flux (*dash*), kinetic energy flux (*dash dot*), sub grid scale flux (*dot*), and total flux (*solid*) converted to luminosity and normalised to the solar luminosity

and the processes opposing these motions, i.e. the heat diffusivity and viscous stresses. If *Ra* is less than a given critical value (typical values in laboratory experiments are around 1000), then the amount of energy transmitted to the system is not large enough to trigger the convection instability and the fluid remains at rest. On the contrary above this threshold, convection sets in. Thus the a priori knowledge of the Rayleigh number of a system is sufficient to characterise its convection state (i.e. stable or unstable). In the Sun this number is huge $Ra \sim 10^{12}$, about 9 orders of magnitude above the threshold found in laboratory experiments and confirmed by linear analysis (Chandrasekhar 1961) and it is thus clearly expected that the Sun is in a state of turbulent convection resulting in peculiar heat transport. This highly turbulent state is difficult to study analytically or through linear perturbation analysis. Further the Sun rotates and it is magnetised and this further complicates the study of the solar convection zone. One must thus rely on nonlinear multi-D numerical simulations in order to gain new insights on such a complex magnetohydrodynamical system.

To illustrate the recent progresses made with 3-D global solar models, we display in Fig. 1 the energy fluxes and in Fig. 2 the convection patterns realised in a simulation of turbulent convection in a spherical shell representative of the bulk of the solar convection zone (see Brun and Toomre 2002; Miesch et al. 2000, 2008 for more details). Figure 1 shows the energy fluxes converted to luminosity and normalised by the solar luminosity as a function of radius. One can clearly see that the enthalpy or convective flux (i.e. correlation $\langle v_r' T' \rangle$) is dominant over most of the domain except at the boundaries where radiative or diffusive processes take over. It reaches up to 170% of the solar luminosity, and contrary to classical mixing length theory assuming that the convective luminosity equals that of the star (Hansen and Kawaler 1995), we clearly see that in our 3-D self-consistent simulations this is not the case. The origin of this 'extra luminous' convective flux comes from the fact that it must compensate the strong and negative (inward) kinetic energy flux. This is typical of highly turbulent and stratified convection and it is linked to the asymmetry between the intense downward plumes and the broad slow upflows mentioned earlier (see also Cattaneo et al. 1991).

Such asymmetry is clearly visible in Fig. 2 where we display the radial velocity and temperature fluctuations near the top of the domain. The narrow downflow lanes correlated well with the network of cool material while the broad upflows are relatively warm. At the interstices of the downflow network, intense cold spinners are seen. This is a direct consequence

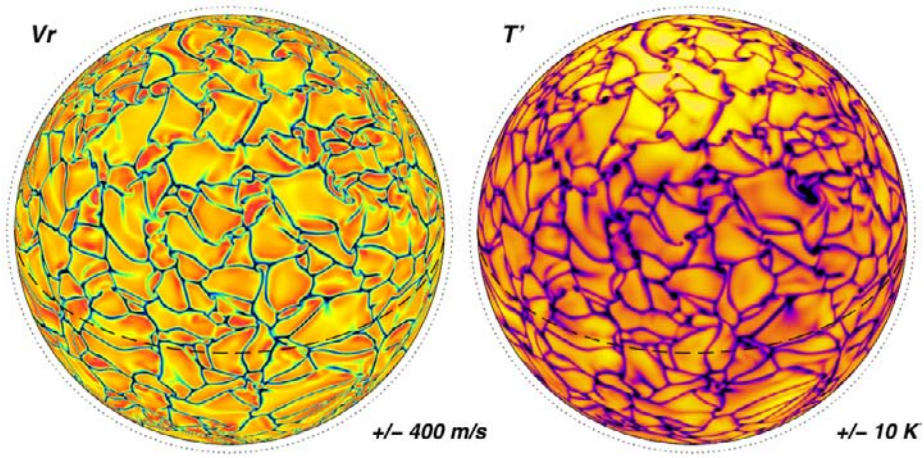


Fig. 2 Convection patterns near the surface realised in the same simulation as featured in Fig. 1. Radial velocities with downflows in dark colours are shown in the *left panel*. Temperature fluctuations with negative fluctuations in dark colours are shown in the *right panel*

of the Coriolis force acting on the convective motions as they converge toward the downflow lanes. As the parcel moves from high pressure points to low pressure valley, the local force balance between the pressure gradient and the Coriolis force leads to the generation of 'cyclones' as in the Earth's atmosphere, i.e. the vortices rotate counter-clockwise in the northern hemisphere. The convective patterns are more isotropic at high latitude than near the equator. At low latitude the convective cells tend to align with the rotation axis. The convection varies significantly over time with convection cells continuously merging, splitting, emerging and being advected by the background large scale flow (Brun and Toomre 2002; Miesch et al. 2008). Looking more closely at the temperature field we notice a banded structure with hot poles, cold mid latitudes and a warm equator. Accompanying these latitudinal temperature variations are the corresponding pressure and entropy gradients. These variations are due to the latitudinal heat transport that naturally arise in convective system under the influence of rotation. The latitudinal heat transport is mostly dominated by the poleward latitudinal enthalpy flux (i.e. correlations $\langle v'_\theta T' \rangle$) that is compensated by the equatorward latitudinal entropy flux. The mean latitudinal heat flux associated with the meridional circulation ($\langle v_\theta \rangle \langle T \rangle$, with the last bracket being the axisymmetric average of the three dimensional temperature fluctuations) and the latitudinal kinetic energy flux both mean $\langle v_\theta \rangle \langle v^2 \rangle$ and fluctuating $\langle v'_\theta v'^2 \rangle$ (where in all the above expressions we omit dimensional quantities for clarity) play a negligible role in the overall balance found in the models (Elliott et al. 2000; Brun and Toomre 2002). These variations in latitude of the entropy and temperature correspond to a self-established thermal wind associated as we will see below with the longitudinal large scale flow present in the convection zone, i.e. the differential rotation.

A detailed analysis of the energy budget integrated over the whole volume of the simulation, reveals that the kinetic energy is smaller by at least 6 orders of magnitude compared to the internal or potential energies. The kinetic energy itself can be split into its mean and fluctuating components. We find that most of the kinetic energy is in the differential rotation and in the non-axisymmetric convective motions with a very small part left for the meridional flow.

4 Maintenance of Large Scale Flows

In order to understand the maintenance of large scale flows it is beneficial to start with the mean field decomposition briefly outlined above. Splitting the velocity field as $\boldsymbol{v} = \langle \boldsymbol{v} \rangle + \boldsymbol{v}'$ allows to derive from the momentum equation a set of equations for the mean flows $\langle \boldsymbol{v} \rangle$. In spherical geometry it is most convenient to decompose the equations into an angular momentum transport equation (ϕ-component of momentum transport) and the longitudinal component of the vorticity equations that combines both meridional components of the momentum equations. This decomposition leads to the system ($\langle v_\phi \rangle = \Omega r \sin\theta$):

$$\varrho (r \sin\theta)^2 \frac{\partial \Omega}{\partial t} = -\operatorname{div} \boldsymbol{F} \tag{1}$$

$$\frac{\partial \langle \omega_\phi \rangle}{\partial t} + [\ldots] = r \sin\theta \frac{\partial \Omega^2}{\partial z} - \frac{g}{c_p r} \frac{\partial s}{\partial \theta} \tag{2}$$

In (2) the coordinate θ is the colatitude, r the radius and z is oriented along the direction of the axis of rotation. s is the specific entropy, c_p the heat capacity at constant pressure, g the gravitational acceleration and $[\ldots]$ denotes advection terms, meridional components of the Reynolds stress and magnetic terms that were omitted to enhance the clarity of the presentation. The most important terms in (2) are the two terms on the right hand side describing a thermal wind balance of differential rotation. The first term arises from the Coriolis force, the second term from pressure and buoyancy forces. In the absence of any entropy variation in latitude Ω has to be close to constant on cylinders (Taylor Proudman state). In the angular momentum equation (1) $\boldsymbol{F}$ denotes the angular momentum flux with the components:

$$F_i = r \sin\theta \varrho \left\{ \underbrace{\langle v_i' v_\phi' \rangle}_{\text{Reynolds stress}} + \underbrace{\langle v_i \rangle \Omega r \sin\theta}_{\text{Meridional flow}} - \frac{1}{4\pi\varrho} \underbrace{(\langle B_i' B_\phi' \rangle + \langle B_i \rangle \langle B_\phi \rangle)}_{\text{Maxwell stress}} \right\} \tag{3}$$

Angular momentum is transported by correlations of small scale turbulent motions (Reynolds stresses), large scale meridional flow ($\langle v_r \rangle$, $\langle v_\theta \rangle$), viscous effects (omitted since in the real Sun they are small) as well as Maxwell stresses (Brun et al. 2004). Here we also decomposed the magnetic field into mean and fluctuating parts $\boldsymbol{B} = \langle \boldsymbol{B} \rangle + \boldsymbol{B}'$ and separated the Maxwell stress into the contributions from the mean field and fluctuating field. The magnetic pressure does not enter here, since we consider the longitudinal average to compute the mean, which eliminates terms arising from gradients in the ϕ direction. We discuss first the purely hydrodynamic situation and come back to the role of the magnetic field in the end of this section.

A stationary solution for Ω requires that the divergence of the total angular momentum flux vanishes. Since in general the Reynolds stress by itself is not divergence free, a meridional flow is required as an additional degree of freedom in the system. The primary driver of the meridional flow is a small difference in the two terms on the right hand side of (2), a deviation from the thermal wind balance of the differential rotation. In a stationary state terms on the right hand side of (2) have to be in balance with the (not explicitly shown) advection and Reynolds stress terms. These terms can be estimated as $(v/d)^2$, with a typical velocity v and length scale d, while terms on the right hand side are of order Ω^2. Deviations from a thermal wind balance have to be of the order $(v/\Omega d)^2 \ll 1$ (using $d = 200$ Mm and $v \sim 100$ m/s gives a value ~ 0.03).

An important question is what determines the structure of the meridional flow. Since the meridional flow has to be such that the divergence of the total angular momentum flux

vanishes it is very closely tied to the structure of the turbulent Reynolds stress. This has two eminent consequences: 1. The meridional flow exhibits the same amount of turbulent fluctuations the Reynolds stress does; 2. The long term averaged flow shows patterns on a length scale similar to a typical scale of the long term averaged Reynolds stress, i.e. a single flow cell per hemisphere can only be expected if the long term averaged Reynolds stress shows a dominant hemispheric pattern.

Which property of the Reynolds stress determines the flow direction of the meridional flow? The angular momentum transport by the meridional flow $\boldsymbol{v_m}$ can be written as $\boldsymbol{F_m} = \varrho \boldsymbol{v_m} l$, where l denotes the specific angular momentum $l = \Omega (r \sin\theta)^2$. Since $\mathrm{div} \varrho \boldsymbol{v_m} = 0$ there is no mass flux across an almost cylindrical surface defined by $l = \mathrm{const}$. Since by definition l is constant on this surface, there is also no angular momentum flux through this surface. If we now apply this to a volume between two surfaces defined by $l_1 = \mathrm{const}$ and $l_2 = \mathrm{const}$ and assume the limit $l_1 \to l_2$, we see that the net effect of meridional flow is a redistribution of angular momentum parallel to iso-l surfaces. It therefore couples most strongly with the component of the Reynolds stress that transports angular momentum parallel to iso-l surfaces (approximately in the direction of the axis of rotation), while the component of the Reynolds stress perpendicular to the iso-l surfaces is the primary driver for differential rotation. A meridional flow poleward at the surface and equatorward at the base of the convection zone can be expected if the Reynolds stress has a strong component that transports angular momentum inward along the axis of rotation.

Since differential rotation tends to shear and amplify magnetic field and therefore converts kinetic energy into magnetic energy it is expected that the effect of the Maxwell stress is a reduction of differential rotation. This overall behaviour has been seen in early 3-D dynamo simulations of Gilman (1983) and also in the more recent work by Brun et al. (2004), see also Sect. 7. While in the latter the microscopic Maxwell stress $\langle B_i' B_\phi' \rangle$ dominates the feedback, mean field dynamo models typically consider only the macroscopic Maxwell stress $\langle B_i \rangle \langle B_\phi \rangle$. In both cases the result is a reduction of the mean differential rotation; periodic mean field dynamo solutions impose additionally a cyclic component of the zonal flows (torsional oscillations). Formally Maxwell stresses also enter in (2), leading to the possibility of magnetically induced deviations from the Taylor Proudman state. However, an estimate by Rempel (2007a) showed that the meridional component of the Maxwell stress is not sufficient in the bulk of the convection zone to cause significant deviations from the Taylor Proudman state unless the poloidal field is stronger than 10 kG.

We have seen in (2) that the profile of differential rotation cannot be explained by considering the momentum equation alone. Also thermal effects are crucial as they can change the latitudinal distribution of entropy in the convection zone. The latitudinal entropy profile in the convection zone can be influenced by anisotropic convective heat transport as well as coupling to the tachocline, which we will discuss in more detail in Sect. 5.1. Heat transport results formally from the second order correlation terms $\langle v_r' T' \rangle$ and $\langle v_\theta' T' \rangle$. While in non-rotating convection only the former exists, modifications introduced by rotation can also lead to latitudinal heat transport (Kitchatinov et al. 1994; Rüdiger et al. 2005a; Elliott et al. 2000), which is typically poleward directed (deflection of heat transport toward axis of rotation, see Sect. 3). The role of entropy perturbations in differential rotation is not restricted to just shaping the profile of differential rotation, thermal perturbations can have a significant influence on the amplitude of DR and can even be the origin of differential rotation in some extreme cases.

5 Differential Rotation

5.1 Results from Mean Field Models

Modelling of differential rotation using the mean field approach heavily depends on the parametrisation of the terms $\langle v_r' v_\phi' \rangle$ and $\langle v_\theta' v_\phi' \rangle$, which are typically decomposed in a diffusive component and a non-diffusive component called Λ-effect after the work of Kitchatinov and Rüdiger (1993). Applying these parametrisation to model rotation in stellar convection zones it was quickly realised that the main difficulty in explaining solar-like differential rotation lies in how avoid the Taylor-Proudman state with cylindrical differential rotation. While first models not considering the meridional flow were able to reproduce solar-like differential rotation, more sophisticated models taking into account the meridional flow obtained only solutions close to the Taylor-Proudman state unless a very large (unreasonable) value for the eddy viscosity were used. The problem was referred to also as the 'Taylor-number puzzle' after Brandenburg et al. (1990). Kitchatinov and Rüdiger (1995) showed that an alternative solution of this problem can be given if the anisotropic convective energy transport is considered. The baroclinic term on the right hand side of (2) can allow for stationary solutions with differential rotation deviating from the Taylor-Proudman state. A equator-pole temperature difference of about 10 K (hotter pole) is sufficient to obtain solar-like differential rotation with lines of constant angular velocity inclined to the axis of rotation by about 25°. All of the recent mean field models by Küker and Stix (2001), Küker and Rüdiger (2005a, 2005b, 2007) are successful in avoiding the Taylor Proudman state by considering anisotropic convective energy transport, however, in some cases the baroclinic term can be also too strong and lead to 'disk-like' differential rotation profiles.

Anisotropic convective energy transport is automatically considered in global 3-D simulations, but in most cases it turns out to be not strong enough for obtaining solar-like differential rotation as we explain in Sect. 5.2. However, deviations from the Taylor-Proudman state seen in these simulations are typically in a baroclinic balance.

Recently Rempel (2005) showed that coupling between the tachocline and convection zone can also provide the latitudinal entropy variation needed to explain the observed profile of solar differential rotation. A typical solution from that model is shown in Fig. 3, displaying differential rotation (a), corresponding entropy perturbation (b) and the stream function of the meridional flow (c). Panel (d) shows for comparison the profile of differential rotation obtained if the effects of the entropy perturbation displayed in (b) are neglected. In this model uniform rotation is imposed at the lower boundary, leading to the formation of a tachocline like shear layer, which is located in a subadiabatic (stable) stratification. According to (2) the strong deviations from the Taylor-Proudman state in this region drive a meridional flow which leads in a subadiabatic stratification to the establishment of an entropy perturbation balancing the term $\sim \partial \Omega / \partial z$ (note: this adjustment toward baroclinic balance only works in a subadiabatic region). The result is a tachocline shear layer with a strong latitudinal variation of entropy as required to keep the terms on the right hand side of (2) in balance. As a result the tachocline imposes a latitudinal entropy variation as boundary condition at the base of the convection zone that is transported into the convection zone by convection and allows for a significant deviation from the Taylor-Proudman state. While the process of establishing a thermal wind balance in the tachocline as well as the spreading of thermal perturbations through the convection zone is found to be very robust (see also the results from 3-D global simulations discussed in Sect. 5.2), the unknown strength of coupling between tachocline and convection zone is the biggest uncertainty in this picture, since it cannot be easily determined from first principles and has a strong influence on the amplitude of the entropy variation in the convection zone.

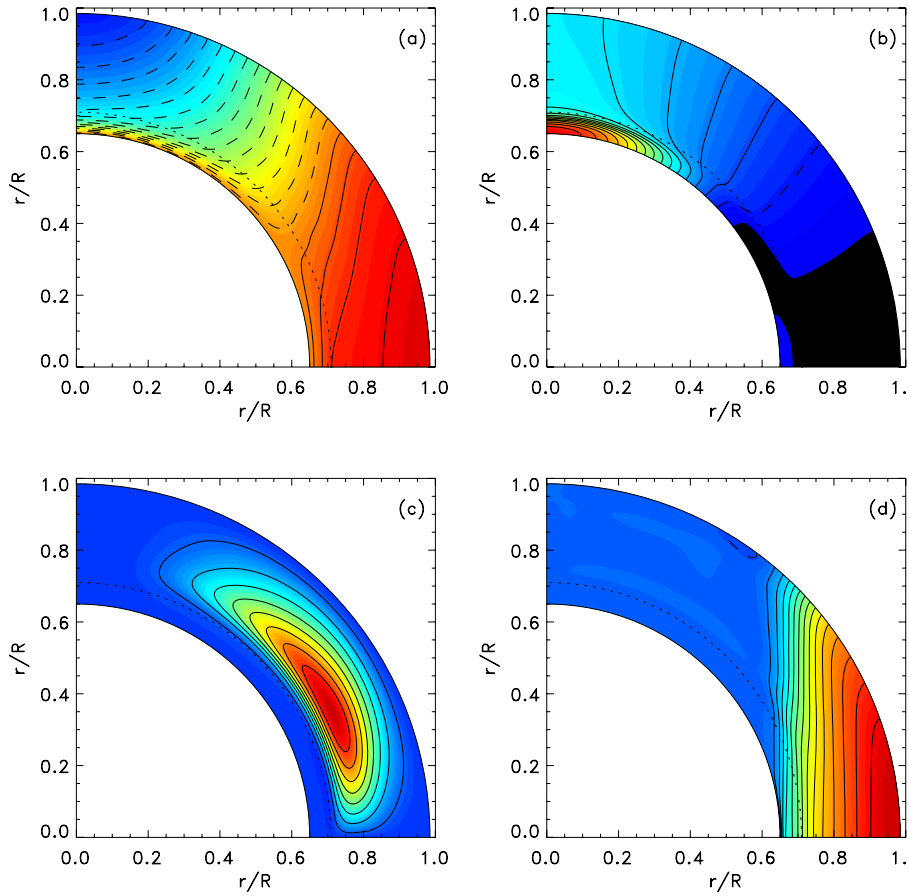


Fig. 3 Contour plots of differential rotation (**a**), entropy perturbation (**b**) and stream function of meridional flow (**c**) using the mean field model of Rempel (2005). Panel (**d**) shows the differential rotation profile obtained using the same parametrisation of the Reynolds stress but neglecting the effects of baroclinicity

While there is a general agreement that thermal effects are essential for solar-like differential rotation, it is still unclear whether the required latitudinal entropy variation is a consequence of anisotropic convective energy transport, imposed by the tachocline, or a combination of both.

5.2 Angular Velocity and Momentum Redistribution in 3-D Global Models

As explained in Sect. 3, global models of a rotating convection zone naturally develop large scale flows. These simulations are thus very useful to study and understand in detail the maintenance of these flows without the need as in mean field model to parametrise the non-linear transport processes of angular momentum and heat. In Fig. 4 we display the angular velocity established in such 3-D models of the solar convection zone. The angular velocity profiles realised in these simulations, in particular in cases AB and AB3 displayed in Fig. 4 (see Brun and Toomre 2002; Miesch et al. 2006), are in good agreement with helioseismic inversions (Thompson et al. 2003): They possess an equator to pole contrast $\Delta\Omega/\Omega_0$ of

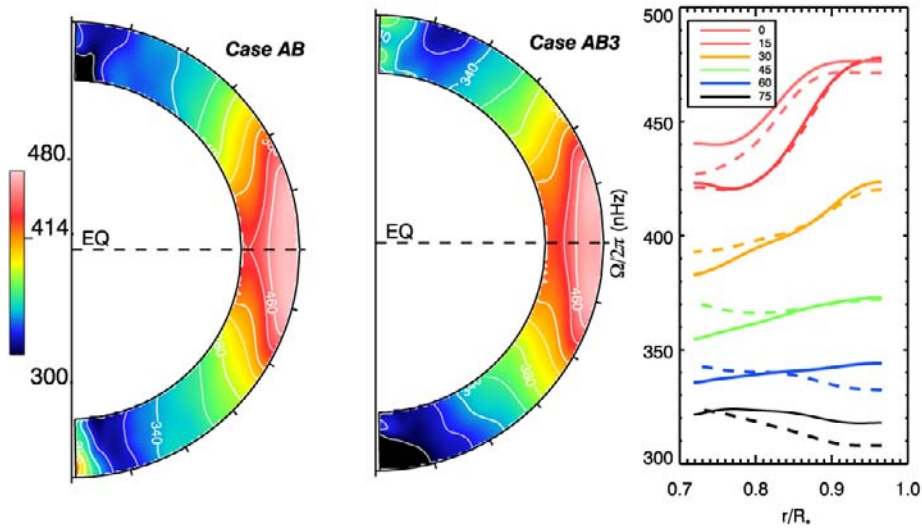


Fig. 4 Contour plots of differential rotation for global 3-D models without or with entropy forcing (Miesch et al. 2006). Also shown are radial cuts at indicated latitudes of Ω (using *solid lines* for case AB and *dashed lines* for case AB3 with thermal forcing). Note the tilt toward a conical differential rotation between the two cases

$\sim$30%, some constancy along radial lines at mid latitude, a monotonic decrease of the differential rotation toward the polar regions and slow poles. However it is clear that the differential rotation profiles in case AB3 is more conical than in case AB and thus more solar-like. As stated earlier, the Taylor-Proudman constraint (that a rotating fluid is quasi-2D and aligns with the rotation axis such as $\partial\langle v_\phi\rangle/\partial z \sim 0$) is hard to break, often yielding profiles close to cylindrical rotation. One way of relaxing the Taylor-Proudman constraint in these 3-D simulations is to establish latitudinal heat transport and/or latitudinal Reynolds stresses. We actually found that in both cases these dynamical processes are present and act continuously to maintain a differential rotation against viscous effect. In 3-D models Reynolds stresses are found to transport angular momentum equatorward opposed by both the meridional circulation and viscous effects, the latter tending to erode angular velocity gradient (i.e. to make the model rotate rigidly; Brun and Toomre 2002). In more turbulent simulations Miesch et al. (2008), it is found that the viscous terms play a negligible role and the balance is mostly between Reynolds stresses and meridional flows when magnetic effects are omitted.

Turning back to the mildly turbulent cases shown in Fig. 4, we find that in the bulk of the convection zone the thermal wind balance is realised almost everywhere with $\partial\langle v_\phi\rangle/\partial z \propto \partial s/\partial\theta$. Both models possess latitudinal variation of temperature and entropy very close to each other with case AB3 having the largest contrast (10 vs. 9 K) at the base of the convection zone (BCZ). However they also have significant departure from an exact thermal wind balance, in places where Reynolds stresses are large and the shearing rate is high (close to the boundary layers). What then distinguishes the two models? In case AB3 an entropy forcing was imposed as if there was a tachocline (i.e. large variation of the angular profile at the BCZ). The small increase in entropy and temperature contrast that results from this forcing (here about 1 K) improves the agreement with the observations. The 3-D results are for that matter in good agreement with the results obtained in mean field theory and discussed in Sect. 5.1 above, even though it seems that in 3-D runs the overall contrast at

the BCZ is more important than the quality of the thermal wind balance achieved to set the shape of the angular velocity profiles. It is important to remember that at the base of the solar convection zone, the mean temperature is $\sim 2.2 \times 10^6$ K, thus a difference of only 1 K between the two models is extremely small. In a set of models published in Miesch et al. (2006), the model possessing the largest imposed temperature contrast ($\sim$13 K), yields a 'disk-like' rotation profile. It is thus very interesting to see that a difference of about 10 K (3 K vs. 13 K at the BCZ) suffices to go from a 'cylindrical-like' profile to a 'disk-like' profile. Such a high degree of sensitivity makes the exact profile of the differential rotation of solar like stars hard to predict. The fast equator and slow poles behaviour are more robust than the conical shape. Indeed in highly turbulent simulations of the solar convection it is found that the latitudinal contrast of entropy and temperature are reduced (going from 9 K to $\sim$6 K at the BCZ) and as a consequence these turbulent models are more cylindrical than their laminar counterpart. Still these models retain a fast equator and slowly rotating mid latitude, confirming the dynamical origin of the solar angular velocity profile (i.e. fast equator/slow poles). We also find that as the degree of turbulence is increased in the model the thermal forcing is less efficient to modify the iso-contours of Ω. We believe that this is in part linked to the Péclet number $Pe = VL/\kappa$ (with κ the thermal diffusivity) being larger in more turbulent cases, and to the corresponding reduced influence of the term involving the diffusion of entropy with respect to the advection term in the heat equation in determining the evolution of the entropy fluctuations. As the flow becomes more turbulent heat diffusion is less efficient and heat advection can locally overcome the slow diffusion of heat coming from the imposed forcing at the bottom boundary. In the 2-D simulations of Rempel (2005) the turbulent convective heat flux is parameterised as a diffusive process and the coupling strength is highly dependent on the overlap of the assumed turbulent diffusivity and the tachocline and therefore difficult to estimate with certainty.

This less efficient transport in turbulent case is also certainly due to the disappearance of the so-called banana-cells seen in laminar models of the solar convection (Gilman 1979; Miesch et al. 2000; Brun and Toomre 2002). As the degree of turbulence is increased the flow is becoming more chaotic and the correlations $\langle v'_\theta v'_\phi \rangle$ and $\langle v'_\theta T' \rangle$ are weaker. The appearance in highly turbulent convection state of strong convection plumes that span the whole convection depth and are tilted away from the meridional plane and from the local radial (gravity) direction may rebuild those correlations and even make them stronger (see Brummell et al. 1998). It is thus likely that by pushing the degree of turbulence in 3-D global simulations of the solar convection zone even more will result in larger latitudinal contrast of temperature and entropy (we are here speaking of few degree K) by having strong turbulent plumes dominating the transport of angular momentum and heat over the more disorganised background fields. Of course the presence of the tachocline may also help in that respect as explained above. It remains to be seen in which proportion but 3-D simulations seem to indicate that it will be at the level of a couple of degree K at the BCZ or about 10–20% of the latitudinal temperature contrast. The thinness and the sharpness of the shear in the tachocline may thus influence partly the final shape of the differential rotation realised in stars, very sharp tachoclines leading to more conical or 'disk-like' profiles (Brun 2007). Finally we wish to stress that the previous considerations omit the influence of the magnetic field on the redistribution of angular momentum, heat and the structure of the tachocline, but it is obvious that in the magnetic Sun these dynamical effects can not be neglected. We will thus address the influence of a magnetic field either self-consistently generated by dynamo action in the solar convection zone or through mean field theory and the large-scale Lorentz force in section Sect. 7 below.

6 Meridional Flow

6.1 Results from Mean Field Models

We explained already in Sect. 4 the very close connection between turbulent Reynolds stress and meridional flow, especially that the component of the Reynolds stress that transports angular momentum parallel to the axis of rotation couples most strongly with the meridional flow. The parametrisation of the Reynolds stress used in most of the mean field models discussed above has in common that in the limit of fast rotation (as defined through the Coriolis number $\Omega^* = 2\Omega\tau_c \gg 1$, with τ_c a correlation time for the turbulence) turbulent angular momentum transport becomes aligned with the axis of rotation and is inward directed. As a consequence all of these models show in the deep convection zone ($\Omega^* \approx 5$) a single celled equatorward directed flow. Some differences occur in the upper layers of the convection zone where in some parametrisation the direction of the radial component of angular momentum flux changes sign and leads to a second flow cell, which is equatorward at the surface in contradiction with observations. A recent generalisation of the Λ-effect by Kitchatinov and Rüdiger (2005) including an additional anisotropy parameter that only matters in the slow rotation regime leads to single cell flows throughout the entire convection zone. While mean field models seem to strongly prefer a single flow cell, the situation is more complicated in the case of global 3-D simulations as we discuss in the following section.

6.2 Amplitude and Profile of Meridional Circulation in 3-D Models

The meridional circulation established in 3-D models of the solar convection zone is highly time dependent (Brun and Toomre 2002; Miesch et al. 2008). A snapshot at a given instant indicates a highly intricate and small scale meridional flow but several month long temporal averages yield a much more structured flow. Its amplitude at the surface is of order 20 m/s in good agreement with surface observations. This very weak flow possesses only about 0.5% of the total kinetic energy present in the convection zone and owes its origin to a small imbalance between several large forces (buoyancy forces, Reynolds stresses, latitudinal pressure gradients and Coriolis force acting on the differential rotation; Brun and Toomre 2002; Miesch 2005). As a consequence its inner profile is hard to predict. Most 3-D models exhibits multi cellular meridional flows patterns both in latitude and radius. In these realisations, the flow is poleward near the surface up to above 45 degree and return cells are found at mid depth. The comparison with observations is hard to make since most inversions using local helioseismology techniques cease to be accurate at about 60 Mm below the surface, while most simulations predict a return flow deeper down. Still some recent models, with weakened viscous effects in the overall dynamic balance, do possess a dominant large scale circulation in the bulk of the convection zone, poleward near the surface with small counter cells at the boundaries (Miesch et al. 2008). Current 3-D numerical simulations thus seem to favour a highly time dependent multi-cellular meridional circulation, that could over long temporal averages reduce to a small number of dominant large scale cells. We also find that in 3-D models of the solar convection zone the transport of heat and kinetic energy in latitude associated with meridional circulation is weak and is not at the origin of the entropy and temperature gradients found in these simulations. As stated in Sect. 3 it is the latitudinal (poleward) enthalpy flux that establishes such gradients with the latitudinal entropy flux trying to erode them. Again we must stress that magnetic effects may modify the balance found in purely hydrodynamical models of the solar convection. In the next section we will discuss how magnetic fields may influence the meridional flows under certain conditions.

6.3 Variability of Meridional Circulation and Observational Constraints

Observations indicate that at the solar surface the meridional circulation is directed poleward with an amplitude of about 20 m/s (see Howe et al. 2006b; Meunier and Zhao, these proceedings, for more details). Local helioseismic methods (time-distance or ring diagram analysis) can probe with high accuracy meridional flows down to about 0.9–0.95 $R_\odot$, their systematic error bars getting larger with increasing depth. These inversions reveal that the flow in the near surface shear layer is relatively steady both in amplitude and direction. Deeper down the meridional circulation is found to possess larger fluctuations that can lead to the appearance of counter cells (Haber et al. 2002). Further it is found that active regions modify locally the amplitude and direction of the meridional circulation, leading to horizontal flow converging toward the active region at the surface and diverging deeper down (Haber et al. 2004; Hindman et al. 2004; Svanda et al. 2008). In Sect. 4 we pointed out that the meridional flow is closely linked to the Reynolds-stress that transports angular momentum and should show a variation comparable to the turbulent Reynolds-stress. 3D simulations exhibit a significant degree of variation as summarised above. A detailed comparison with observations is currently difficult for two main reasons: (a) most global simulations have the top of their domain located between 0.96 and 0.98 $R_\odot$, limiting the region of overlap where quantitative comparison with Doppler and local helioseismic observations can be made. High accuracy inversions of the deep structure ($r < 0.9R_\odot$) of the solar meridional flow are still missing and will help constraining furthermore the 3-D models. (b) Most helioseismic observations are averaged over substantial time intervals, i.e. the nominal temporal resolution is around a couple of days or slightly less but often an additional two-month temporal average is applied to the data to get steadier horizontal flow maps (Haber et al. 2002; Howe et al. 2006b) and are therefore not very sensitive to short term flow variability. Of course such observational constraints are not present in numerical simulations and we can choose to either consider a snapshot or temporally averaged flow. When averaging over say few weeks, the relatively high temporal fluctuations seen in consecutive snapshots are significantly reduced leading to steadier flows as with observations. Recently Rempel (2007b) pointed out that meridional and zonal flow variability are strongly coupled through the Coriolis force and therefore constraints on the variability of zonal flows also constrain the potential variability of meridional flows. Using an axisymmetric mean field model with random forcing in the Reynolds-stress parametrisation they were able to show that the 3 month averaged torsional oscillation signal would be overwhelmed by noise if the 1σ fluctuation of Reynolds-stress and meridional flow on a weekly time scale would exceed 50% of their mean values.

7 Magnetic Feedback on Mean Flows

7.1 The Influence of a Dynamo Induced Magnetic Field on the Mean Flows

In the above sections, we have discussed the establishment of the large scale flows in 3-D simulations of the solar convection in the purely hydrodynamical limit. While this has helped us to understand the subtle solar dynamics it lacks the influence of both the organised and disorganised magnetic field so obvious in observations of the solar surface. Currently it is believed that the Sun operates a hydromagnetic dynamo in its interior. This dynamo is the source of all the magnetic fields present in the Sun and that variously appear at the surface as small scale flux, active regions, coronal loops, etc.... It is a real challenge to explain

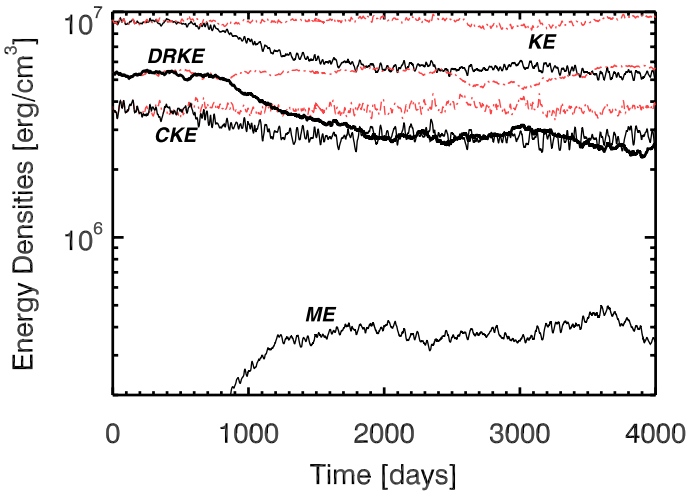


Fig. 5 Evolution of the kinetic energy (KE) split into its mean (axisymmetric) toroidal (DRKE) component (we omit the poloidal MCKE) and its non axisymmetric components (CKE), when a magnetic field is introduced or not (*solid black* vs. *dash dot red lines*) within a simulation of convection. The strong dynamo effect that develops leads to a magnetic energy (ME) of about 8% of KE. Notes the strong decrease of DRKE as ME grows above 0.1% of KE (Brun 2004)

the large variety of the solar magnetic field but it is largely admitted that the Sun runs two types of dynamo: a large scale dynamo located in and at the base of the convection zone (in the tachocline), that organises the field into a 22 yr cycle of magnetic activity, and a small scale dynamo, driven by turbulent convection that generates field over the whole velocity spectrum. The latter has been studied in detail by Cattaneo (1999). The former involves the development of global 3-D MHD models of the type that the ASH code can model. It is thus very instructive to introduce a weak magnetic field in 3-D simulations of the solar convection discussed above and to evolve self-consistently the convection model with the nonlinear feedback of the magnetic field (Brun 2004; Brun et al. 2004).

In such models, the introduction of a weak seed magnetic field leads to the generation, amplification and maintenance of a strong magnetic field, through the convection and the large scale differential rotation, when the magnetic Reynolds number ($Rm = vL/\eta$, with η the magnetic diffusivity) exceeds a certain threshold of order of a few hundreds. Of course such conditions are amply realised in the Sun, but they are harder to reach in numerical simulations. When such conditions are realised, the magnetic energy grows by many orders of magnitude to reach about 10% of the kinetic energy (KE) of the system (but this depends strongly on the rotation rate; Brun et al. 2005; Brown et al. 2007). This increasingly strong magnetic field reacts back on the flow since the Lorentz force is quadratic in B, leading to the saturation of the field strength and to a modification on how the kinetic energy is distributed between its different components (DRKE, MCKE and CKE). Figure 5 displays the temporal trace of the kinetic and magnetic energies in a successful dynamo run (with $Rm \sim 450$, Brun et al. 2004). It is clear that most of the energy transferred from the kinetic energy reservoir toward the magnetic energy reservoir comes from the decrease of the energy contained in the differential rotation (DRKE). As a direct consequence the contrast of differential rotation found in the magnetised convective case is reduced both in radius and latitude with respect to the purely hydrodynamical progenitor model, since the magnetic field tends to make the rotation rigid. The convective motions are less affected (a reduction of CKE by 20% vs

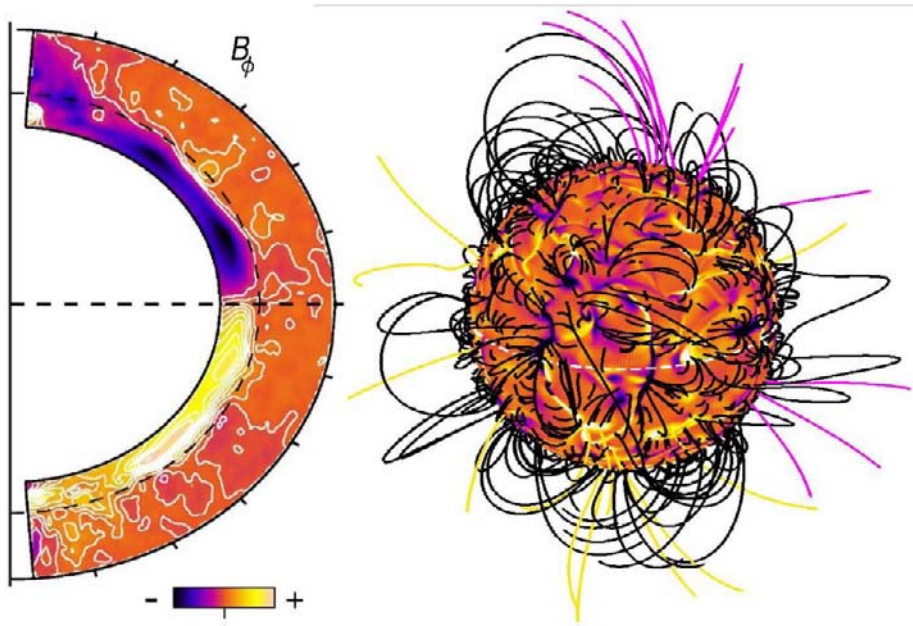


Fig. 6 *Left*: Contours of the mean (axisymmetric) toroidal field obtained in a simulation including a tachocline at the BCZ (Browning et al. 2006). Note the antisymmetric and large scale nature of the field in the tachocline. *Right*: Radial component of the magnetic field near the surface of a convective dynamo simulation including a potential reconstruction of the magnetic fields in the corona above (Brun et al. 2004)

50% for DRKE) and for that matter global models differ from their Cartesian counterpart in which most of the energy comes from the convection since these local simulations do not possess large scale flows. We thus see that the introduction of magnetic effects leads to a different angular velocity profile.

The thermal wind balance is modified by the presence of strong magnetic field, which can break the Taylor-Proudman balance. In practice the reduction of the differential rotation contrast observed in the magnetised case is associated with a smaller latitudinal temperature contrast in keeping with the balance described in Sect. 4. This suggests that it is the dynamical (Reynolds and magnetic stresses) rather than the thermal processes that drive the large scale flows by establishing fast equator and slow poles and that the thermal fields adjust accordingly and not the reverse. To better characterise the influence of magnetic field in the establishment of the differential rotation we have studied in detail the full MHD angular momentum redistribution (Brun 2004; Brun et al. 2004). In such global MHD models the contribution to the transport of angular momentum by the magnetic field is dominated by the Maxwell stresses $\langle B_r' B_\phi' \rangle$ and $\langle B_\theta' B_\phi' \rangle$ rather than by the large scale magnetic torque $\langle B_r \rangle \langle B_\phi \rangle$ and $\langle B_\theta \rangle \langle B_\phi \rangle$. This can be explained by the nature of the spectrum of magnetic field realised in the convection zone as shown in Fig. 6 (right panel), where we display B_r near the top of the domain along with the extrapolated coronal potential field lines. Strong magnetic concentrations of mixed polarity of the radial field are found in the downflow lanes and they tend to oppose locally the Reynolds stresses, reducing the vorticity generation and leading to a less efficient angular momentum transport (Brun 2004). The large scale axisymmetric fields represent less than 3% of the total magnetic energy, their dynamical role is thus very weak. However these simulations do not yet possess a cyclic axisymmetric field

and instead drive a small scale turbulent dynamo. Indeed in simulations of purely unstable convective layers, the organisation of the toroidal field into strong ribbons is not operating efficiently and the mean toroidal field remains weak (read however Brown et al. 2007 for a different story/result in the case of rapidly rotating Suns). Recently, Browning et al. (2006) have thus introduced a stable layer below the unstable zone and imposed a tachocline of shear in that region. The resulting organisation of the magnetic field is strikingly different as illustrated in Fig. 6 (left panel). In the convection zone one finds similar results to those with the purely unstable models, e.g. that the field is disorganised and rather small scale (with a spectrum peaking around $\ell \sim 20$–30). By contrast in the tachocline a large anti-symmetric, with respect to the equator, toroidal field is found. In that stable sheared layer, the toroidal field is about 10 times stronger that in the unstable convection zone above. The mean poloidal magnetic field in the solar convection zone seems to be stabilised by the presence of strong toroidal structures in the tachocline and its corresponding large scale poloidal component. While in purely unstable layers the mean poloidal field reverses too promptly (every 400 days or so), in the global simulations including a tachocline it seems to vary on a longer time scale, much closer (or even longer) than the 11-yr solar cycle. Thus the profile of the large scale differential rotation and its interaction with the magnetic fields via the ω-effect both in stable and unstable layers is key to determining the non-linear evolution of the solar convection zone, its flows and the organisation of the mean field.

Some indication of the weakening of the solar differential rotation in the presence of strong magnetic field have been observed recently (Ambroz 2004), with a weaker zonal flow during the peak of magnetic activity (see also Brajsa et al. 2007). Eddy et al. (1976) have also shown that during the Maunder minimum the Sun had a greater angular velocity contrast. Brun (2004) has shown that this is in qualitative agreement with a reduction of the magnetic energy contained in the solar convection the exact amount being model dependent.

The presence of magnetic field can also potentially modify the maintenance of the meridional circulation. However, the presence of turbulent magnetic fields do not make much difference since the meridional circulation is already extremely time dependent and the addition of new forces just make the overall inner profile of the meridional flow even more complicate to predict. The energy contained in the meridional circulation (MCKE) is even smaller than the magnetic energy (0.5% vs. $\sim$10% of KE) and do not change significantly over time contrary to DRKE that dropped by a factor of 2. However given the importance of the meridional flow profile in recent flux transport Babcock-Leighton models determining the profile of the meridional circulation and its variability with the level of magnetic activity is very important (Dikpati et al. 2004; Jouve and Brun 2007). Recent observations (Haber et al. 2002; Svanda et al. 2007, 2008; see also Meunier and Zhao, these proceedings) have shown that the surface meridional flow varies with the solar cycle and is completely dominated around active regions by the local magnetic fields. In these studies the meridional circulation is found to clearly possess a multi-cellular structure, and to vary significantly over time. Recently Jouve and Brun (2007) have studied the influence of multi cellular meridional circulations on the solar dynamo flux transport model and their consequence for the 22 yr activity cycle and resulting butterfly diagram of magnetic activity. They showed that the presence in the convection zone of several counter cells must not persist for too long if one wishes to retain the reasonably good agreement between flux transport models and the solar observations, or the solar mean field dynamo model must be reconsidered.

7.2 Results from Mean Field Models Possessing a Cyclic Magnetic Field

Observations show cyclic variations of differential rotation (torsional oscillations) with an amplitude of a few nHz. This signal was first detected through surface Doppler measurements by Labonte and Howard (1982). A more recent publication by Ulrich and Boyden (2005) shows the signal over two solar cycles. The torsional oscillation signal derived from surface Doppler measurements shows primarily an equatorward propagating branch starting at around 50 to 60 degrees latitude with a clear relation to the active region belt. Measurements in higher latitudes are less certain due to projection effects. Helioseismic inversions show in addition to the low latitude branch also a poleward propagating high latitude branch with about twice the amplitude (4 nHz) (Antia and Basu 2001; Howe et al. 2005). The latter is found to penetrate with almost constant amplitude to the base of the convection zone, while the former is found to be more shallow. Also helioseismic inversions have less accuracy in high latitudes; however, a recent analysis by Howe et al. (2006a) has shown that most of the high latitude features can be trusted at least on a qualitative level if OLA and RLS inversions are combined.

Since we do not have to date a 3-D global dynamo simulation showing cyclic activity we summarise here theoretical explanations based on mean field models of the solar dynamo. Within these approaches one can identify three classes of models depending on how the magnetic field drives large scale zonal flow variations: Macroscopic Lorentz force feedback, microscopic Lorentz force feedback, and thermal forcing. Formally thermal forcing could be classified as one of the former two (Lorentz force induced changes in convective energy transport), however, we prefer to separate it out since the flows driven by thermal forcing can have quite different properties.

The idea of macroscopic Lorentz force feedback (computed from the large scale magnetic mean field of the solar dynamo) was originally proposed by Schüssler (1981) and Yoshimura (1981) and has been incorporated into dynamo models more recently by Covas et al. (2000, 2004, 2005). While these models address the non-linear Lorentz force feedback using a simplified equations of motion (considering only the longitudinal component), models by Jennings (1993) and Rempel (2006) consider the Lorentz force feedback also in the meridional plane. The model of Rempel et al. (2005), Rempel (2006) is along the lines of the $\alpha\Lambda$-models by Brandenburg et al. (1990, 1991, 1992), Moss et al. (1995), and Muhli et al. (1995) (coupling mean field models for differential rotation, meridional flow and magnetic field evolution), but puts more emphasis on the role of the meridional flow leading to a flux-transport dynamo (see e.g. Dikpati (2005) for a recent review on the development of flux-transport dynamos).

Microscopic Lorentz force feedback (quenching of turbulent transport processes driving differential rotation 'Λ-quenching') has been addressed by Kitchatinov and Pipin (1998), Kitchatinov et al. (1999), and Küker et al. (1999).

Very recently Spruit (2003) proposed a thermal origin of the low latitude branch of torsional oscillations, driven through enhanced radiative losses in the active region belt. This theory also predicts an inflow into the active region belt, which has been observed by Komm et al. (1993), Komm (1994), and Zhao and Kosovichev (2004). The difficulty with this explanation is that the equatorward propagating torsional oscillation signal is already present in mid-latitudes a few years before the first active regions appear, which questions the idea of surface driven thermal effects. However, thermal effects might still play an important role in explaining torsional oscillations.

All Lorentz force driven models summarised above consider the longitudinal component of the Lorentz force (ϕ-direction), which is expected to be the dominant contribution in a

dynamo in which strong toroidal magnetic field is built up by the shear of differential rotation. Also the models that consider microscopic Lorentz force feedback through quenching of turbulent angular momentum transport have a longitudinal forcing as primary driver of zonal flows, since the angular momentum transport enters only in the zonal direction. On the other hand, thermal forcing drives flows through the meridional components of the momentum equation (pressure force and buoyancy). This leads to a subtle but important difference that becomes obvious in the vorticity equation (2), which takes into account only terms in the meridional plane. Since torsional oscillations are flows that vary on a time scale long compared to dynamical time scales in the convection zone, they have to be close to a thermal wind balance given by

$$r \sin\theta \frac{\partial \Omega^2}{\partial z} = \frac{g}{c_p r} \frac{\partial s}{\partial \theta} \tag{4}$$

Decomposing Ω as $\Omega_r + \Omega_t$ and s as $s_r + s_t$, with the index 'r' indicating the reference mean state and 't' indicating the perturbations associated with the torsional oscillations leads to

$$2r \sin\theta \left(\Omega_r \frac{\partial \Omega_t}{\partial z} + \Omega_t \frac{\partial \Omega_r}{\partial z} \right) = \frac{g}{\gamma r} \frac{\partial s_t}{\partial \theta} \tag{5}$$

Here the first term on the left hand side is the dominant one (typically at least a factor of 5 larger than the second one for the examples discussed here), relating directly the z derivative of Ω_t to the entropy perturbation. Therefore, if the right hand side of (5) is zero, the torsional oscillation pattern has to be very close to the Taylor-Proudman state with $\partial \Omega_t / \partial z = 0$. This conclusion is independent from the fact that the reference state differential rotation is not in the Taylor-Proudman state, since the entropy perturbation s_r drops out in this balance. Since only forces in the meridional plane potentially enter this equation, torsional oscillations driven through microscopic and macroscopic Lorentz force as summarised above have to be in the Taylor-Proudman state, while thermal forcing allows for deviations. A more detailed analysis of Rempel (2007a) concluded that while (longitudinal) Lorentz-force driving of the high latitude branch is consistent with observations, the low latitude branch requires at least some thermal contribution to explain the observed deviations from the Taylor-Proudman state. In principle Maxwell or Reynolds stresses in the meridional plane could also cause deviations from the Taylor-Proudman state, however, in most dynamo models the r and θ components of the Lorentz force are too small and to our knowledge driving of torsional oscillations by meridional Reynolds stresses has not been investigated so far. We emphasise that several of the models listed above obtain quite respectable torsional oscillations patterns (both high and low latitude branches) by neglecting completely the meridional components of the momentum equation—we expect that especially the low latitude flow patterns would change strongly if the proper momentum balance is considered. This situation is similar to early models of differential rotation that got quite respectable results but neglected the meridional flow.

In Fig. 7 we show the results of a model by Rempel (2006) that solves the full axisymmetric HD equations and couples the flows with a flux transport dynamo. The left panel shows the torsional oscillation pattern obtained by only considering macroscopic Lorentz force. Only a poleward propagating high latitude branch is present. The right panel shows results that assume a thermal perturbation in the active region belt parameterising the idea proposed by Spruit (2003). In this case also a low latitude branch is present, which is driven by a thermal perturbation of about 0.1 K. We emphasise that the origin of this thermal perturbation through enhanced radiative losses in the active region belt is controversial; however,

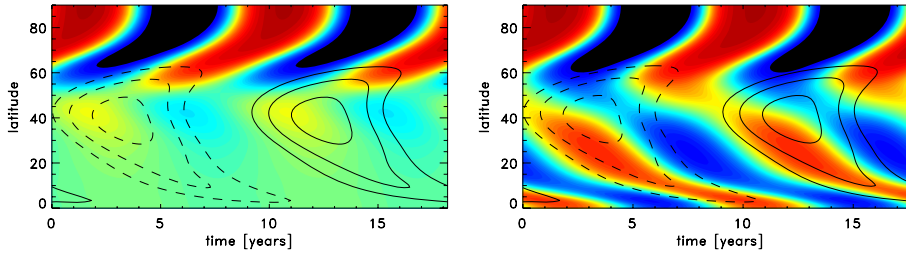


Fig. 7 Torsional oscillations (TO) obtained in non-kinematic flux-transport dynamo model. *Left*: TO patterns resulting from macroscopic Lorentz force feedback. *Right*: TO patterns resulting from including thermal forcing in the active region belt. Note: Associated temperature perturbations in active region belt are of order 0.1 K

a perturbation of only 0.1 K could also be the result of magnetically induced changes in the convective energy flux within the convection zone.

The high latitude branch is a consequence of shearing up magnetic field in the bulk of the convection zone by latitudinal shear. We strongly expect that any distributed $\alpha\Omega$ dynamo that operates on latitudinal shear in the convection zone should provide a similar pattern. The almost constant amplitude of the high latitude torsional oscillations with depth is a consequence of the Taylor-Proudman theorem that tries to minimise that variation along the axis of rotation, which does not differ too much from the radial direction in high latitudes. Since in that way the Taylor-Proudman theorem shields the exact position of the driving, it is non-trivial to try to invert the torsional oscillation signal to obtain information on distribution of magnetic field in the convection zone. The exact origin of the low latitude branch still remains a mystery, we can say however that any longitudinal forcing alone is unlikely due to the constraints arising from the Taylor-Proudman theorem.

Since the Coriolis force strongly couples zonal and meridional motions, it can be expected that zonal flow variations in the solar convection zone are accompanied by meridional flow variations with comparable amplitude. Helioseismic as well as surface Doppler measurements show on average an inflow into the active region belt of the order of 5 m/s (Komm et al. 1993; Komm 1994; Zhao and Kosovichev 2004). Thermal forcing as proposed by Spruit (2003) predicts such a flow with correct amplitude and sign close to the surface. Recently Gizon and Rempel (2008) compared results from local helioseismology with the model of Rempel (2006) and found a qualitative agreement. Both, observations and the model, indicate an outflow at about 50 Mm depth; however, the amplitude of the outflow is about an order of magnitude weaker in the model than observed. There is currently the discussion of whether the active region belt inflows are just the cumulative effect of flows around active regions, or if a there is a meridional flow component left that is independent from active regions. While carefully filtering out areas that are influenced by active regions does not change the torsional oscillation pattern, it appears that most of the meridional flow variation in the active region belt disappears when active regions are masked out, which could indicate that zonal and meridional flow variations do not have a common origin.

As a general note we emphasise here that observations of meridional flow variations associated with torsional oscillations are essential to understand the origin of torsional oscillations. As we explained above, the Taylor-Proudman theorem makes it very difficult to 'invert' torsional oscillations to obtain information on magnetic field in the convection zone. If, however, zonal and meridional flow variations are available such an inversion is more likely too succeed. As discussed in detail by Rempel (2007a), meridional flow patterns reflect in general more the location where the forcing takes place than zonal flows patterns.

Moreover, the phase between meridional and zonal flow patterns allows to discriminate between meridional and zonal forcing.

8 Conclusions

We conclude this paper by summarising the main findings in a short list bullets pointing out the similarities and differences in results obtained from mean field models and 3-D simulations.

- *Heat and energy transport in convection and resulting latitudinal gradient:*
  Both, mean field models and 3-D simulations predict a significant latitudinal (poleward directed) convective energy flux that leads to pole equator differences of several K throughout the entire convection zone. The meridional flow plays a weak role in transporting heat or kinetic energy. A realistic description of turbulence and its associated correlations is thus key to understanding the establishment of the latitudinal gradients.
- *Dominant terms for maintenance of differential rotation and meridional flow:*
  The important terms for maintaining differential rotation and meridional flow are turbulent angular momentum transport and latitudinal heat flux. While the former leads directly to an angular momentum transport, the latter influences angular momentum transport through modification of the meridional flow. The meridional flow is driven through Coriolis forces resulting from differential rotation as well as buoyancy forces resulting from latitudinal entropy variations. The profile of the meridional flow is set through the constraint that the total angular momentum flux has to be divergence free in a stationary state. In 3-D simulations as well as most mean field models the turbulent Reynolds stress is main driver of differential rotation, while thermal perturbations primarily correct the differential rotation profile (deviation from Taylor-Proudman state; Brun and Toomre 2002; Miesch et al. 2008), but the influence of the thermal forcing coming from the tachocline seems to depend on the Peclet number $Pe = VL/\kappa$ realised in the 3-D simulations. A more complicated situation is encountered in mean field models using the Λ-effect parameterizations of Kitchatinov and Rüdiger (1993) and Kitchatinov and Rüdiger (2005) that predict an angular momentum flux parallel to the axis of rotation in the limit of fast rotation. Such an angular momentum transport can be offset exactly by the meridional flow, preventing the build up of any differential rotation. In this situation a latitudinal variation of entropy is essential for obtaining differential rotation in the first place.
- *Role of anisotropic heat transport vs. tachocline in setting the contrast and profile of differential rotation:*
  The relative contribution of both effects is currently unknown. 3-D models seem to indicate a contribution of a couple of degree K or 10–20% to the overall entropy and temperature latitudinal gradients but this may depend on the actual thickness of the solar tachocline (Brun 2007). While in the mean field models of (Küker and Stix 2001; Küker and Rüdiger 2005a, 2005b, 2007) anisotropic heat transport alone is sufficient to explain the observed deviation from the Taylor-Proudman state, this is not the case for most 3-D simulations (Miesch et al. 2006). Here an additional source of latitudinal entropy variation such as the tachocline source proposed by Rempel (2005) is required. On the other hand the contribution of the tachocline is very sensitive to the coupling between tachocline and convection zone, i.e. the detailed properties of the overshoot region, which is still under discussion.

- *Uni vs. multi cellular meridional flows and the current model predictions, potential problems for Babcock-Leighton flux transport dynamo models:*
  Most mean field models predict unicellular meridional flow with a direction (poleward at top and equatorward at bottom of convection zone) very favourable for Babcock-Leighton flux transport dynamo models. In contrast to this most 3-D simulations prefer multi cellular flow; however, the high resolution run of Miesch et al. (2008) is also supporting more the unicellular flow topology. The robustness of the latter result has to be confirmed through more high resolution runs that will become feasible in the near future. The high time variability of the meridional flow is not a significant problem for Babcock-Leighton flux transport dynamo models as it has been investigated by Charbonneau and Dikpati (2000) and Rempel (2007b). However, long lasting counter cells may have a visible impact on the butterfly diagram or the activity cycle frequency (Jouve and Brun 2007). Independent from the detailed structure of the meridional flow the relative contribution of turbulent transport processes (anisotropic magnetic diffusivity and turbulent pumping) is still an open, but very crucial question (flux transport dynamos rely on the assumption that turbulent transport is weak compared to advective transport).
- *Magnetic feed back on mean flows both through dynamo induced field and from large scale Lorentz torques:*
  Both, large and small scale Maxwell stresses have the tendency to reduce the amount of differential rotation (Brun 2004). In 3-D models of magnetised convection it is found that large scale mean magnetic torques are weak compared to Maxwell stresses which play a significant role in the angular momentum transport balance (Brun et al. 2004). This balance is thus modified by the presence of magnetic fields and may lead to a different meridional circulation profiles than in the purely hydrodynamical case, since more than two terms (Reynolds stresses and meridional flow) now enter the balance. Including a tachocline at the BCZ helps generating large scale and axisymmetric magnetic fields and as a consequence the large scale magnetic torques become more important leading to a more subtle angular momentum balance involving now four terms Browning et al. (2006). In mean field models with a cyclic large scale dynamo the large scale component of the Maxwell stress (dominated by the $r\phi$ and $\theta\phi$ components) leads additionally to a cyclic variation of differential rotation (torsional oscillations) and meridional flow. If the meridional flow is considered self-consistently the cyclic component of zonal flows has to be in the Taylor-Proudman state. Significant deviations of zonal flows from the Taylor-Proudman state as observed in low latitudes require additional forcing terms beside the zonal component of the Lorentz force (e.g. thermal forcing).
- *Future perspective: what's next in 2-D and 3-D simulations:*
  With the increasing computing power, 3-D simulations are expected to become the dominant tool for understanding solar and stellar convection zone dynamics (see for instance Brun et al. 2005; Brown et al. 2007). While mean field models are very convenient in exploring a wide range of parameter ranges, they are restricted by the parametrisation of turbulent transport processes used. Especially large scale coherent flow structures as they are found in global 3-D simulations are difficult to incorporate in mean field models and are currently not considered. Improved physical description or constraints on the parametrisation may be incorporated in mean field models by extracting from 3-D high resolution simulations averaged turbulent quantities and profiles of key processes such as turbulent angular momentum transport processes (Λ-effect, turbulent viscosity) and turbulent induction effects (α-effect, turbulent magnetic pumping, flux emergence).

Acknowledgements The authors are thankful to the organisers of the ISSI workshop "Origin and Evolution of Solar Magnetism" held in Bern in January 2008 for their invitation. A.S.B. is grateful to J. Toomre,

M.S. Miesch and J.-P. Zahn for more than 10 years of fascinating discussions and collaborations on solar and stellar fluid dynamics. M.R. acknowledges that the National Center for Atmoshperic Research is sponsored by the National Science Foundation.

References

P. Ambroz, Sol. Phys. **224**, 61 (2004)
H.M. Antia, S. Basu, Astrophys. J. Lett. **559**, 67 (2001)
R. Brajsa, H. Wohl, D. Ruzdjak, B. Vrsnak, G. Verbanac, L. Svalgaard, J.-F. Hochedez, Astron. Not. **328**, 1013 (2007)
A. Brandenburg, I. Tuominen, D. Moss, G. Ruediger, Sol. Phys. **128**, 243 (1990)
A. Brandenburg, D. Moss, G. Ruediger, I. Tuominen, Geophys. Astrophys. Fluid Dyn. **61**, 179 (1991)
A. Brandenburg, D. Moss, I. Tuominen, Astron. Astrophys. **265**, 328 (1992)
B.P. Brown, M.K. Browning, A.S. Brun, M.S. Miesch, N. Nelson, J. Toomre, Am. Inst. Phys. Conf. Proc. **948**, 271 (2007)
M. Browning, M.S. Miesch, A.S. Brun, J. Toomre, Astrophys. J. **648**, L157 (2006)
N.H. Brummell, N.E. Hurlburt, J. Toomre, Astrophys. J. **493**, 955 (1998)
A.S. Brun, Sol. Phys. **220**, 333 (2004)
A.S. Brun, Astron. Not. **328**, 329 (2007)
A.S. Brun, J. Toomre, Astrophys. J. **570**, 865 (2002)
A.S. Brun, M.S. Miesch, J. Toomre, Astrophys. J. **614**, 1073 (2004)
A.S. Brun, M. Browning, J. Toomre, Astrophys. J. **629**, 461 (2005)
B. Castaing, G. Gunaratne, L. Kadanoff, A. Libchaber, F. Heslot, J. Fluid Mech. **204**, 1 (1989)
F. Cattaneo, Astrophys. J. **515**, L39 (1999)
F. Cattaneo et al., Astrophys. J. **370**, 282 (1991)
S. Chandrasekhar, *Hydrodynamic and Hydromagnetic Stability* (Oxford University Press, Oxford, 1961)
P. Charbonneau, M. Dikpati, Astrophys. J. **543**, 1027 (2000)
T.L. Clune, J.R. Elliott, G.A. Glatzmaier, M.S. Miesch, J. Toomre, Parallel Comput. **25**, 361 (1999)
E. Covas, R. Tavakol, D. Moss, A. Tworkowski, Astron. Astrophys. **360**, L21 (2000)
E. Covas, D. Moss, R. Tavakol, Astron. Astrophys. **416**, 775 (2004)
E. Covas, D. Moss, R. Tavakol, Astron. Astrophys. **429**, 657 (2005)
M. Dikpati, Adv. Space Res. **35**, 322 (2005)
M. Dikpati, P.S. Cally, P.A. Gilman, Astrophys. J. **610**, 597 (2004)
B.R. Durney, H.C. Spruit, Astrophys. J. **234**, 1067 (1979)
J.A. Eddy, P.A. Gilman, D.E. Trotter, Sol. Phys. **46**, 3 (1976)
J.R. Elliott, M.S. Miesch, J. Toomre, Astrophys. J. **533**, 546 (2000)
P.A. Gilman, Astrophys. J. **231**, 284 (1979)
P.A. Gilman, Astrophys. J. Suppl. Ser. **53**, 243 (1983)
L. Gizon, M. Rempel, Sol. Phys. **249**, 58 (2008)
D.A. Haber, B.W. Hindman, J. Toomre, R.S. Bogart, R.M. Larsen, F. Hill, Astrophys. J. **570**, 855 (2002)
D.A. Haber, B.W. Hindman, J. Toomre, M.J. Thompson, Sol. Phys. **220**, 371 (2004)
C.J. Hansen, S.D. Kawaler, *Stellar Interiors: Physical Principles, Structure, and Evolution* (Springer, New York, 1995)
D. Hathaway, Astrophys. J. **276**, 316 (1984)
B.W. Hindman, L. Gizon, T.L. Duvall Jr., D.A. Haber, J. Toomre, Astrophys. J. **613**, 1253 (2004)
R. Howe, J. Christensen-Dalsgaard, F. Hill, R. Komm, J. Schou, M.J. Thompson, Astrophys. J. **634**, 1405 (2005)
R. Howe, M. Rempel, J. Christensen-Dalsgaard, F. Hill, R.M. Komm, R.W. Larsen, J. Schou, M.J. Thompson, Astrophys. J. **649**, 1155 (2006a)
R. Howe, R.M. Komm, F. Hill, R. Ulrich, D.A. Haber, B.W. Hindman, J. Schou, M.J. Thompson, Sol. Phys. **235**, 1 (2006b)
B.J. Labonte, R. Howard, Sol. Phys. **75**, 161 (1982)
R.L. Jennings, Sol. Phys. **143**, 1 (1993)
L. Jouve, A.S. Brun, Astron. Astrophys. **474**, 239 (2007)
M. Küker, M. Stix, Astron. Astrophys. **366**, 668 (2001)
L.L. Kitchatinov, V.V. Pipin, Astron. Rep. **42**, 808 (1998)
L.L. Kitchatinov, G. Rüdiger, Astron. Astrophys. **276**, 96 (1993)
L.L. Kitchatinov, G. Rüdiger, Astron. Astrophys. **299**, 446 (1995)
L.L. Kitchatinov, G. Rüdiger, Astron. Nachr. **326**, 379 (2005)

L.L. Kitchatinov, V.V. Pipin, G. Rüdiger, Astron. Nachr. **315**, 157 (1994)
L.L. Kitchatinov, V.V. Pipin, V.I. Makarov, A.G. Tlatov, Sol. Phys. **189**, 227 (1999)
R.W. Komm, Sol. Phys. **149**, 417 (1994)
R.W. Komm, R.F. Howard, J.W. Harvey, Sol. Phys. **147**, 207 (1993)
M. Küker, G. Rüdiger, Astron. Astrophys. **433**, 1023 (2005a)
M. Küker, G. Rüdiger, Astron. Nachr. **326**, 265 (2005b)
M. Küker, G. Rüdiger, Astron. Nachr. **328**, 1050 (2007)
M. Küker, R. Arlt, G. Rüdiger, Astron. Astrophys. **343**, 977 (1999)
M. Lesieur, *Turbulence in Fluid* (Kluwer, Dordrecht, 1997)
M.S. Miesch, Living Rev. Sol. Phys. **2**, 1 (2005)
M.S. Miesch, J.R. Elliott, J. Toomre, T.L. Clune, G.A. Glatzmaier, P.A. Gilman, Astrophys. J. **532**, 593 (2000)
M.S. Miesch, A.S. Brun, J. Toomre, Astrophys. J. **641**, 618 (2006)
M.S. Miesch, A.S. Brun, M. Derosa, J. Toomre, Astrophys. J. **673**, 557 (2008)
D. Moss, D.M. Barker, A. Brandenburg, I. Tuominen, Astron. Astrophys. **294**, 155 (1995)
P. Muhli, A. Brandenburg, D. Moss, I. Tuominen, Astron. Astrophys. **296**, 700 (1995)
M. Rempel, Astrophys. J. **622**, 1320 (2005)
M. Rempel, Astrophys. J. **647**, 662 (2006)
M. Rempel, Astrophys. J. **655**, 651 (2007a)
M. Rempel, Astron. Nachr. **328**, 1096 (2007b)
M. Rempel, M. Dikpati, K. MacGregor, in *CS13 Proceedings*, ESA SP-560 (2005), p. 913
G. Rüdiger, R. Hollerbach, *The Magnetic Universe: Geophysical and Astrophysical Dynamo Theory* (Wiley-VCH, New York, 2004) (ISBN 3-527-40409-0)
G. Rüdiger, P. Egorov, L.L. Kitchatinov, M. Küker, Astron. Astrophys. **431**, 345 (2005a)
G. Rüdiger, P. Egorov, U. Ziegler, Astron. Nachr. **326**, 315 (2005b)
M. Schüssler, Astron. Astrophys. **94**, L17 (1981)
H.C. Spruit, Sol. Phys. **213**, 1 (2003)
M. Svanda, A.G. Kosovichev, J. Zhao, Astrophys. J. **670**, L69 (2007)
M. Svanda, A.G. Kosovichev, J. Zhao, Astrophys. J. **680**, L161 (2008)
M.J. Thompson, J. Christensen-Dalsgaard, M.S. Miesch, J. Toomre, Ann. Rev. Astron. Astrophys. **41**, 599 (2003)
R.K. Ulrich, J.E. Boyden, Astrophys. J. **620**, 123 (2005)
H. Yoshimura, Astrophys. J. **247**, 1102 (1981)
J. Zhao, A.G. Kosovichev, Astrophys. J. **603**, 776 (2004)

Photospheric and Subphotospheric Dynamics of Emerging Magnetic Flux

A.G. Kosovichev

Originally published in the journal Space Science Reviews, Volume 144, Nos 1–4, 175–195.
DOI: 10.1007/s11214-009-9487-8

Abstract Magnetic fields emerging from the Sun's interior carry information about physical processes of magnetic field generation and transport in the convection zone. Soon after appearance on the solar surface the magnetic flux gets concentrated in sunspot regions and causes numerous active phenomena on the Sun. This paper discusses some properties of the emerging magnetic flux observed on the solar surface and in the interior. A statistical analysis of variations of the tilt angle of bipolar magnetic regions during the emergence shows that the systematic tilt with respect to the equator (the Joy's law) is most likely established below the surface. However, no evidence of the dependence of the tilt angle on the amount of emerging magnetic flux, predicted by the rising magnetic flux rope theories, is found. Analysis of surface plasma flows in a large emerging active region reveals strong localized upflows and downflows at the initial phase of emergence but finds no evidence for large-scale flows indicating future appearance a large-scale magnetic structure. Local helioseismology provides important tools for mapping perturbations of the wave speed and mass flows below the surface. Initial results from SOHO/MDI and GONG reveal strong diverging flows during the flux emergence, and also localized converging flows around stable sunspots. The wave speed images obtained during the process of formation of a large active region, NOAA 10488, indicate that the magnetic flux gets concentrated in strong field structures just below the surface. Further studies of magnetic flux emergence require systematic helioseismic observations from the ground and space, and realistic MHD simulations of the subsurface dynamics.

Keywords Solar magnetism · Magnetic flux · Active regions · Sunspots · Helioseismology

1 Introduction

The current paradigm is that the solar magnetic fields are generated by a dynamo action deep in the convection zone, presumably, at the bottom, in a thin rotational shear layer

A.G. Kosovichev (✉)
Stanford University, Stanford, CA 94305, USA
e-mail: AKosovichev@solar.stanford.edu

M.J. Thompson et al. (eds.), *The Origin and Dynamics of Solar Magnetism*.
DOI: 10.1007/978-1-4419-0239-9_11

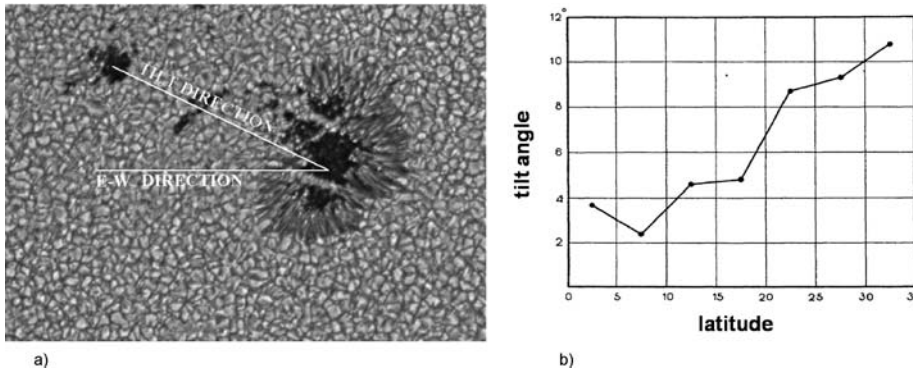


Fig. 1 (**a**) Illustration of the Joy's law (courtesy of D. Hathaway). (**b**) The tilt angle as a function of latitude (Hale et al. 1919)

called tachocline. In the tachocline the solar differential rotation changes from the differential rotation of the convection zone to a solid-body rotation of the radiative core. Most of the tachocline is located in a convectively stable zone mixed by convective overshoot (Kosovichev 1996). The combination of strong shearing flows and stability makes possible generation and storage of magnetic field in the tachocline (Parker 1993).

When the magnetic field is sufficiently strong it becomes buoyant and emerges in the form of toroidal flux ropes ("Ω-loops") oriented in the East–West direction forming bipolar active regions on the surface. The rising magnetic loops are affected by the Coriolis force, which induces retrograde flows, directed from the leading part of the toroidal tube towards its following part. The Coriolis force acting on these flows causes deflection of the flux tubes to higher latitudes and also a tilt with respect to the equator. However, observations show that the magnetic flux emerges mostly at mid and low latitudes, and calculations demonstrate that to explain this the magnetic field generated at the bottom of the convection zone must be very strong, 60–160 kG (D'Silva 1992; D'Silva and Howard 1994; Parker 1994). This is significantly higher than the field strength estimated from energy equipartition with convective motions. Whether this is possible is under debate (e.g. Schüssler 2005).

In addition, the Coriolis force causes a tilt in the orientation of emerging flux tubes with respect to the equator. This phenomenon is well-known as the Joy's law. Statistical studies by A.H. Joy (Hale et al. 1919) of long series of sunspot drawings showed that the following spot of a bipolar group tend to appear farther from the equator than the preceding spot, and that the higher the latitude the greater the inclination of the axis to the equator Fig. 1. The tilt of bipolar magnetic groups appears only statistically. The orientation of individual active regions may vary quite significantly. This effect is important for understanding the process of magnetic flux emergence. Also, it is a key element of solar-cycle theories (e.g. Wang and Sheeley 1991).

Theories of rising magnetic flux tubes seem to explain the tilt (Schmidt 1968). In these theories, the tilt angle is determined by the torque balance between magnetic tension and the Coriolis forces, and, thus, depends on both the amount of magnetic flux and the emergence latitude. Fisher et al. (1995) and Caligari et al. (1995) found from an analysis of sunspot group data that the theoretically predicted flux dependence of the tilt angle was consistent with the data. Tian and Liu (2003) updated these results using magnetic fluxes instead of polarity separations for a more direct comparison with the theory. Their results showed

a less clear agreement between the rising flux tube theories and the observations. These theories also predict that after the rising phase the magnetic flux tubes become stationary, and since the Coriolis force disappears the tilt should decrease under the action of magnetic tension in the East–West direction. However, the tilt of bipolar groups does not disappear after the emergence (Howard 2000). Moreover, there is a tendency of the tilt angle to rotate towards the averaged value defined by the Joy's law. This cannot be explained by the Coriolis effect, and is probably related to a complicated interaction between the magnetic structures and flows below the surface. An alternative explanation suggested by Babcock (1961) is that the tilt is due to the spiral orientation of the magnetic field lines below the surface, wrapped around the Sun by the differential rotation. However, his mechanism assumes that the toroidal field is generated close to the surface, not in the tachocline.

In general, understanding of the observed properties of emerging magnetic flux is closely related to the depth of the main dynamo process in the Sun. While the modern theories assume that the solar dynamo operates in the tachocline there is no convincing observational evidence to support this. Also, there are theoretical difficulties. The pros and cons of the dynamo mechanisms operating in the tachocline and in the bulk of the convection zone, or perhaps even in the near-surface shear layer, are discussed by Brandenburg (2005). From the observational point of view, helioseismology observations for the whole solar cycle from the Michelson Doppler Imager (MDI) instrument on Solar and Heliospheric Observatory (SOHO) (Scherrer et al. 1995) and Global Oscillation Network Group (GONG) (Harvey et al. 1996) do not provide a convincing evidence for solar-cycle variations of the solar rotation rate in the tachocline (Howe et al. 2007). Such variations are expected because of the back reaction of the strong dynamo-generated magnetic fields on the turbulent Reynolds stresses and, hence, on the differential rotation (which is maintained by the Reynolds stresses). Moreover, the comparison of the rotation rate of long-living complexes of activity (which are the sources of repeated flux emergence) with the internal differential rotation deduced from helioseismology showed that the roots of the complexes of activity are probably located in the near-surface shear layer (Benevolenskaya et al. 1999). Determination of the depth of the solar dynamo is one of the most important problems of solar magnetism.

Observations show that emerging magnetic flux plays an important role in initiation of solar flares and coronal mass ejections. Thus, it is important to develop predicting capabilities for flux emergence. This problem can be addressed by helioseismology but initial attempts to detect the magnetic flux in the interior before it becomes visible on the surface showed that this is difficult because of the high emergence speed in the upper 20 Mm (Kosovichev et al. 2000). Thus, it is important to investigate large-scale flow patterns, which may provide indication of the flux emergence and development of large magnetic regions in the interior.

In general, investigation of emerging magnetic flux includes the following questions:

- How deep is the source of emerging magnetic flux?
- Does emerging magnetic flux become disconnected from the source?
- Why does magnetic flux tend to emerge in the same areas, forming complexes of activity?
- What is the plasma dynamics associated with emerging flux?
- How does emerging flux interact with the existing magnetic fields, and triggers flares and CMEs?
- Can we predict emerging magnetic flux before it become visible on the surface?
- Can we predict evolution of active regions?

This paper discusses some of these questions and presents recent results of investigation of surface and subsurface characteristics of the magnetic flux emergence process obtained

from SOHO/MDI and GONG. In particular, I discuss a new analysis of the Joy's law for the emerging flux, dynamics of the photospheric plasma prior and during the flux emergence, methods and results of acoustic tomography of wave-speed perturbations and mass flows below the visible surface, and also future observational projects and perspectives.

2 Observations of Emerging Magnetic Flux in the Photosphere

2.1 Joy's Law and Magnetic Flux Transport

The tilt of bipolar sunspot groups with respect to the equator (the Joy's law) is one of the fundamental properties of solar magnetism. This phenomenon is closely related to the dynamo mechanism and the process of flux emergence. The key question is whether the tilt is caused by the Coriolis force acting on magnetic flux tubes radially moving from the bottom of the convection zone (Schmidt 1968), or it reflects the orientation of subsurface magnetic field lines stretched by the differential rotation (Babcock 1961), or it is created by subphotospheric shearing flows after the emergence (Howard 1996). Previous studies of the Joy's law were based on daily white light images or magnetograms of sunspot group. These data did not have sufficient temporal resolution to investigate variations of the tilt angle during the flux emergence process.

Using a series of 96-min cadence magnetograms from SOHO/MDI, Kosovichev and Stenflo (2008) attempted to investigate the tilt angle and its statistical relationships to the region latitude, the amount of emerging flux, the emergence rate and the separation between the magnetic polarities. The magnetograms obtained almost uninterruptedly for almost the whole solar cycle, from May 1996 until October 2006, have been analyzed. During this period the MDI instrument on SOHO observed more than 2000 active regions, and 715 active regions, which emerged within 30 degrees from the central meridian, are selected for this study.

The analysis method is pretty straightforward. Each active region is remapped into the heliographic coordinates. The tilt angle and the separation between the magnetic polarities are calculated for their centers of gravity. The period of the growth of the total magnetic flux (lasting usually 2–3 days) is divided into 5 intervals, and the statistical relations are calculated for each interval separately, and for the whole emergence phase.

A typical example of this data analysis is shown in Fig. 2 for active region NOAA 8167 emerged in the Southern hemisphere at about 26° latitude. The magnetic flux of both polarities rapidly and simultaneously increased and reached a maximum during the first 2 days after the initial appearance of the bipolar region on the surface (Fig. 2c). The tilt angle (Fig. 2d) shows rapid variations at the beginning of emergence, reaches a maximum of about 15° and then stabilizes at about 7–8°. The separation between the polarities (Fig. 2e) increases during the emergence phase and continues after the total flux reaches the maximum. It starts decreasing as the active region decays. Figures 2f–h show the relationships among these properties. In this example the tilt angle is established rather quickly during the emergence in accordance with the Joy's law (the blue polarity is closer to the equator than the red polarity in Fig. 2b). However, the title angle varies significantly in our sample of emerging active regions, and the Joy's law holds only statistically.

Figures 3a–c show the distributions of the tilt angles with latitude for three periods of emergence (the total flux growth): the initial appearance (emergence interval 1), the mid interval (3) and at the end of emergence (interval 5). It appears that at the beginning of emergence the tilt angle is randomly distributed, and the mean tilt angle is about zero (Fig. 3a).

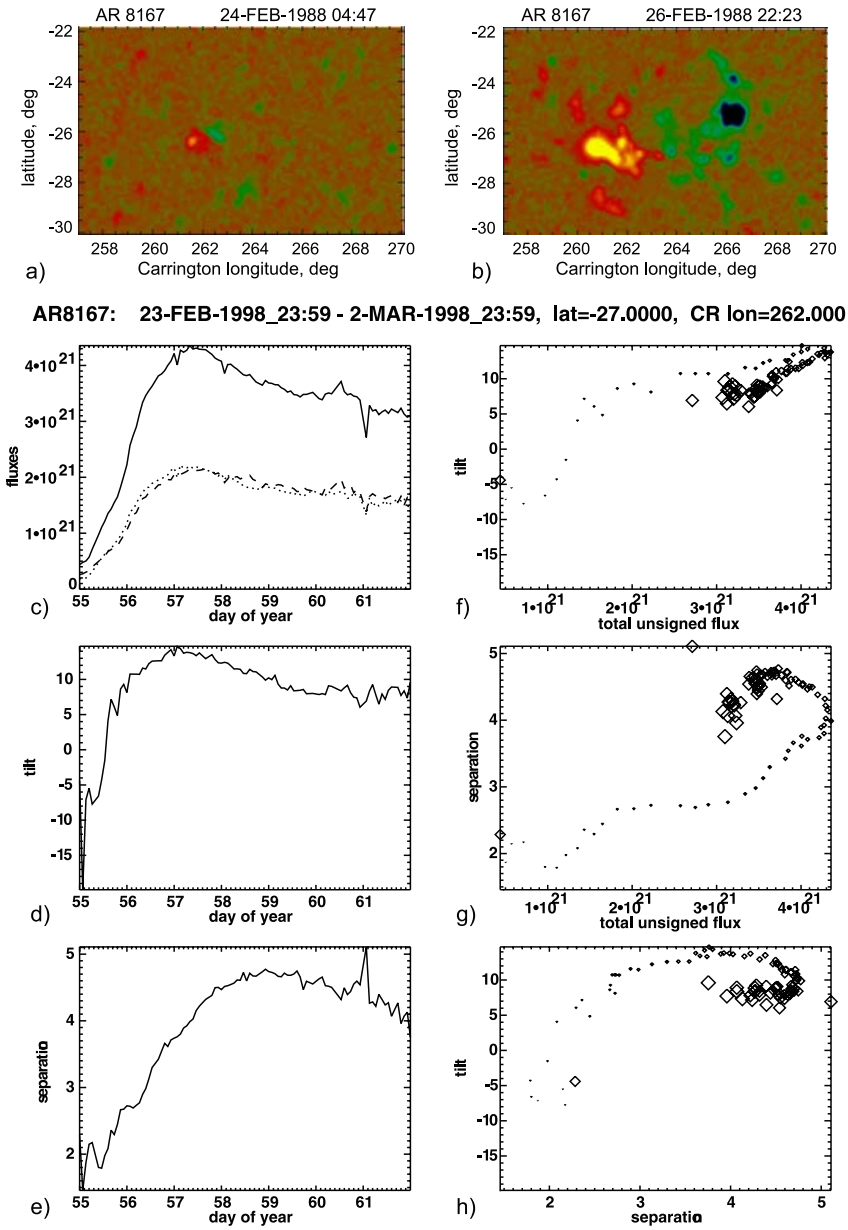


Fig. 2 Magnetograms of active region NOAA 8167: (**a**) at the beginning of flux emergence and (**b**) at the end of emergence. The *green and blue colors* show the negative polarity; the *red and yellow colors* show the positive polarity. The evolutions of: (**c**) magnetic fluxes in Mx (*dashed curve*—positive polarity, *dotted curve*—negative polarity, the *solid curve*—the total unsigned flux); (**d**) the tilt angle (in degrees); (**e**) the separation between the polarities (in heliographic degrees). The relationships between: (**f**) the tilt angle and total unsigned flux, (**g**) the polarity separation and the total flux, (**h**) the tilt angle and the separation. The symbol size is proportional to time from the start of emergence

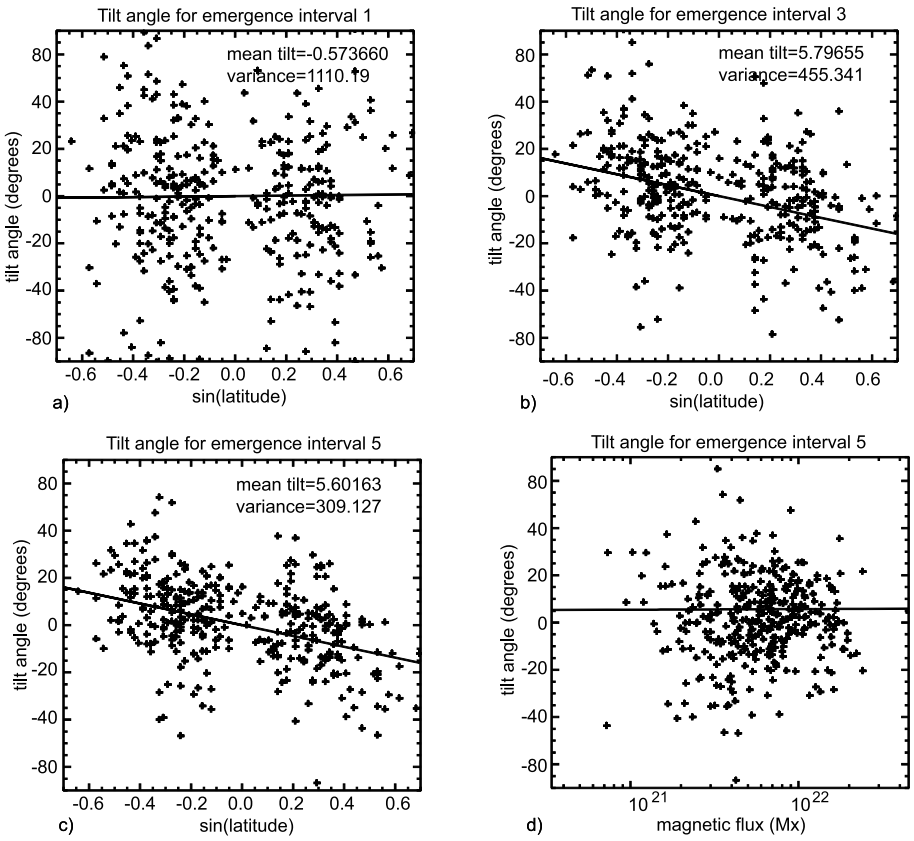


Fig. 3 The distribution of the tilt angle with sine latitude at the beginning of emergence (**a**), at the middle of the emergence interval (**b**), and at the end of emergence (**c**); the distribution of the tilt angle with the total magnetic flux at the end of emergence

However, at the middle of the emergence period the distribution of the tilt angle clearly follows the Joy's law (Fig. 3b) with the latitudinal dependence and the mean tilt angle of about 6 degrees. At the end of the emergence period the Joy's law distribution becomes more pronounced as the variance of the deviation from the linear dependence on the sine latitude decreases (Fig. 3c). However, these data show no significant correlation between the tilt angle and the total magnetic flux at the end of emergence (Fig. 3d).

These results show that the tilt of bipolar magnetic regions becomes statistically significant during the emergence process. This means that the tilt is established in subsurface layers. Among the mechanisms suggested to explain the observed tilt are: the spiral orientation of the subsurface toroidal field lines wrapped around the differential rotation (Babcock 1961), the effect of the Coriolis force acting on the flux tubes moving from the bottom of the convection zone (Schmidt 1968), and large-scale subsurface motions associated with the differential rotation and meridional circulation (Howard 1996). The most popular explanation that the tilt is caused by the Coriolis force acting on the flows inside an emerging flux rope was questioned by Howard (1996) who investigated variations of the tilt after the emergence and found that the tilt angle moves towards the Joy's law orientation instead of relaxing to the East–West direction as expected from this theory when the radial flux rope motion stops

(and, thus, the Coriolis force vanishes). In addition, our results do not show a significant dependence of the tilt on the magnetic flux, predicted by the Coriolis force theories (e.g. Fan et al. 1994). Howard (1996) suggested that the tilt angle may be established after the emergence due to the action of the depth dependent differential rotation and meridional flow. However, our results indicate that the bipolar magnetic flux regions emerge already tilted in accordance with the Joy's law. The helioseismology results show that the emerging magnetic structures propagate very fast in the upper convection zone (Sect. 3.2), and thus do not support the Howard's idea. Perhaps, we should go back to the Babcock's mechanism that the tilt is caused by the spiral structure of the subsurface toroidal flux tubes.

2.2 Mass Flows

One can expect that when a large magnetic flux rope emerges on the solar surface it drives significant upflows and outflows, which may be detectable when the flux rope is still below the surface. Figure 4 shows the magnetograms and Dopplergrams obtained from

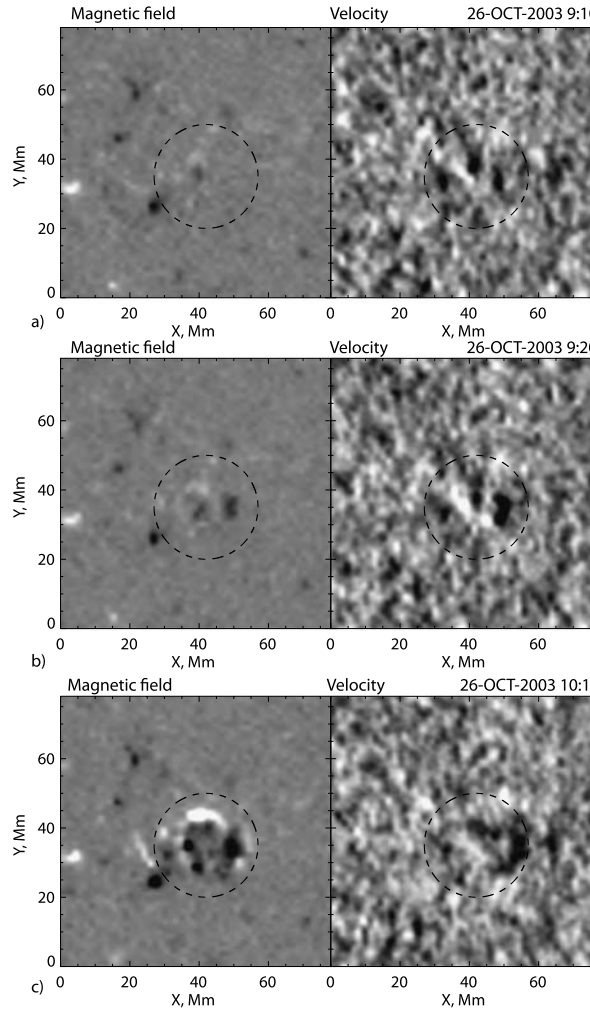


Fig. 4 Maps of the line-of-sight magnetic field (*left panels*) and Doppler velocity (*right panels*) on the solar surface obtained from SOHO/MDI at the beginning of emergence of AR 10488, October 26, 2008, (**a**) 7:30 UT, (**b**) 7:40 UT, (**c**) 7:50 UT. The range of the magnetic field strength is [−180 G, 180 G]. The range of the Doppler velocity is [−600 m/s, 600 m/s]. The *dark color* shows upflows, and the *white color* shows downflows. The *dashed circle* outlines the area of the initial magnetic flux emergence

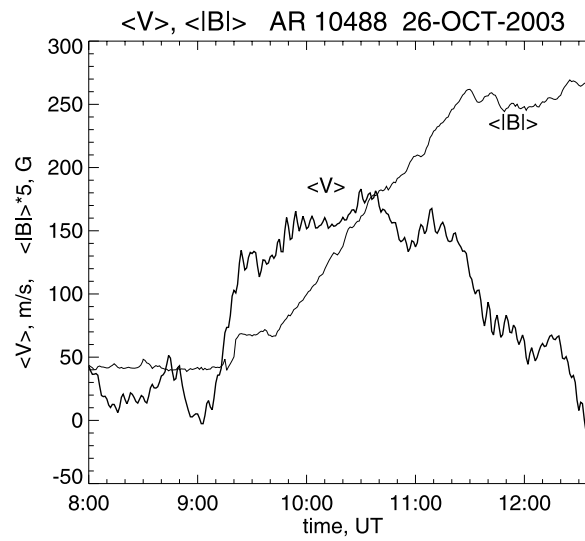


Fig. 5 The mean Doppler velocity and the total magnetic flux as a function of time in the region of the initial emergence of AR 10488. The positive velocity values correspond to the plasma motions towards the observer (upflow)

SOHO/MDI on October 26, 2003, during the initial emergence of AR 10488, which later grew in to one of the largest active regions of Solar Cycle 23. These data reveal strong localized upflows (dark features in the velocity images) in the places of the initial magnetic flux growth at 9:10 UT (Fig. 4a). Ten min later the upflow velocity reaches a peak of about 800 m/s, and the magnetic flux starts appearing on the surface (Fig. 4b). The strong upflow is mostly concentrated in the leading part of the emerging flux. In the following part, the data show the Doppler shift of the opposite sign corresponding to downflows. After the appearance of the magnetic flux and its initial growth we observe the similar flow pattern with upflows in the leading part, but the velocity amplitude decreases (Fig. 4c).

The evolution of the mean Doppler velocity and the total magnetic flux in the area of emergence (Fig. 5) shows a sharp rise of upflows, which continue to be strong for about 2 hours during the initial emergence. The mean velocity reaches ∼150 m/s, well above the mean velocity fluctuations of about 50 m/s in similar-sized quiet-Sun regions. After the initial emergence phase, we do not observe significant upflows despite the continuing growth of the active regions.

The strong photospheric plasma flows associated with the magnetic flux emergence of this active region have been also detected by Grigor'ev et al. (2007). It is unclear if such surface flows are typical for emerging magnetic flux and if their strength corresponds to the amount of emerging flux. Pevtsov and Lamb (2006) studied plasma flows in fifteen emerging active regions using Dopplergrams, magnetograms, and white light observations from SOHO/MDI. They observed no consistent plasma flows at the future location of an active region before its emergence.

Also, in the case of AR10488 there is a systematic upflow in the leading magnetic polarity and a downflow in the following polarity. This can interpreted as flows inside the emerging flux tubes, driven by the Coriolis force (e.g. Fan et al. 1994). However, Pevtsov and Lamb (2006) found the asymmetric flows only in three active regions. In two regions, flows are directed from the following to leading polarity, and in one region material flows from the leading to the following polarity. Thus, more detailed statistical studies of the flow dynamics in emerging magnetic flux are necessary.

3 Investigations of Emerging Flux by Helioseismology

Methods of local helioseismology developed in recent years allow us to probe physical conditions of the solar plasma below the surface and detect the wave-speed structures and mass flows associated with the emerging magnetic flux.

3.1 Method of Time–Distance Helioseismology

Time–distance helioseismology measures travel times of acoustic waves propagating to different distances, and uses these measurements to infer variations of the wave speed along the wave paths. Turbulent convection excites acoustic waves which propagate deep into the solar interior. Because the sound speed increases with depth these waves are refracted and come back to the solar surface. The wave speed depends on the temperature, magnetic field strength and flow velocity field in the region of the wave propagation. By measuring reciprocal travel times of acoustic waves propagating along the same ray paths in opposite directions, and then taking the mean and the difference of these travel times, it is possible to separate the flow velocity (advection) effect from temperature and magnetic field perturbations (Kosovichev and Duvall 1997). However, in order to disentangle the contributions of temperature variations and magnetic field to the mean travel times it is necessary to measure the travel-time anisotropy, and this has not been accomplished. Therefore, the current helioseismic results represent maps of sub-photospheric variations of the magneto-acoustic wave speed and flow.

The travel times are typically measured from a cross-covariance function of solar oscillation signals for various distances and time lags. When for a given distance the time lag corresponds to the propagation time of acoustic waves for this distance, a wavepacket-like signal appears in the cross-covariance function. The cross-covariance plotted as a function of the distance and the time lag displays a set of ridges formed by the wave-packet signals, representing an analog of a solar "seismogram". Since the solar oscillations are stochastic it is necessary to use the oscillation signals at least 2–8 hours long and also average them over some surface (typically, circular) areas in order to obtain a sufficient signal-to-noise ratio. Then, the travel times are determined by fitting a wavelet to this function (e.g. Kosovichev and Duvall 1997), or by measuring displacement of the ridges (Gizon and Birch 2002). Two general observing schemes, so-called 'surface-focusing' and 'deep-focusing', have been used in the travel-time measurements. In the surface-focusing scheme the travel times are measured for acoustic waves traveling between a central point and surrounding annuli. In the deep-focusing scheme, the travel times are measured for the acoustic wave packets traveling between the opposite parts of the annuli, the ray paths of which cross each other in a point located below the surface.

The relationship between the observed travel-time variations and the internal properties of the Sun is given by so-called sensitivity kernels (illustrated in Fig. 6 for both surface- and deep focusing) through integral equations. These integral equations are solved by standard mathematical inversion techniques such as Least Square QR Decomposition (LSQR) and Multi-Channel Deconvolution (MCD) (Kosovichev 1996; Jensen et al. 2001; Couvidat et al. 2006). The sensitivity functions are calculated using a ray theory or more complicated wave perturbation theories, e.g., the Born approximation, which takes into account the finite wavelength effects. These theories can also take into account stochastic properties of acoustic sources distributed over the solar surface (Gizon and Birch 2002; Birch et al. 2004).

The vertical structure of the computational grid and a sample of acoustic ray paths, used in this paper, are illustrated in Fig. 7. The travel times are measured for waves traveling between a central location and surrounding annuli with different radial distances from the

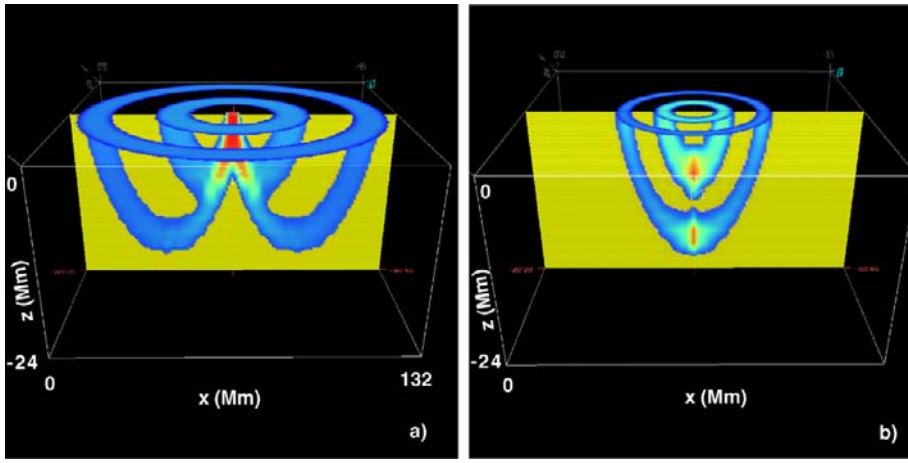


Fig. 6 Illustration of time–distance sensitivity kernels calculated in the ray-path approximation: (**a**) surface-focusing scheme; (**b**) deep-focusing scheme

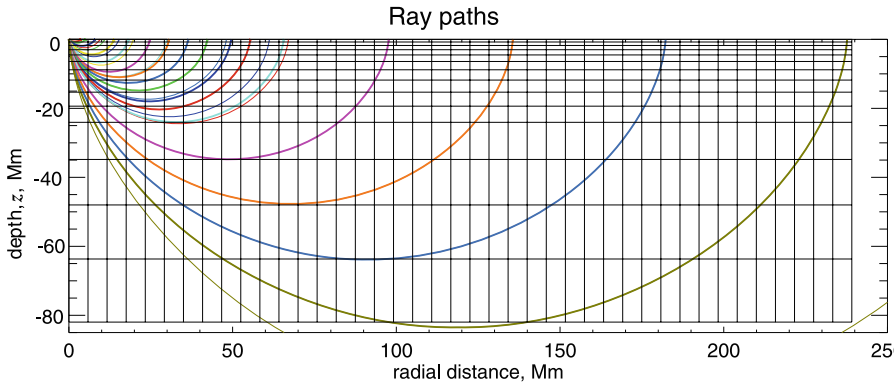


Fig. 7 A vertical cut through the 3D data inversion grid and a sample of acoustic ray paths

central point. The width of the annuli is larger for larger distances in order to improve the signal-to-noise ratio. A set of 17 annuli covering the distance range from 0.54 to 24.06 heliographic degrees (from 6.5 to 292 Mm) was used. The acoustic waves traveling to these distances sample the Sun's interior up to the depth of 95 Mm. The central locations of the time–distance measurements are chosen on a uniform 256×256 grid with the grid step of 2.9 Mm. A total of 1.1×10^6 travel time measurements are made to obtain each wave-speed image of the interior. For the flow velocity, the number of the measurements is three times larger, because in this case in addition to the travel times for a whole annulus it is necessary to measure also the travel times for waves traveling North–South and East–West. This is done by dividing each annulus into four sectors. A part of the inversion grid and the ray paths of the acoustic waves are illustrated in Fig. 7). The horizontal step in this figure is twice as large as the step of the travel-time measurements. The vertical grid, 82 Mm deep, is non-uniform with the step size increasing with depth, from 0.7 Mm at the top to 17 Mm at the bottom. The inversion procedure for all three components of flow velocity is described by Kosovichev and Duvall (1997).

3.2 Tomographic Imaging of Wave-Speed Perturbations

Figure 8 shows the results for the emerging active region, NOAA 8131, of January, 1998, obtained by Kosovichev et al. (2000). This was a high-latitude region of the new solar cycle which began in 1997. The distribution of the wave speed variations in a vertical cross-section

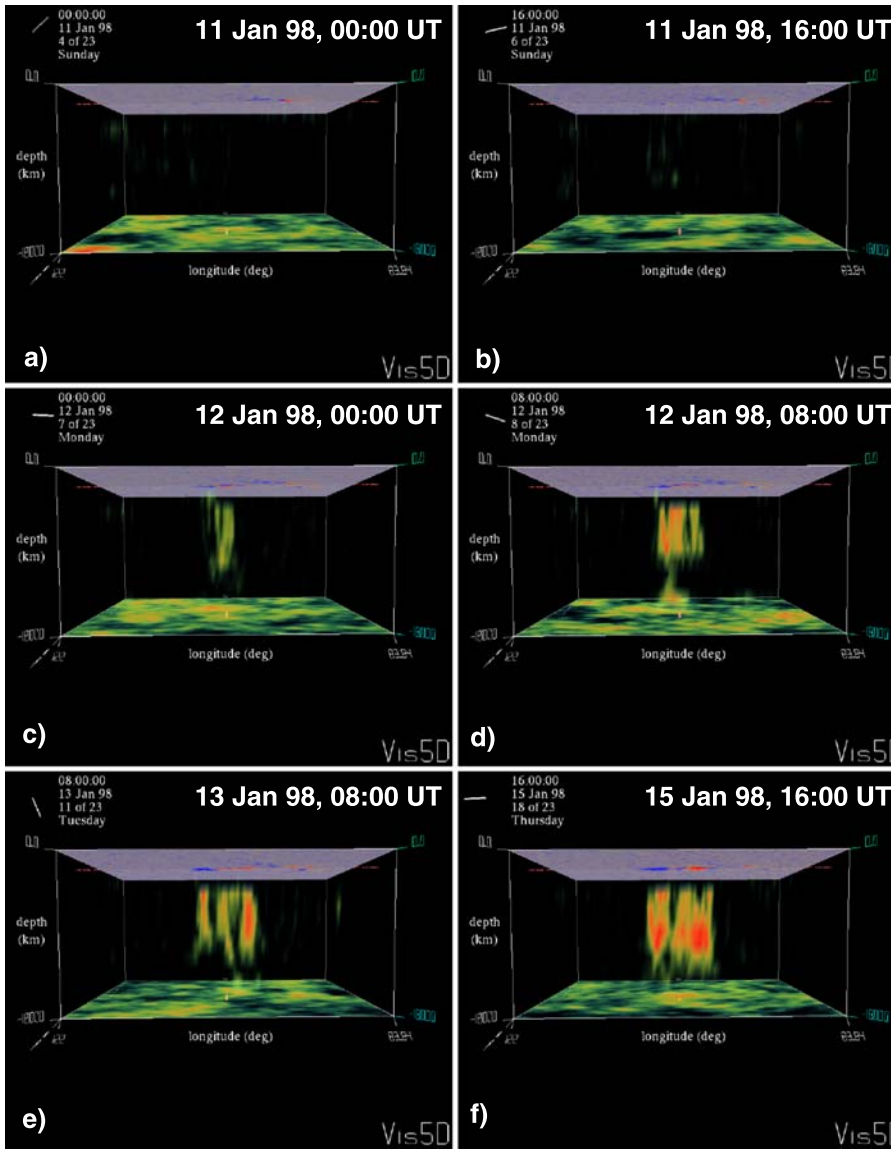


Fig. 8 The wave-speed perturbation in the emerging active region, NOAA 8131. The horizontal size of the box is approximately 38 degrees (460 Mm), the vertical size is 18 Mm. The *panels on the top* are MDI magnetograms showing the surface magnetic field of positive (*red*) and negative (*blue*) polarities. The perturbations of the wave speed are approximately in the range from -1 to $+1$ km/s. The positive variations are shown in *red*, and the negative ones in *blue*

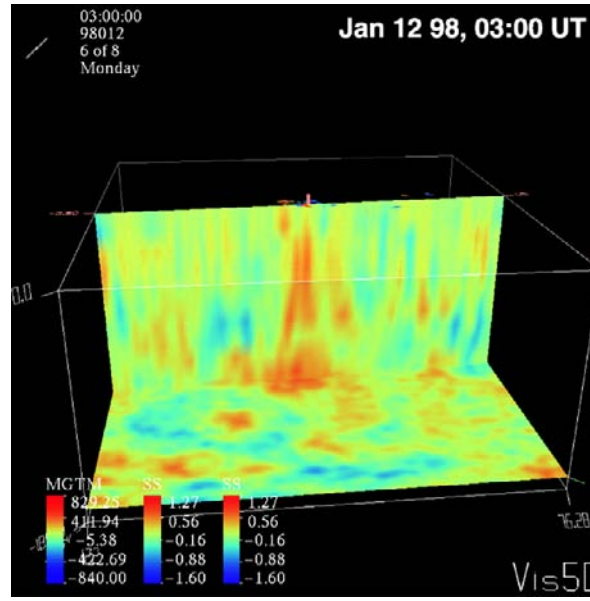


Fig. 9 Image of the magneto-acoustic wave speed in an emerging active region (AR 8131) in the solar convection zone obtained from the SOHO Michelson Doppler Imager (MDI) data on January 12, 1998, from 02:00 to 04:00 UT, using time–distance helioseismology. The horizontal size of the box is approximately 560 Mm, and the depth is 18 Mm. The (mostly) *transparent panel on the top* is an MDI magnetogram showing the surface magnetic field of positive (*red*) and negative (*blue*) polarities stronger than 200 gauss. The *vertical and bottom panels* show perturbations of the wave speed which are approximately in the range from -1.6 to $+1.3$ km/s. The positive variations are shown in *red*, and the negative ones in *blue*. A large active region formed at this location within a day after these observations

in the region of the emerging flux and in a horizontal plane at a depth of 18 Mm are shown for six 8-hour consecutive intervals. The perturbations of the magnetosonic speed shown in this figure are associated with the magnetic field and temperature variations in the emerging magnetic ropes and in the surrounding plasma. The panel (a) shows no significant variations in the region of emergence, which is at the middle of the vertical plane. The MDI magnetogram shown at the top indicates only very weak magnetic field above this region. The panel (b) shows a slight positive perturbation associated with the emerging region. During the next 8 hours (panel c) the perturbation becomes stronger and occupies the whole range of depths and continue to increase. These results show that the emerging flux propagates very quickly through the upper 18 Mm of the convection zone. We have also analyzed the data for 2-hour intervals at the start of emergence from 2:00 UT to 4:00 UT, January 12, 1998, (Fig. 9) and concluded that the emerging flux propagated through the characteristic depth of 10 Mm in approximately 2 hours. This gives an estimate of the speed of emergence $\approx$1.3 km/s. This speed is similar to the speed predicted by the theories of emerging flux ropes. The typical amplitude of the wave-speed variation in the emerging active region is about 0.5 km/s. After the emergence we observed a gradual increase of the perturbation in the subsurface layers, and the formation of sunspots (Fig. 8d–f). The observed development of the active region seems to suggest that the sunspots are formed as a result of the concentration of magnetic flux close to the surface.

Several large active regions emerged on the Sun in October–November 2003. This period represents one of the most significant impulses of solar activity in solar cycle 23. During this period the MDI instrument on SOHO was in the full-disk mode (“Dynamics program”), taking full-disk Dopplergrams and line-of-sight magnetograms every minute. Thus, it was able to capture the emergence of active regions NOAA 10488, and also obtain data for two other large active regions, 10484 and 10486. The active region, 10488, emerged in the Northern hemisphere (at 291 deg Carrington longitude and 8 deg latitude) approximately at

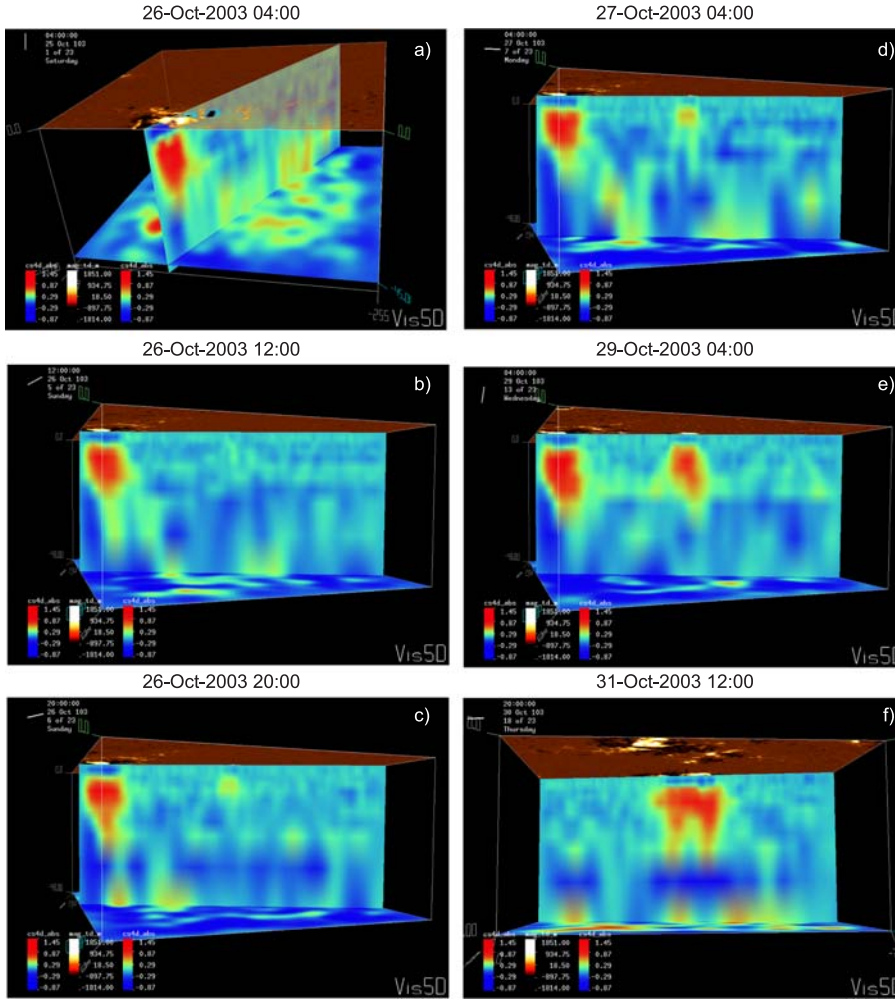


Fig. 10 Subsurface magnetosonic wave-speed structures of the large complex of activity of October–November 2003, consisting of active regions NOAA 10486 (in the *left-hand part* of the images), and 10488 (emerging active region in the middle). *Red color* shows positive wave-speed variations relative to the quiet Sun; the *blue color* shows the negative variations, which are concentrated near the surface. The *upper semi-transparent panels* show the corresponding MDI magnetograms; the *lower panel* is a horizontal cut 48 Mm deep. The horizontal size is about 540 Mm. The vertical cut goes through both active regions, approximately in the North–South direction crossing the equator, except the image in the *right bottom panel*, (**f**), where it goes only through AR 10488 in the East–West direction

the same longitude as AR 10486, which emerged earlier and had a very complex magnetic configuration resulting in several strong flares. It is possible that these two active regions had a common nest in the interior (Zhou et al. 2007).

Using the method of time–distance helioseismology, we have obtained wave-speed and flow velocity maps for 8 days, 25–31 October, 2003 (Kosovichev and Duvall 2008). The maps are obtained using 8-hour time series with a 2-hour shift. Total 96 wave-speed and flow maps were obtained. Figure 10 shows a sample of the wave-speed images. The vertical

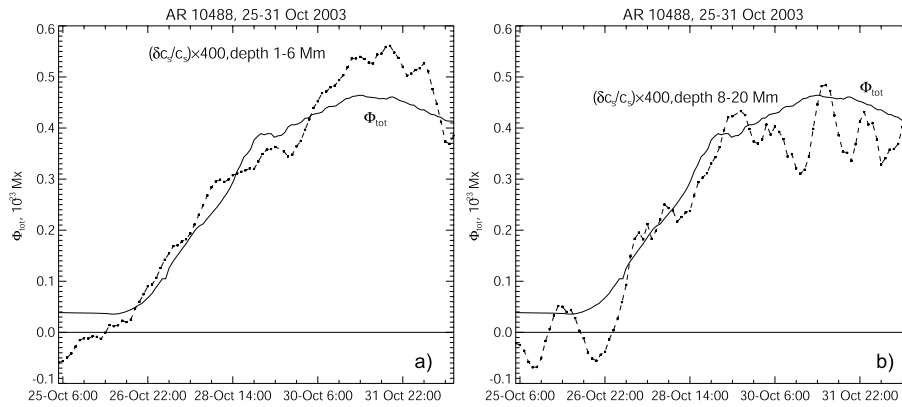


Fig. 11 Evolution of the total unsigned photospheric magnetic flux (*solid curve*) and the mean relative wave-speed variation (*dotted curve with stars*) at the depth of 1–6 Mm (*left*) and 8–20 Mm (*right*) in the region of the flux emergence of AR 10488

cut through these images (except the image of 31 Oct 2003, 12:00) is made through both AR 10486 and 10488 in approximately the North–South direction, and for the image of 31 Oct 2003, 12:00 (Fig. 10f) it is made in the East–West direction. The depth of the image box is 48 Mm, and the horizontal size is about 540 Mm.

The results show that the first wave-speed signal below the surface appeared in the image obtained on 26 October, 2003, for the time interval centered at 12:00 UT (Fig. 10b). This is slightly ahead of the growth of the total magnetic field flux, which started to grow at about 20:00 UT (Fig. 11, left, solid curve; however, the first magnetic field signal appeared approximately at the same time). During the next 8 hours, between 12:00 and 20:00, the wave-speed perturbation rapidly grows, and is most visible in the subsurface layers, about 10 Mm deep. In the deeper interior, we do not detect a clear signal above the noise level at this time. This may be because the relative perturbation in these layers is too weak, and also may indicate that the formation of magnetic flux concentrations starts in the subsurface layers. During the next 8 hours the signal extends into the deeper layers and continues to grow (Fig. 10d). The typical two-layer structure with lower wave speed in the top 4–5 Mm, and higher wave speed in the deeper layers is formed (Kosovichev et al. 2000; Jensen et al. 2001; Couvidat et al. 2006). During the following 5 days of the MDI observations, the wave-speed perturbation below the active region becomes larger and stronger, and in the East–West direction it forms a loop-like structure (Fig. 10f). This structure can be traced to the depth of about 30 Mm, and then it is lost in noise.

In Fig. 11 we compare the evolution of the total (unsigned) magnetic flux of the active regions and the mean wave-speed signal in the two depth intervals, 1–6 Mm and 8–20 Mm. In both cases, the signals correspond well to the evolution of the surface magnetic flux. There is possibly an indication of a slight lead of the wave-speed signal at 1–6 Mm, but there is no significant time lag. At greater depths the noise level is higher, and it is even more difficult to see the difference in the time evolution relative to the surface magnetic flux.

These results show that the magnetic flux emerges very rapidly from the interior, and that there is no significant (on the scale of few hours) time difference between the evolution of the wave-speed variations associated with the emerging active region and the photospheric magnetic flux. There are indications that the process of the magnetic field concentration,

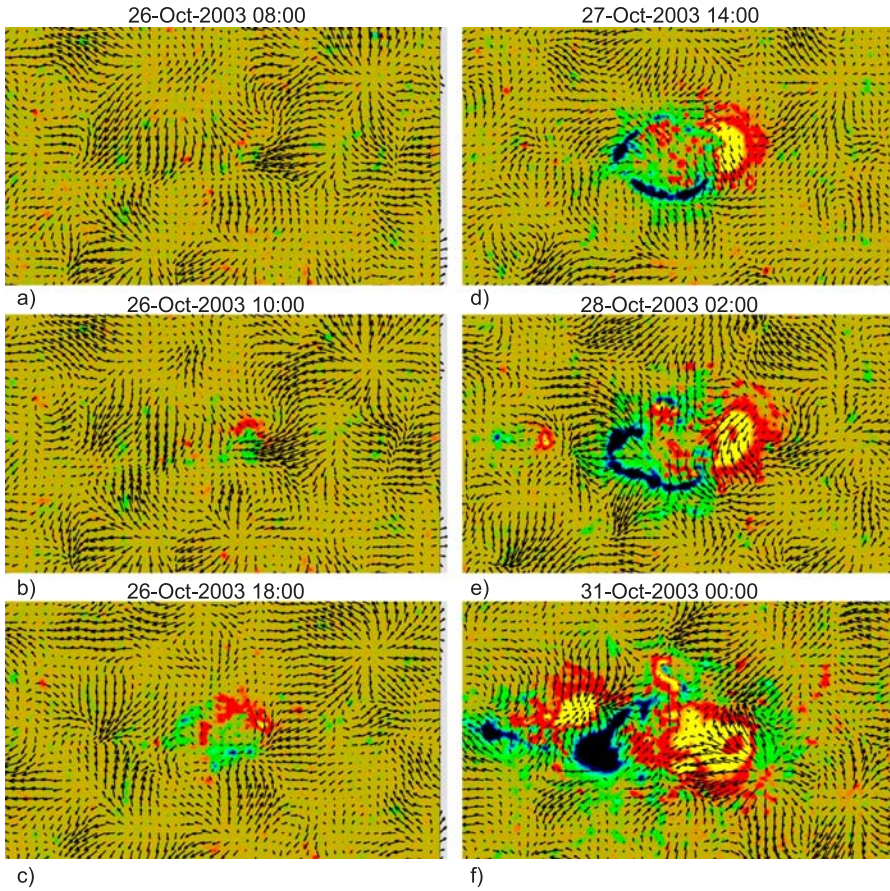


Fig. 12 Evolution of subsurface flows at the depth of 2 Mm below the photosphere during the emergence and growth of AR 10488, on 26–31 October, 2003. The flow maps are obtained by the time–distance technique using 8-hour time series of full-disk Doppler images from SOHO/MDI. The maximum horizontal velocity is approximately 1 km/s. The background image is the corresponding photospheric magnetogram (*red* and *bright yellow areas* show regions of positive polarity, and *blue* shows negative polarity of the line-of-sight magnetic field)

which forms the active region, first occurs in the subsurface layers, and that then the active region grows because of subsequent flux emergence in this area.

3.3 Subsurface Flows

The helioseismology measurements of subsurface flows are obtained from the reciprocal travel times, and generally, are less affected by various kind of uncertainties. They may provide better indicators of the development of active region structures inside the Sun. Figure 12 shows six flow maps at the depth of about 2 Mm for various stages of evolution of the active region, NOAA 10488, before the emergence, during the initial emergence, and during the developed state. The background color maps show the corresponding magnetograms.

Prior to the emergence, the maps do not show any specific flow pattern that would indicate development of a large magnetic structure below the surface, except, perhaps, a small

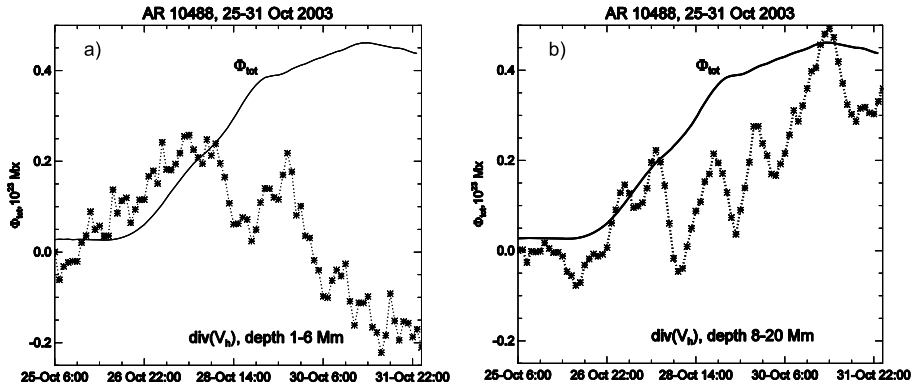


Fig. 13 The evolution of the total unsigned photospheric magnetic flux (*solid curve*) and the mean divergence of the horizontal flow velocity (*dotted curve with stars*) at the depth of 1–6 Mm (*left*) and 8–20 Mm (*right*) in the region of the flux emergence of AR 10488. The units of divV_h are $3 \cdot 10^{-7}$ s^{-1}

shearing flow feature, which appeared near the first magnetic field signal in the center of Fig. 12a–b. During the next 8 hours (Fig. 12c), this feature disappears, and a ring-like magnetic field structure is formed. Within this structure the flows are clearly suppressed, and they remain suppressed during further evolution. Also, at the same type a diverging flow pattern starts developing at the boundaries of the magnetic structures. This pattern is consistent with the expectation that the emerging magnetic structure pulls plasma outside. The divergent flow field becomes stronger as the active region grows (Fig. 12e), but later, it is replaced by a converging flow pattern around the sunspots (Fig. 12f), which was previously observed beneath sunspots (Zhao et al. 2001).

The strength of the divergent flows is obviously related to the development of active regions, and, perhaps, may be even used for predicting their future evolution. The time evolution of the mean horizontal divergence in the two depth intervals and the photospheric flux is shown in Fig. 13. It is quite clear that the divergence at the depth 1–6 Mm started to grow before the magnetic flux, reached maximum in the middle of the flux growth phase, and then was replaced by converging flows. At greater depths, 8–20 Mm (Fig. 13 right), the horizontal flow behavior is not very clear, probably because of higher noise, or because the flow pattern is not as well organized as in the subsurface (6 Mm deep) layer. Perhaps the most significant feature at this depth is the formation of a divergent flow pattern approximately at the time of the formation of convergent flows in the upper subphotospheric layer.

One would expect that during the emergence the plasma is not only pushed outside the magnetic field area but also upward, particularly, in the upper layers. Figure 14a shows the evolution of the mean vertical flow below the active region at the depth 1–6 Mm. Indeed, upflows dominate at the very beginning of the magnetic flux emergence. However, the signal fluctuates, probably reflecting a complicated structure of the vertical flows. After the emergence phase the vertical flow pattern is dominated by downflows, which are organized around the sunspots.

It seems that the horizontal divergence of subsurface flows is the most sensitive characteristic of the emerging magnetic flux. The divergent flows appear before the initial flux emergence, and continue to evolve in correlation with the magnetic flux. Figure 14b shows a comparison of the mean horizontal divergence and the total magnetic rate. Evidently, there were two or three peaks of the magnetic emergence rate. The flow divergence shows two

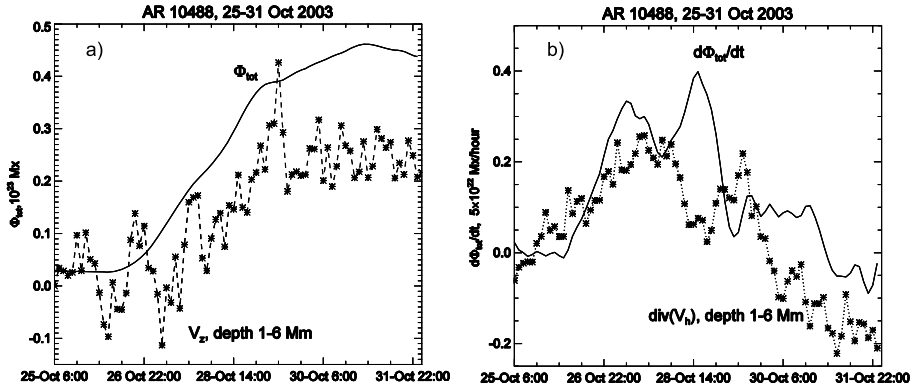


Fig. 14 (**a**) The evolution of the total unsigned photospheric magnetic flux (*solid curve*) and the mean vertical velocity in km/s (*dotted curve with stars*) at the depth of 1–6 Mm in the region of the flux emergence of AR 10488. The negative velocity corresponds to upflows, and the positive velocity corresponds to downflows. (**b**) The corresponding changes of the total emerging flux rate and the mean divergence of the horizontal flow components

peaks, which are shifted relative to the flux rate. It is unclear whether these peaks precede or follow the magnetic flux emergence events. Obviously, this relationship requires further investigation. Similar flow patterns have been studied by Komm et al. (2008) using data from the GONG network and the ring-diagram method of local helioseismology. The initial results are quite encouraging and show the potential of the helioseismic diagnostics.

4 Comparison with Theoretical Models

The time–distance helioseismology measurements provide new information about the structure and dynamics of emerging active regions. In this paper, we presented in detail the results for a large active region, NOAA 10488, which emerged in October 2003. The results show that the formation of the active region takes about 5 days. During this period the total magnetic flux and the corresponding subsurface wave-speed perturbation grow mostly monotonically. However, the magnetic flux rate reveals two or three peaks of intensive flux emergence; each is about one day long. It appears that the active region is formed by multiple magnetic flux emergence events. The initial magnetic flux emerged very rapidly without any significant perturbation of the Sun's thermodynamic structure or flow field in the place of emergence. There seems to be a short lead in the growth of the subsurface wave speed perturbation relative to the mean magnetic flux, but this relationship is rather uncertain. A localized shearing flow seems to be formed few hours before the initial flux emergence, and then disappears soon after the emergence. The active region has a elliptical shape, with the magnetic field concentrated at the boundaries. The plasma flows are suppressed inside this structure. In the outer region, the plasma flows are dominated by divergent flow, driven by the expanding magnetic structures. The flow divergence at the depth of 1–6 Mm shows a few-hour lead relative to the magnetic flux. It grows during the emergence phase until approximately the mid-point of the flux growth curve. After this, the divergence is sharply reduced and then is replaced by predominantly converging flows around sunspots. Approximately, at the same time a divergent flow pattern is formed in the deeper interior (depth 8–20 Mm). The vertical flow pattern is quite complicated. The results do not show strong

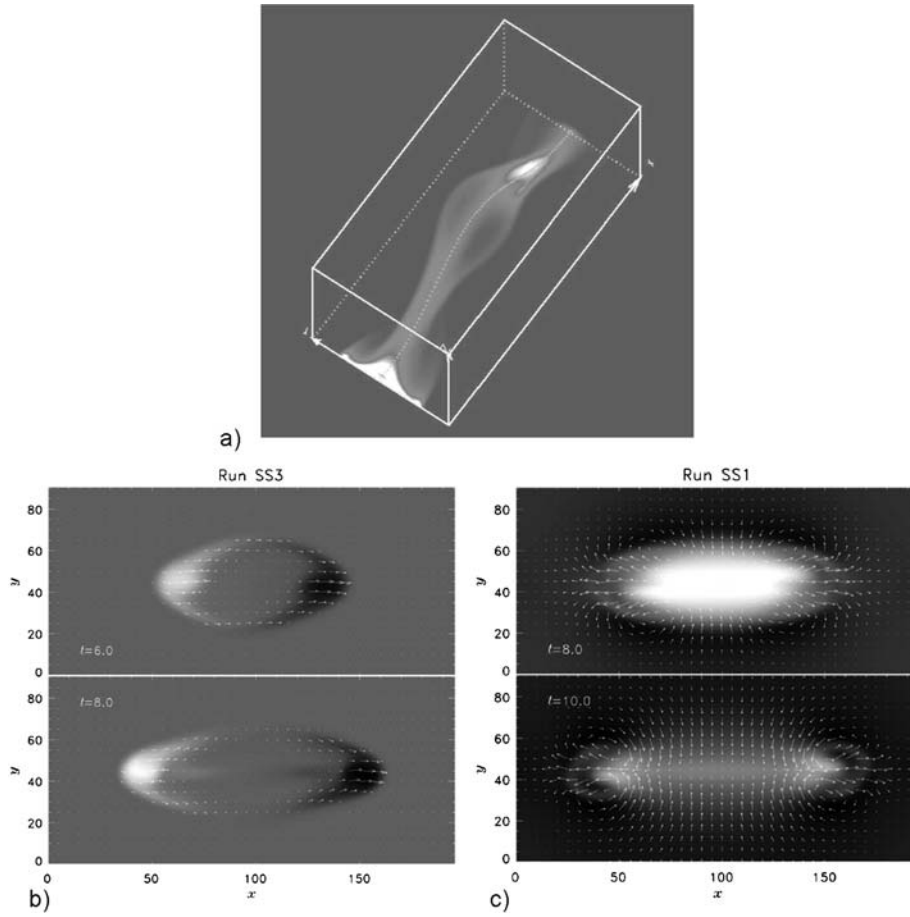


Fig. 15 A theoretical MHD model of emerging magnetic flux tube (Abbett et al. 2000): (**a**) Volume rendering of the magnetic field strength. (**b**) Vector magnetogram images. The gray-scale background represents the vertical component of the magnetic field (positive values are indicated by the light regions), and the *arrows* represent the transverse components. (**c**) Velocity field for the same slice. The gray-scale background represents the vertical component of the velocity (light regions indicate upflows), and the *arrows* represent the transverse components of the velocity field

upflows prior to the emergence, as one might expect. In general, the vertical flow pattern is highly intermittent. There is an evidence of predominant upflows during the initial stage of emergence, but after this the mean flow beneath the active region is directed downwards. It seems that there is an interesting correlation with some time lag between the flow divergence and the flux emergence rate. However, at this stage it is unclear whether the changes in the flow divergence precede or follow the flux rate.

It is interesting that the magnetic structure of this active region, in particular, its elliptical shape is very similar to a model of emerging magnetic flux tube of Abbett et al. (2000) (Fig. 15). The model also predicts a divergent flow pattern similar to the observed one. However, the strong upflow at the beginning of emergence (Fig. 15c, top panel) is not detected in our observations. Thus, the process of emergence and formation of active regions requires further observational and theoretical studies.

5 Discussion and Future Perspectives

The recent observations and modeling reveal some interesting features of the properties of emerging magnetic flux and associated dynamics on the solar surface and in the upper convection zone. In particular, the new statistical study of the variations of the tilt angle of bipolar magnetic regions during the flux emergence questions the current paradigm that the magnetic flux emerging on the solar surface represents large-scale magnetic flux ropes (Ω-loops) rising from the bottom of the convection zone. The flux rope models predict that the tilt angle is a result of the Coriolis effect acting on a plasma flow inside the flux tube, and thus the tilt should depend on latitude, the amount of magnetic flux and relax after the emergence when the Coriolis force vanishes. The observations indeed show the predicted latitudinal dependence (the Joy's law) and indicate that the tilt is formed below the surface. However, there is no evidence of the dependence on the amount of magnetic flux and no evidence for the relaxation of the tilt angle towards the East–West direction. Contrary, the tilt angle tends to relax to the Joy's law value. Perhaps, the Joy's law reflects not the dynamics of the rising flux tubes but the orientation of the toroidal magnetic field lines below the surface as suggested by Babcock (1961).

The observations of the surface flows from SOHO/MDI prior and during the emergence of a large active regions, AR 10488, in October 2003, show strong localized vertical flows just prior the flux emergence and during the initial stage. It is curious that the direction of the flows, namely, an upflow in the area of the leading polarity and a downflow in the following polarity, is consistent with the predictions of the rising flux rope theories. However, observations of some other active regions do not show this (Pevtsov and Lamb 2006). Also, the data do not show large-scale flow patterns on the surface, which would indicate emergence of a large flux-rope structure.

The local helioseismology results obtained by both, the time–distance and ring-diagram techniques, show large-scale outflows beneath the surface during most of the emergence phase, and also formation of converging flows around the magnetic structure of sunspots. However, the structure of the vertical flows remains unclear. There is an indication of upflows mixed with downflows at the beginning of emergence, but then the downflows dominate. In the case of AR 10488, there were two or three major flux emergence events. The photospheric magnetic flux rate and subsurface flow divergence show two or three peaks, which are not in phase, but it is unclear if the flux rate precedes the variation of the flow divergence or follows it.

From the observations it is obvious that the multiple flux emergence events over several days plays important role in the formations and maintaining the magnetic structure of the large active region. This reminds the idea of a common 'nest' in the deep interior (Castenmiller et al. 1986). However, such nests have not been found in the helioseismic images of the subphotospheric magnetosonic wave speed variations, which are currently obtained up to the depth of 40–50 Mm. The wave speed images reveal that the emerging magnetic flux structures travel very fast in the upper convection zone, with a speed of at least 1 km/s. This makes very difficult the detection of these structures before the magnetic field becomes visible on the surface. Thus, it is difficult to use the helioseismology measurements for advanced predictions of emerging active regions. However, it should be possible to use the measurements of both, the wave speed variations and flow velocities, for predicting the growth and decay of active regions and, perhaps, the complexity of their magnetic structure. This task will require a substantial statistical analysis of emerging active regions by methods of local helioseismology.

Thus, despite the significant new information from helioseismology and magnetography the main questions formulated in Introduction about the origin and physical properties of the emerging magnetic remain unanswered. The recent results from the SOHO spacecraft and GONG network show that for further investigations it is necessary to improve the local helioseismology techniques, extending their coverage into the deep convection zone, carry out statistical studies using uninterrupted solar oscillation data (such as will be available from the Solar Dynamics Observatory mission), and also develop realistic MHD numerical simulations for understanding the physics of magnetic structures in the turbulent convection zone and for supporting the helioseismology observations.

Acknowledgements I wish to thank all the organizers and participants of this workshop, and, in particular, Dr Andre Balogh. This work was supported by the International Space Science Institute (Bern).

References

W.P. Abbett, G.H. Fisher, Y. Fan, Astrophys. J. **540**, 548 (2000)
H.W. Babcock, Astrophys. J. **133**, 572 (1961)
E.E. Benevolenskaya, J.T. Hoeksema, A.G. Kosovichev, P.H. Scherrer, Astrophys. J. Lett. **517**, L163 (1999)
A.C. Birch, A.G. Kosovichev, T.L. Duvall Jr., Astrophys. J. **608**, 580 (2004)
A. Brandenburg, Astrophys. J. **625**, 539 (2005)
P. Caligari, F. Moreno-Insertis, M. Schussler, Astrophys. J. **441**, 886 (1995)
M.J.M. Castenmiller, C. Zwaan, E.B.J. van der Zalm, Sol. Phys. **105**, 237 (1986)
S. Couvidat, A.C. Birch, A.G. Kosovichev, Astrophys. J. **640**, 516 (2006)
S. D'Silva, in *The Solar Cycle*, ed. by K.L. Harvey. Astronomical Society of the Pacific Conference Series, vol. 27 (1992), p. 168
S. D'Silva, R.F. Howard, Sol. Phys. **151**, 213 (1994)
Y. Fan, G.H. Fisher, A.N. McClymont, Astrophys. J. **436**, 907 (1994)
G.H. Fisher, Y. Fan, R.F. Howard, Astrophys. J. **438**, 463 (1995)
L. Gizon, A.C. Birch, Astrophys. J. **571**, 966 (2002)
V.M. Grigor'ev, L.V. Ermakova, A.I. Khlystova, Astron. Lett. **33**, 766 (2007)
G.E. Hale, F. Ellerman, S.B. Nicholson, A.H. Joy, Astrophys. J. **49**, 153 (1919)
J.W. Harvey, F. Hill, R. Hubbard, J.R. Kennedy, J.W. Leibacher, J.A. Pintar, P.A. Gilman, R.W. Noyes, A.M. Title, J. Toomre, R.K. Ulrich, A. Bhatnagar, J.A. Kennewell, W. Marquette, J. Patrón, O. Saá, E. Yasukawa, Science **272**, 1284 (1996)
R.F. Howard, Sol. Phys. **167**, 95 (1996)
R.F. Howard, J. Astrophys. Astron. **21**, 119 (2000)
R. Howe, J. Christensen-Dalsgaard, F. Hill, R. Komm, J. Schou, M.J. Thompson, J. Toomre, Adv. Space Res. **40**, 915 (2007)
J.M. Jensen, T.L. Duvall Jr., B.H. Jacobsen, J. Christensen-Dalsgaard, Astrophys. J. Lett. **553**, L193 (2001)
R. Komm, S. Morita, R. Howe, F. Hill, Astrophys. J. **672**, 1254 (2008)
A.G. Kosovichev, Astrophys. J. Lett. **469**, L61 (1996)
A.G. Kosovichev, T.L. Duvall Jr., in *SCORe'96: Solar Convection and Oscillations and their Relationship*, ed. by F.P. Pijpers, J. Christensen-Dalsgaard, C.S. Rosenthal. Astrophysics and Space Science Library, vol. 225 (1997), p. 241
A.G. Kosovichev, T.L. Duvall Jr., in *Subsurface and Atmospheric Influences on Solar Activity*, ed. by R. Howe, R.W. Komm, K.S. Balasubramaniam, G.J.D. Petrie. Astronomical Society of the Pacific Conference Series, vol. 383 (2008), p. 59
A.G. Kosovichev, J.O. Stenflo, Astrophys. J. Lett. **688**, L115 (2008)
A.G. Kosovichev, T.L. Duvall Jr., P.H. Scherrer, Sol. Phys. **192**, 159 (2000)
E.N. Parker, Astrophys. J. **408**, 707 (1993)
E.N. Parker, Astrophys. J. **433**, 867 (1994)
A. Pevtsov, J.B. Lamb, in *Solar MHD Theory and Observations: A High Spatial Resolution Perspective*, ed. by J. Leibacher, R.F. Stein, H. Uitenbroek. Astronomical Society of the Pacific Conference Series, vol. 354 (2006), p. 249
P.H. Scherrer, R.S. Bogart, R.I. Bush, J.T. Hoeksema, A.G. Kosovichev, J. Schou, W. Rosenberg, L. Springer, T.D. Tarbell, A. Title, C.J. Wolfson, I. Zayer, MDI Engineering Team, Sol. Phys. **162**, 129 (1995)

H.U. Schmidt, in *Structure and Development of Solar Active Regions*, ed. by K.O. Kiepenheuer. Proc. IAU Symposium No. 35 (1968), p. 95
M. Schüssler, Astronomische Nachrichten **326**, 194 (2005)
L. Tian, Y. Liu, Astron. Astrophys. **407**, L13 (2003)
Y.-M. Wang, N.R. Sheeley Jr., Astrophys. J. **375**, 761 (1991)
J. Zhao, A.G. Kosovichev, T.L. Duvall Jr., Astrophys. J. **557**, 384 (2001)
G. Zhou, J. Wang, Y. Wang, Y. Zhang, Sol. Phys. **244**, 13 (2007)

The Topology and Behavior of Magnetic Fields Emerging at the Solar Photosphere

B.W. Lites

Originally published in the journal Space Science Reviews, Volume 144, Nos 1–4, 197–212.
DOI: 10.1007/s11214-008-9437-x

Abstract The nature of flux emerging through the surface layers of the Sun is examined in the light of new high-resolution magnetic field observations from the Hinode space mission. The combination of vector magnetic field data and visible-light imaging from Hinode support the hypothesis that active region filaments are created as a result of an emerging, twisted flux system. The observations do not present strong evidence for an alternate hypothesis: that the filaments form as a result of localized shear flows at the photospheric level. Examination of the vector magnetic field at very small scales in emerging flux regions suggests that reconnection at the photospheric level and below, followed by submergence of flux, is a likely and essential part of the flux emergence process. The reconnection and flux submergence are driven by granular convection.

Keywords Sun · Magnetic fields · Flux emergence

1 Introduction

The initial movies of the entire solar disk seen in X-rays by the Yohkoh satellite revealed that the coronal loops in and around solar active regions undergo a continual expansion (Uchida et al. 1992). This overall behavior of the solar corona reflects the constant emergence of active region magnetic fields at the solar surface (the photosphere). The phenomenon of an expanding active region corona is but one aspect of the dynamics of the outer solar atmosphere influenced by the emergence of magnetic fields from the solar interior. In the past two decades it has also been recognized that many of the emerging magnetic fields carry substantial twist, or helicity (Rust and Kumar 1994; Pevtsov et al. 1995); that the helicity must be a nearly conserved quantity on large scales due to the high conductivity of the solar plasma (Berger 1984); that moderate degrees of twist allow magnetic "flux ropes" to rise

The National Center for Atmospheric Research is sponsored by the National Science Foundation.

B.W. Lites (✉)
High Altitude Observatory, ESSL, National Center for Atmospheric Research, P.O. Box 3000, Boulder, CO 80307, USA
e-mail: lites@ucar.edu

M.J. Thompson et al. (eds.), *The Origin and Dynamics of Solar Magnetism*.
DOI: 10.1007/978-1-4419-0239-9_12

through the convection zone on their way to the solar surface without being fragmented by convective instabilities (Abbett et al. 2000); and that highly twisted (sheared) magnetic fields may develop instabilities in the stratified solar atmosphere that lead to the sudden release of magnetic energy in solar flares and coronal mass ejections (Rust and Kumar 1996; Low 2001).

To some degree, the attributes and evolution of magnetic fields at the solar surface reflect the state of the fields within the solar interior. Furthermore, the fields emerging through the solar photosphere drive all the dynamic and energetic phenomena of the upper solar atmosphere that are responsible for variability of the Sun and the near-Earth space environment. To a great extent, our understanding of solar magnetic phenomena depends on how well we can characterize in quantitative detail the emergence of magnetic fields at the solar photosphere.

As a result of modern solar polarimetry, we are fortunate to be able to infer the magnetic field vector at the level of the solar photosphere with significant precision and certainty. Launched in September 2006, the Hinode satellite (Kosugi et al. 2007) gives us our clearest view yet of photospheric magnetic fields. Much of this section will present new results from Hinode that allow us to characterize the topology of the emerging flux and its evolution. These new observations have not only helped to solidify some notions based on earlier, lower resolution and sporadic ground-based measurements, they have also revealed to us new aspects of the emergence process that we are only beginning to understand.

In this section we focus on two aspects of emerging flux upon which recent observations from Hinode have shed new light. First we consider the appearance of twisted magnetic flux in active regions that is associated with the formation of active region filaments in the chromosphere above. The evolution of such structures suggests the emergence of twisted magnetic flux from below, not the consequence of shearing motions and reconnection, underlies the filament phenomenon. Second, we consider the behavior and characteristics of the small scale magnetic field in regions where flux is actively emerging. These observations suggest that reconnection followed by flux submergence might be an essential aspect of the emergence process, in that the submerging reconnected flux allows the field to unload the substantial mass of the solar plasma at the photospheric level.

2 Emergence of Twisted Magnetic Flux and the Formation of Filament Channels

Solar prominences[1] are structures occasionally seen above the solar limb when viewed in opaque spectral lines forming in the solar chromosphere. Prominences are known to be relatively cool (10^4 K) plasma embedded in the surrounding hot (10^6 K) plasma of the corona. Magnetic fields both support the dense, cool material of the prominence above the solar surface and isolate it thermally from the hot corona. Because the plasma in a prominence is so much cooler than its surroundings, the plasma contained therein will have a scale height similar to that of the solar chromosphere: a few hundred km, yet the vertical extent of prominences as seen above the limb are frequently 10,000 km or more. In order that the prominence be visible for long periods, the visible structure must represent an ensemble of a multitude of individual magnetic field lines, stacked one upon another, and each field line supporting relatively cool chromospheric material in the lowest few hundred km of local dips in the field lines. The questions then arise: how does such a field geometry form, and how does the dense, cool mass come to reside on those field lines?

[1] When seen on the solar the disk, prominences are known as filaments.

An increasing body of evidence indicates that the geometry of the magnetic field that supports prominences is a "flux rope"—a magnetic field that has some twist. Supporting this notion are the empirical facts that: (1) when it has been possible to directly measure the orientation of the magnetic field in a prominence, their fields are most commonly observed to have "inverse" magnetic configuration (Athay et al. 1983; Leroy 1989): the component of the magnetic field parallel to the solar surface in the prominence is directed from regions of negative polarity toward regions of positive polarity, (2) magnetograms of active regions containing filaments invariably show them to occur over the polarity inversion line (PIL: the separator between locally upward- and locally downward-directed vertical components of the magnetic field), (3) in cases where the vector magnetic field under active region filaments has been observed with good precision and high angular resolution, the photospheric field also shows inverse configuration (Lites et al. 1995; Lites 2005; Okamoto et al. 2008), (4) measurements of the field in the vicinity of quiescent filaments also suggest inverse configuration (López Ariste et al. 2006) and (5) the topology of the magnetic field in the corona over active region filaments, as inferred from EUV or X-ray imaging, often shows a "sigmoid" shape characteristic of twisted magnetic flux (Gibson et al. 2002, 2006). The inverse configuration, coupled with the belief that the prominence material resides at local minima in the height of magnetic field lines, strongly suggests a helical field geometry of the field that is locally "concave upward" at the PIL. Because the field makes only a small angle with respect to the axis of the filament, we surmise that the field is not highly twisted; that is, it executes only one or two turns in the azimuthal direction traversing the length of the filament along the PIL. Simulations of flux ropes in the corona based on the magnetic field observed at the photosphere (Bobra et al. 2008) also depict a slightly twisted magnetic field configuration.

To date, there are only a few observational studies (cited above) that have attempted to characterize the vector magnetic field in the photosphere under filaments. Quantitative measures of the magnetic field vector at high angular resolution are needed to resolve the components of the field locally horizontal to the solar surface (i.e., not just perpendicular to the line-of-sight). Furthermore, also needed is the evolutionary history of the field when the conditions become favorable for the appearance of a filament in the corona above an active region. This history gives us important clues as to how that field structure comes to be. Active region filaments are particularly important as indicators of the process of flux rope formation for the following reasons: (1) the magnetic fields locally in the photosphere of active regions are much stronger than those associated with quiescent filaments in the quiet Sun, thus they more readily lend themselves to measurement and interpretation using the Zeeman effect; (2) for many growing active regions harboring filaments, the fields are measurable continuously over a large region, so that one is able to discern the evolution of the vector field under the filament in the context of its magnetic surroundings; (3) the flux rope associated with an active region filament may reside at relatively low heights above the photosphere so that the influence of the twisted rope structure is much more likely to be exerted on the photosphere, hence become measurable at photospheric heights; and (4) important dynamic phenomena such as flares and coronal mass ejections (CMEs) associated with twisted or sheared magnetic flux occur with higher frequency in active regions.

Two scenarios have been put forth to explain the formation of flux ropes in the corona. The first relegates the formation of the twist to the solar interior, forming either as part of the generation of the field by a dynamo at the bottom of the convection zone (and apparently some twist is needed to allow the flux to rise coherently to the surface Abbett et al. 2000), or imparted to the field as it rises buoyantly through the differentially-rotating convection zone. In this scenario—the "emerging twisted flux" scenario—the magnetic flux would first

emerge as bipolar fields having the normal orientation, forming arch filament structures in the corona as observed at sites of actively emerging flux. Then as the rope moves bodily into the atmosphere the field transitions to being aligned with the PIL, and finally forms the inverse configuration after the axis of the rope—where the field strength is largest—resides in the corona. The fields below the axis of the rope have the concave-upward geometry that can support the cool filament material.

The second scenario (the "flow system" scenario) involves the interaction of horizontal flows at or just below the photospheric level with pre-existing magnetic flux rooted in those flows. The motion of the magnetic foot points must be systematic along much of the length of the PIL, resulting from either a converging flow perhaps due to diffusion of fields (van Ballegooijen and Mackay 2007) or a systematic shear flow with respect to the PIL (Aulanier et al. 2002). The latter may arise, at least on larger scales, as a result of differential rotation. In these scenarios, reconnection of the field occurs in the atmosphere above the PIL leading to a twisted field geometry that can support a prominence.

Both scenarios have their advantages and detractors. The emerging twisted flux does not require the specific conditions that must be assumed of the flow system at the photosphere, but it must somehow allow for the unloading of the enormous mass of the photosphere and sub-photosphere that must be trapped on the concave-upward portions of the field lines.

2.1 Hinode Observations of Filament Channel Formation/Destruction

The Spectro-Polarimeter (SP) for the Solar Optical Telescope (SOT, Tsuneta et al. 2008; Suematsu et al. 2008; Ichimoto et al. 2008; Shimizu et al. 2008) aboard Hinode provides precision measures of the magnetic field vector of active regions. This instrument has been used to record the history of development of a filament channel (Okamoto et al. 2008) observed simultaneously with the SP and with the narrow-band filter instrument (NFI) on Hinode observing in the core of the hydrogen Balmer α (Hα) line. Those observations were interpreted as indicative of a flux rope rising through the photosphere. Here we present another particularly well-observed example of the formation and disappearance of a filament channel observed with Hinode during 2007 December 9 to 12 in NOAA region 10978 (Fig. 1: the region of interest for the filament channel is the larger outlined area). For these observations we did not observe the formation of the filament directly in Hα, but observations in the Ca II H-line at 396 nm using the broad-band filter instrument (BFI) for SOT clearly indicate the presence of a filament. The SP vector magnetograms taken every 4–6 hours were accompanied by frequent BFI images in Ca II and the G-band, plus NFI magnetograms in the Na I D1 line at 589.6 nm. Movies constructed from the BFI and NFI sequences serve to provide the evolutionary context of the formation and decay of an active region filament. Figure 2 presents some snapshots from these movies at the times of SP vector magnetic field maps. That figure shows both intrinsic field strength from the SP maps and intensity from the BFI Ca II H-line filter.

The filament channel forms in a somewhat disjoint fashion. The leftmost images at 06 h on 10 December show the region before the formation of the filament channel, where sub-kiloGauss fields are commonplace and there are small areas of sheared field near the PIL, but no coherent inverse configuration exists. By 11 h on 10 December (second panel of Fig. 2) the channel has formed over the consolidated PIL as evidenced by the darkening along the PIL in the Ca II image, and a significant length of the PIL has inverse configuration. By 16 h on the 10th the channel is fully formed and has reached is maximum width in both the field strength and Ca II images. The weaker fields are flanked by strong ($|\mathbf{B}| = 1300$–1600 Gauss) plage fields. After this phase, the channel narrows and lengthens, and the inverse magnetic

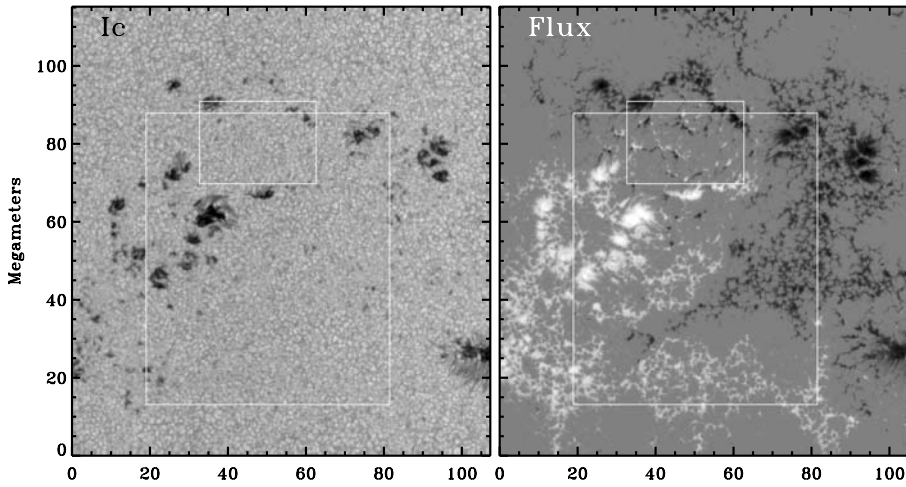


Fig. 1 Observations with the Spectro-Polarimeter (SP) aboard the Hinode satellite of NOAA active region 10978 during December 2007 provide excellent examples of both the evolution of an active region filament and actively emerging flux. These images from a high resolution SP map of the Fe I 630 nm lines obtained between 11:14–12:41 UT on 2007 December 11 show the continuum intensity (*left*) and vertical apparent flux density $B^L_{\rm app}$ (*right*, see Lites et al. 2008 for definition of $B^L_{\rm app}$). The *large central box* outlines the region surrounding the filament channel (at the PIL running from *lower-left* to *upper-right*). The *smaller box* outlines an area of emerging flux. All figures presented in this section are oriented with solar north *up* and solar west to the *right*

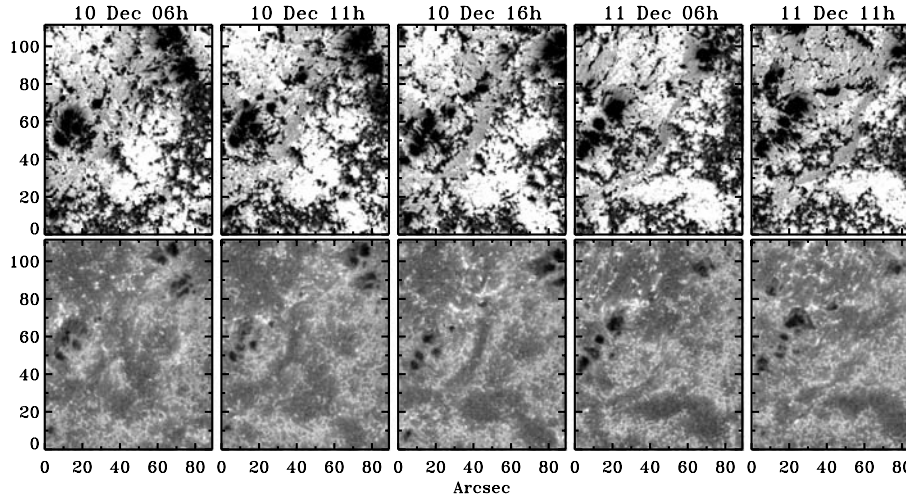


Fig. 2 Hinode/SP vector field observations of NOAA active region 10978 provide measurements of the *intrinsic* magnetic field strength |**B**| within the active region (*top panels*: light-dark scaled from 0–1500 Gauss) as derived from a Milne-Eddington inversion. *White areas* in the field strength images are locations where the polarization signal was weak enough that no inversion was attempted. The *bottom panels* show the Ca II H-line (3 Å width) filter measurements obtained near the midpoint of each SP map. The final map corresponds to the data shown in the larger outlined area of Fig. 1. The filament channel is visible as a dark structure running from *lower left* to *upper right* in the Ca II images, and as a low field strength structure in the *upper panels*

configuration at the photosphere exists along most of the length of the channel. Note that the PIL and the channel are easily identified as an extended lane of weak ($|\mathbf{B}| = 400-600$ Gauss) field. After the final observation shown in Fig. 2 there are no further vector field observations during the following 21 hours, but nearly continuous Ca II imaging shows the filament channel to narrow, break into shorter segments, and dissipate. During the later part of the filament channel evolution, the NFI Na I D-line magnetograms show the flanking, intrinsically strong plage fields to slowly drift toward the PIL and cancel there.

2.2 The Occurrence of a "Naked Bald Patch"

The formation of the filament channel described above is a particularly clear example of the process that has been described previously (Lites 2005; Okamoto et al. 2008). However, one aspect of the December 2007 observations that has not been described previously is the occurrence of an *isolated* length of weak, horizontal, inverse configuration fields apparent in the upper right two panels of Fig. 2. The segment of the filament channel just to the right and above the center of the $|\mathbf{B}|$ images presented for 06 h and 11 h on 2007 December 11 show an absence of strong plage flux immediately flanking the filament channel. A "bald patch" is a term that has been used in solar vector magnetometry for a region of inverse configuration transverse fields. Such locations typically do not show any significant signal in longitudinal magnetograms, hence they are designated "bald". Because this segment of the filament channel is not flanked by strong field plage, we designate this region as a "naked bald patch" (NBP, not to be confused with "network bright point").

Figure 3 shows an expanded view of the region around the filament channel from the upper right panel of Fig. 2. On that figure have been superimposed arrows indicating the direction of the field vector in the plane of the solar surface. The azimuth of the fields are such that selection of the opposite resolution of the azimuth ambiguity nearly anywhere within the map results in the presence of physically unacceptable discontinuities of the field orientation. The inverse field orientation is evident along most of the length of the filament channel, and is especially noticeable in the NBP. The white areas surrounding the NBP are regions where the polarization is weak enough that no inversion was attempted. Examination of the individual Stokes profiles in this region demonstrate that there are no significant longitudinal or transverse fields adjacent to the filament channel.

The occurrence of the NBP is significant because it indicates the presence of a concentration of matter in the atmosphere dense enough to weigh down the buoyant, hectoGauss fields onto the photosphere. For much of the length of this filament channel (and others observed like it) strong plage fields of opposite polarity exist on either side. For those regions, one might argue that the downward magnetic tension force arising from potential-like fields connecting these flanking plage fields and arching over the bald patch might be sufficient to contain a flux rope underneath them. In the case of the NBP, however, there is no flanking plage hence only the weight of the magnetized plasma can maintain the configuration in a quasi-stable state against magnetic buoyancy.

2.3 Emerging, Twisted Flux or Shear-Generated Flux Ropes?

If reconnection in the corona is responsible for creating the flux rope configuration, it would be necessary for enough mass to condense out of the corona onto the field lines forming the flux rope in order to arrive at nearly photospheric densities. This scenario appears unlikely given that prominences and filaments have not been observed to attain higher than chromospheric densities. The other alternative is that the field is sheared by photospheric

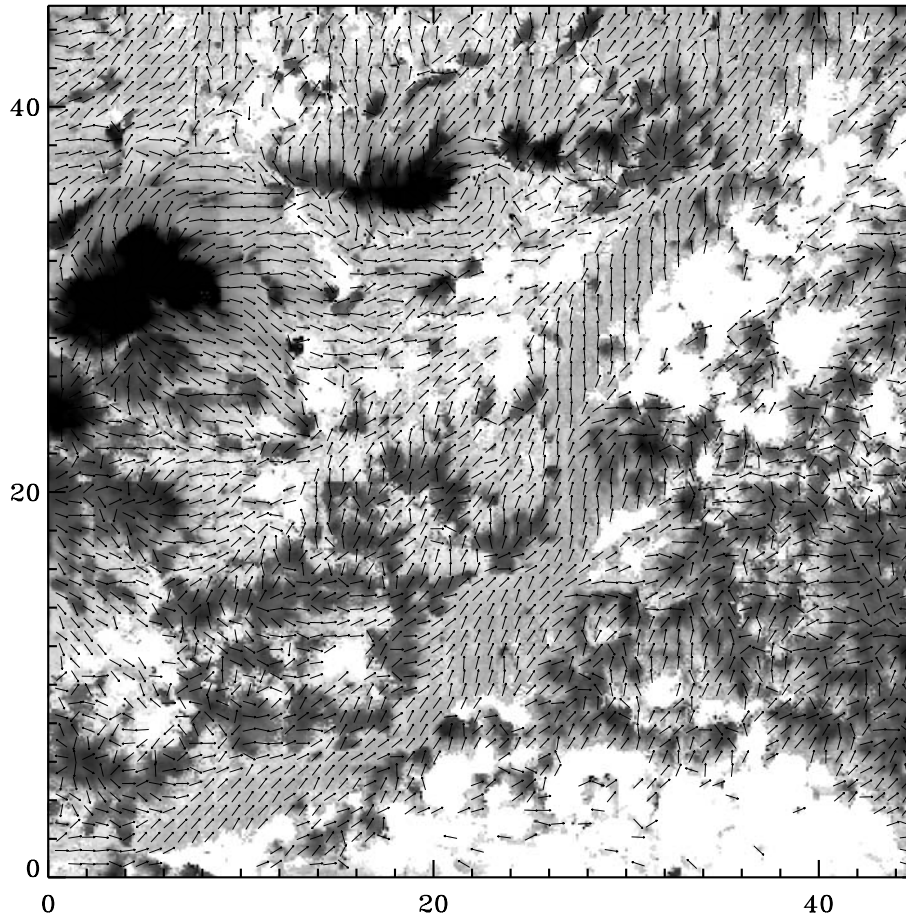


Fig. 3 An expanded view of |**B**| in region around the filament channel observed on 2007 December 11 (rightmost panels of Fig. 2). *Arrows* of equal length showing the orientation of the horizontal component of the magnetic field vector are superimposed. The field is directed mainly along the filament channel running from *lower left* to *upper right*, but over most of its length the field has inverse configuration: the arrows have a component directed from negative polarity (*lower right*) toward positive polarity (*upper left*). This is especially true in the "naked bald patch" region above and to the right of the center of the image where the filament channel is not flanked by strong plage. The gray scale ranges from 0 to 2000 Gauss (*white* to *black*), with low-polarization pixels where no inversion was attempted filled with *white*. *Arrows* are plotted every 6th pixel of the image

motions, then reconnects at, or just above the photosphere, in order to form the flux rope low in the atmosphere. In the case studied here, this scenario also appears to be unlikely because movies of the evolution of this region recorded both in the NFI magnetograms taken in the wing of the Na I D-line and in BFI Ca II H-line filtergrams show no evidence for systematic photospheric flows in the vicinity of, and especially along the full length of the filament channel. Furthermore, a preliminary analysis of photospheric flows determined by correlation tracking from the BFI G-band sequence reveals no systematic flows along the PIL.

In models of prominence formation arising from shear localized near the PIL, an important ingredient is the constraint imposed by the potential-like bipolar arcade loops arching over the sheared region (e.g., Aulanier et al. 2002). It was noted in that paper that normal configuration fields can occur below regions where the bipolar field above is weak. The presence of rather strong inverse configuration at the NBP seemingly argues against the prominence-like configurations evolving in the atmosphere above a sheared arcade.

It must be noted that bipole centers of flux emergence occur throughout the active region during the period leading up to and after formation of the filament channel. Most of the systematic photospheric flows determined by correlation tracking are associated with either these centers of emergence or with the moat flows emanating from the sunspots. Because of the emergence, it is difficult to follow the development of the filament channel unambiguously in the BFI and NFI movie sequences, but in comparison to the filament channel, these emergence events are short-lived and smaller in scale, so that the emergence events do not support the notion of systematic shear along the PIL.

Taking a broader view, one might consider the emerging flux as part of the larger, twisted flux system emerging from below. Many observed properties of the formation of this filament channel appear to support the scenario of emerging, twisted flux. The presence of the NBP with distinctly inverse magnetic configuration unambiguously demonstrates that a low-lying flux rope exists in the solar photosphere that is constrained from erupting into the upper atmosphere mainly by the weight of its dense, photospheric plasma. This situation would arise naturally by the emergence of twisted flux from below. The progression from normal to inverse orientation at the PIL, and the accompanying sequential widening followed by narrowing of the filament channel are all hallmarks of the emerging flux rope as documented in previous studies (Okamoto et al. 2008). The slow convergence and cancellation of opposite polarity plage flux as observed at the PIL after the filament channel is fully formed would also be expected in the case of an emerging flux rope.

3 The Small-Scale Topology of Magnetic Fields at the Site of Emergence

A flux rope rising by its buoyancy into the solar atmosphere must find some way to unload the considerable mass constrained by gravity and the field to reside in the concave-upward (U-loop) field geometry. This issue has often been cited as an argument against the emerging flux rope scenario. The rope must find some means to rid itself of its mass, but the process must be slow and nearly continuous in order that the filament not be destabilized, especially if it is held down primarily by its own weight as in the case of the NBP discussed in the previous section.

We look to the small-scale topology and evolution of the magnetic field in emerging flux regions for clues as to how emerging flux forming U-loops rids itself of mass. High resolution observations from Hinode support the notion that field reconnection at or just below the surface, occurring at very small scales, is a likely avenue for mass-unloading. The concept is that flows on the scale of granulation at the photospheric level and below would force together the vertical sections of the U-loop, making them merge and reconnect. The mass constrained to the field will always fall to the lowest point along the field line. Reconnection then forms a subsurface O-loop, thereby freeing the reconnected section of the remaining U-loop to rise further into the atmosphere. In this section we examine Hinode observations of an emerging flux system in detail.

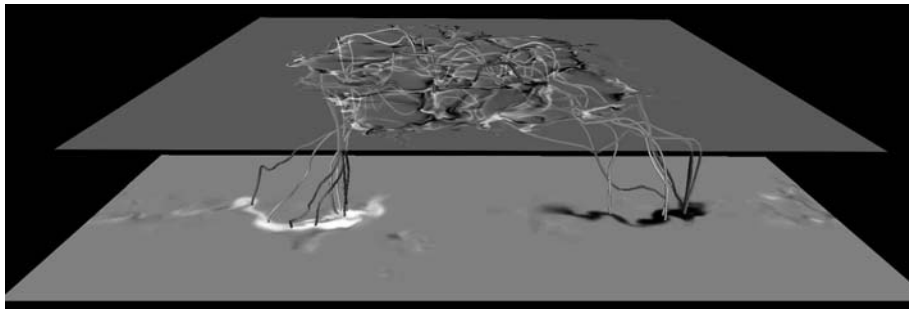


Fig. 4 Simulations of magneto-convection in the presence of a buoyantly rising flux system now have a striking resemblance to emerging regions observed at high resolution with Hinode. This image taken from the simulations of Cheung et al. (2008) shows the magnetic field as a gray scale on two planes: −5 Mm (*lower*) and 0 Mm (*upper*). A few selected field lines are highlighted to illustrate the serpentine nature of the field near the solar surface. At depth, the field is more or less bipolar, but at the surface, granular convection distorts the field into the "cloud" of opposite polarities, like those often seen in Hinode observations

3.1 Hinode SP Observations of Emerging Flux

"Bipolar" magnetic flux does not emerge as a simple bipole. Hinode magnetogram movie sequences reveal that the initial emergence is characterized by a "cloud" of very small-scale, mixed polarity flux. Small scale bipolar flux elements emerge, generally in narrow elongated strands, and the opposite poles rapidly stream away from each other. Most of such emergence occurs along the southeast-northwest line defining the overall poloidal orientation of the emergence event. As the emergence event proceeds, many of the emerged flux elements approach flux of the opposite polarity within the emergence region and appear to cancel in place. We find very little of the process of opposite polarity flux elements streaming by each other as reported by Strous (1994) and Strous et al. (1996). Perhaps that phenomenon occurs on a larger scale and the cancellation process we note here are mainly visible only at the high resolution of Hinode.

Rapid sequences of Hinode SP maps reveal flux emergence events within individual granules in the quiet Sun (Centeno et al. 2007) that appear initially as horizontal fields then develop into bipolar structures. Thus, on the scale of granules the flux appears to emerge as simple bipolar structures (although the spatial resolution of those measurements is limited to 0.3″, so finer scale structure probably exists). On somewhat larger scales, one sees the clouds of mixed polarity emerging flux, both in quiet and active regions, in the Hinode NFI magnetogram sequences. There is good reason to believe that flux emerging on scales larger than individual granules will be strongly influenced by the granular convection. When emerging flux reaches the photosphere, the observed horizontal field strengths are a few hundred Gauss (Lites et al. 1998; Kubo et al. 2003); i.e., roughly in equipartition with the granular convective flows. These fields are therefore susceptible to drastic distortion by convective action. Indeed, magneto-convection simulations of emerging bipolar flux (Cheung et al. 2008) show just such a complex structure; see Fig. 4.

The December 2007 observations of NOAA active region 10978 not only reveal the formation and dissipation of a filament channel as described above, but they also comprise excellent SP observations of an emerging region. The SP map of Fig. 1, obtained on 2007 December 11 between 11:14 and 12:41 UT, is unique in that it covers an actively emerging region at full angular resolution of the SP (0.16″ pixels). We use the data of this region

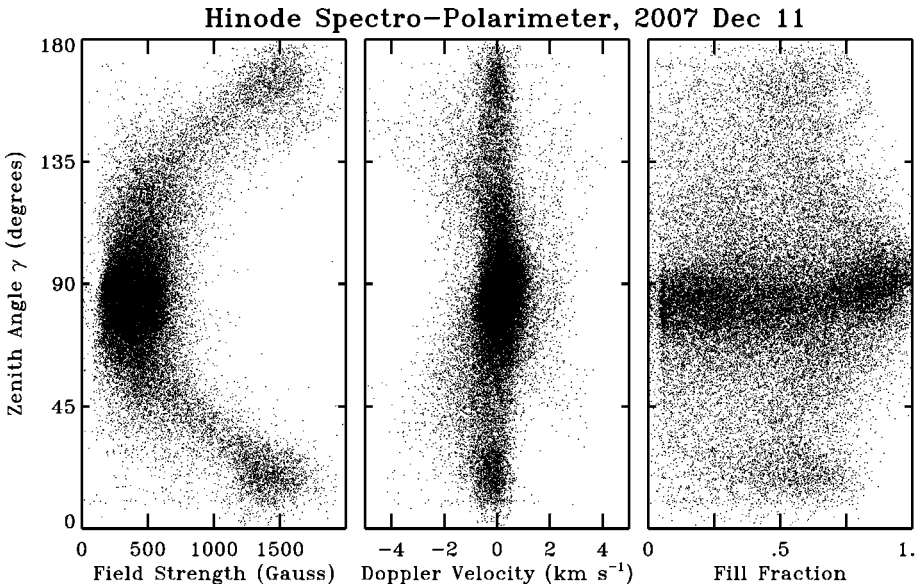


Fig. 5 Scatter plots are shown of the magnetic field zenith angle (relative to the local normal to the solar surface) versus |**B**| (*left*), Doppler velocity of the magnetic elements (*middle*), and fill fraction (*right*) as determined by inversion of the emergence region for NOAA region 10978. Selected points avoid pores and sunspots in the emergence region outlined by the smaller box of Fig. 1. Emerging fields are mostly horizontal and have gentle upward flows

(smaller box in Fig. 1) to illustrate the topology and dynamics of emerging flux at small scales.

3.2 Fine-Scale Properties of Emerging Flux

Inversions of the SP map of NOAA active region 10978 demonstrate that the emergence region outlined by the smaller box of Fig. 1 is dominated by weak (200–600 Gauss), horizontal field. Near the top of the upper right panel of Fig. 2 one sees a large area of weak field punctuated by small concentrations of strong field. The weak horizontal fields have been noted previously at lower resolution (Lites et al. 1998; Kubo et al. 2003). Figure 5 presents scatter plots for the emergence region (avoiding sunspots and pores) comparing field strength, Doppler velocity, and fill fraction versus the field zenith angle (angle relative to the local normal to the photosphere) for the emergence region outlined in Fig. 1. There is a clear preference for locally horizontal fields having strengths of a few hectoGauss, and slightly upward motions. The kiloGauss field concentrations are close to vertical with a slight preference for down flow, and for fill fractions less than unity.

Figure 6 demonstrates the flow structure of the emergence region outlined in Fig. 1. Displayed on the right are continuum intensity (top) and $B^L_{\rm app}$ (bottom, scaled ± 700 Mx cm^{-2}). On the left of Fig. 6 are monochromatic images of Stokes V (hereafter denoted "wing magnetograms", originally applied to Hinode SP data by Ichimoto et al. 2007) measured at ± 260 mÅ from the center of Fe I 630.25 nm. The lower (upper) left image shows the locations of strong, localized up (down) flows in the magnetized regions. The sign of these wing magnetograms has been set so that the polarity of the fields agrees with that of the $B^L_{\rm app}$ image at lower right. The locations of red- (blue-) shifted field concentrations are indicated by

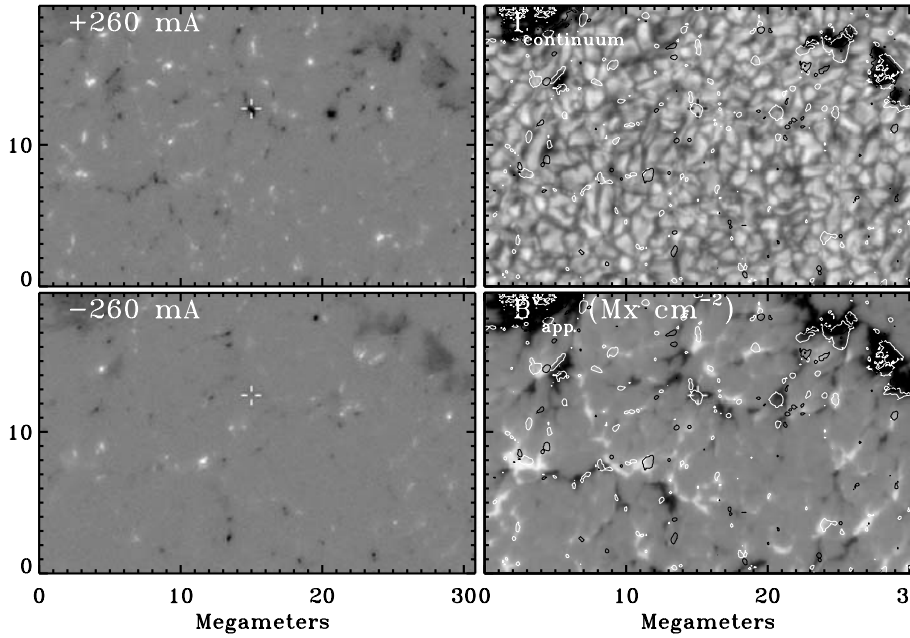


Fig. 6 Hinode/SP observations of the emergence region in NOAA active region 10978 outlined in the small box in Fig. 1 reveal extremely small scale, high speed flows in magnetized regions. The continuum image (*upper right*) and apparent longitudinal flux density $B^L_{\rm app}$ (*lower right*) with gray scale ± 700 Mx cm^{-2} are shown. The panels on the left show the corresponding wing magnetograms at ± 260 mÅ from the center of Fe I 630.25 nm. *White* (*black*) contours on the panels at right indicate concentrations of flux having large redshift (blueshift). The *small cross* just north of the center of the panels denotes the location of the Stokes profiles presented in Fig. 7

white (black) contours. Several conclusions may be drawn from these wing magnetograms: (1) the fast flows in the emergence region are highly localized; that is, generally smaller than granules, (2) strong down flows are more prevalent than strong up flows, and (3) the strongest down flows usually occur near bright points in the intergranular lanes; i.e., where there are kiloGauss flux concentrations.

When we examine Stokes profiles at the site of vigorous down flows in Fig. 6, we find that there are two spatially unresolved components to the field. Figure 7 shows one such Stokes profile measurement (dots) for which we have carried out a Milne-Eddington inversion with two magnetic components and one non-magnetic component (solid lines). These profiles are from the location of the cross superimposed on the images of Fig. 6. One magnetic component is nearly at rest, while the other has a down flow of about 10 km s^{-1}. The fact that two components appear to be present within this single pixel indicates that these flows are smaller than the $0.3''$ Hinode SP resolution. Clearly there are highly concentrated, high velocity flows in some strong magnetic field concentrations in the emergence region.

The $B^L_{\rm app}$ image at lower right of Fig. 6 shows the characteristic highly mixed polarity of the emergence region. Those horizontal fields are characterized by mild up flows. We now examine the topology of the magnetic field around the juxtaposed opposite polarity, strong field elements seen in $B^L_{\rm app}$. Figure 8 shows a perspective plot of the vector magnetic field in the vicinity of the "colliding" fields just to the south and east of the center of $B^L_{\rm app}$ in Fig. 6. This image shows that, like the filament channel discussed in the previous section, the fields

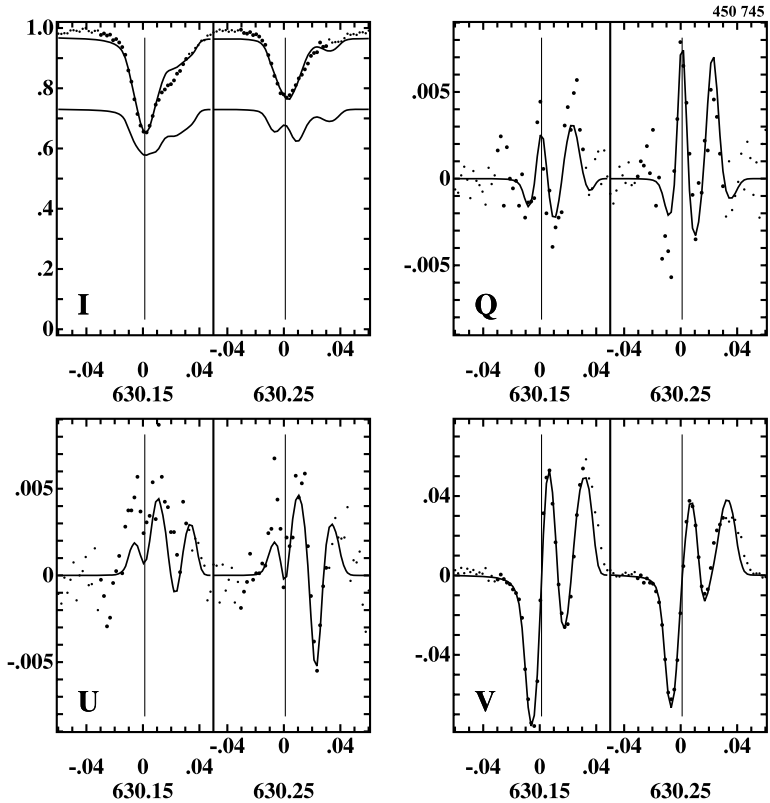


Fig. 7 Observed spectral Stokes profiles (*dots*) are shown for one pixel highlighted by the *cross* just to the north of the center of images in Fig. 6. The *solid curves* represent a 3-component Milne-Eddington fit to these profiles: two magnetic components plus one non-magnetic component. *Vertical lines* show the zero velocity line positions of the Fe I 630 nm lines. The *lower solid curve* in the Stokes *I* panel shows the combined profiles from the two magnetized components. One magnetized component is essentially at rest, while the other is red-shifted by 10.3 km s^{-1}, indicating an intense down flow that is unresolved at 0.3″ resolution

have a concave-upward (U-loop) geometry. Unlike the filament channel, the fields here are more nearly perpendicular than parallel to the PIL (shown as white contours). Away from these local PILs, the pervasive weaker horizontal fields are measured at most locations, as indicated by the arrowheads in Fig. 8.

3.3 The Role of Small Scale Processes in Flux Emergence

The Hinode observations of emerging flux described herein give perhaps the clearest view yet of the processes controlling the emergence of flux at the photosphere:

- The dominant character of an emerging flux region is relatively weak (200–600 Gauss), horizontal fields, with orientation aligned more or less with the overall bipole of the emergence event.
- The horizontal field regions have a tendency to harbor a weak rising motion.
- The emerged flux reaches the surface as a cloud of mixed polarity on small scales. This mixed polarity occurs on the scale of granules, and thus appears to be a direct result of granular convection acting on a larger scale rising Ω-loop.

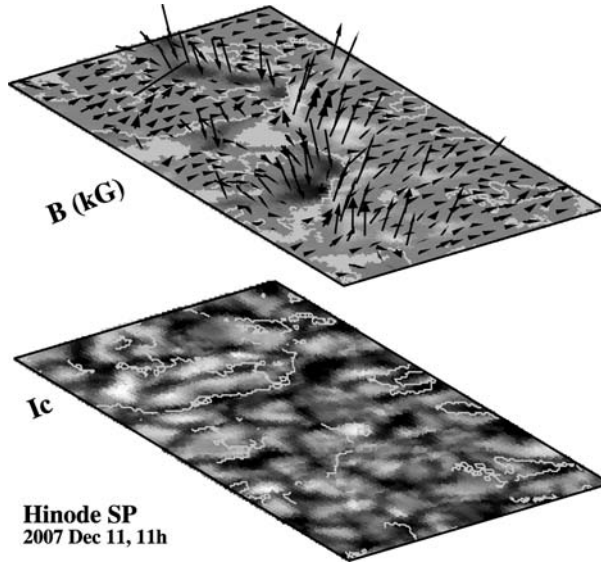


Fig. 8 The continuum intensity (*bottom*) and a perspective view of the vector magnetic field (*top*) is presented for the "colliding field" region just south and east of the center of images in Fig. 6. *White contours* on the images are the locations of the polarity inversion lines (PIL). The *top image* presents the vertical flux density ($f|\mathbf{B}|\cos\psi$, where ψ is the zenith angle of the field) scaled ± 1500 Mx cm^{-2}. The *arrows* shown in perspective represent the orientation of the field vector, and their lengths are proportional to the intrinsic strength of the field $|\mathbf{B}|$. *Arrows* are depicted every fourth observed pixel. North is directed to the *right* along the shorter edge of the displayed rectangle, and west is directed *downward* and to the *right* of the longer edge

- Opposite polarity, strong (kiloGauss) field concentrations within the emergence zone approach each other and cancel in place.
- The vertical flux in the emergence region exists as concentrated photospheric flux tubes residing in the intergranular lanes.
- At the level of formation of the 630 nm Fe I lines, and within the resolution limitations of the Hinode SP (0.3″), the canceling strong field elements have a concave-upward (U-loop) topology.
- High speed, tiny down flows are scattered throughout the emergence region. These flows are associated (but not exactly cospatial) with the kG concentrations of more vertical flux. In regions where opposite polarity flux is canceling, the down flows are located on the sides of the flux elements away from the PIL.

All of these properties are consistent with a buoyantly rising bipolar flux system that is disrupted near the surface as realistically depicted by the simulations of Cheung et al. (2008) as shown in Fig. 4. In their buoyant rise, the field lines carry plasma upward until they reach the region of granular convection near the surface. At that point, the convective flows overcome the field and cause it to take on the serpentine character described by Cheung et al. (2008). The plasma contained in the rising horizontal segments of these field lines descends toward the nearest local minimum of the field lines, which happen to be concentrated in the intergranular lanes as intense flux tubes. These tubes are the spout of the funnel concentrating the draining of the mass that rises with the large regions of nearly horizontal field.

Note that the intense down flows could result from convective collapse and such a process has been noted in the simulations of Cheung et al. (2008). However, those authors offer an alternate explanation for the phenomenon: their simulations indicate that the flows are caused by the Lorentz force due to highly-stressed field lines in the region of flux cancellation. Most of the prominent down flows visible in Fig. 6 occur on the sides of strong flux concentrations away from the PIL, and not coincident with the strongest fields. Similarly, steady supersonic down flows have been observed near the edges of pores. From the present Hinode observations we are unable to determine if these flows persist for long periods. In any case, the observed configuration favors either the draining funnel scenario offered above, or convective collapse. The Lorentz force explanation of Cheung et al. (2008) appears less likely to be operative because the down flows are located away from the PIL and in regions where field gradients and strengths are therefore relatively small. It should be noted that very small-scale, persistent down flows have been seen at the PIL of what apparently is a closed, rising magnetic system (Mártinez Pillet et al. 1994). In that case, the steady nature of the flow precludes transient processes such as reconnection and convective collapse—it is more readily understood as a site of mass drainage in the rising flux system. The time history of the down flow events observed in emerging flux systems is crucially important to understanding the responsible mechanism, so further studies using Hinode and other high resolution instrumentation are needed.

As the opposite polarities approach each other, they reconnect at or below the surface, effectively releasing the dense plasma in the form of a descending "O-loop" constrained within the downward convective plume of an intergranular lane. This action frees the field above to continue its buoyant rise into the chromosphere and corona. This process is similar to that depicted in Fig. 3 of Spruit et al. (1987), although what is described here would resemble that figure when mirrored in the vertical direction.

Acknowledgements I wish to thank all those involved in the international Solar-B/Hinode program for their dedication and work without which the new insights provided by this mission would not have been realized. I thank T.J. Okamoto, R. Shine, and A. Title for discussions and helping with processing of Hinode images. Thanks also go to P. Judge for reviewing the manuscript. Hinode is a Japanese mission developed and launched by ISAS/JAXA, with NAOJ as domestic partner and NASA and STFC (UK) as international partners. It is operated by these agencies in co-operation with ESA and NSC (Norway). The FPP project at LMSAL and HAO is supported by NASA contract NNM07AA01C.

References

W.P. Abbett, G.H. Fisher, Y. Fan, The three-dimensional evolution of rising, twisted magnetic flux tubes in a gravitationally stratified model convection zone. Astrophys. J. **540**, 548–562 (2000)

R.G. Athay, C.W. Querfeld, R.N. Smartt, E. Landi degl'Innocenti, V. Bommier, Vector magnetic fields in prominences. III—He I D3 Stokes profile analysis for quiescent and eruptive prominences. Sol. Phys. **89**, 3–30 (1983)

G. Aulanier, C.R. DeVore, S.K. Antiochos, Prominence magnetic dips in three-dimensional sheared arcades. Astrophys. J. **567**, L97–L101 (2002)

M. Berger, Rigorous new limits on magnetic helicity dissipation in the solar corona. Geophys. Astrophys. Fluid Dyn. **30**, 79–104 (1984)

M.G. Bobra, A.A. van Ballegooijen, E.E. DeLuca, Modeling nonpotential magnetic fields in solar active regions. Astrophys. J. **672**, 1209–1220 (2008)

R. Centeno, H. Socas-Navarro, B. Lites, M. Kubo, Z. Frank, R. Shine, T. Tarbell, A. Title, K. Ichimoto, S. Tsuneta, Y. Katsukawa, Y. Suematsu, T. Shimizu, S. Nagata, Emergence of small-scale magnetic loops in the quiet-sun internetwork. Astrophys. J. **666**, L137–L140 (2007)

M.C.M. Cheung, M. Schüssler, T.D. Tarbell, A.M. Title, Solar surface emerging flux regions: a comparative study of radiative MHD modeling and Hinode SOT observations. Astrophys. J. (2008, submitted)

S.E. Gibson, L. Fletcher, G. Del Zanna, C.D. Pike, H.E. Mason, C.H. Mandrini, P. Démoulin, H. Gilbert, J. Burkepile, T. Holzer, D. Alexander, Y. Liu, N. Nitta, J. Qiu, B. Schmieder, B.J. Thompson, The structure and evolution of a sigmoidal active region. Astrophys. J. **574**, 1021–1038 (2002)

S.E. Gibson, Y. Fan, T. Török, B. Kliem, The evolving sigmoid: evidence for magnetic flux ropes in the corona before, during, and after CMES. Space Sci. Rev. **124**, 131–144 (2006)

K. Ichimoto, R.A. Shine, B. Lites, M. Kubo, T. Shimizu, Y. Suematsu, S. Tsuneta, Y. Katsukawa, T.D. Tarbell, A.M. Title, S. Nagata, T. Yokoyama, M. Shimojo, Fine-scale structures of the evershed effect observed by the solar optical telescope aboard Hinode. Publ. Astron. Soc. Jpn. **59**, S593–S599 (2007)

K. Ichimoto, B. Lites, D. Elmore, Y. Suematsu, S. Tsuneta, Y. Katsukawa, T. Shimizu, R. Shine, T. Tarbell, A. Title, J. Kiyohara, K. Shinoda, G. Card, A. Lecinski, K. Streander, M. Hakagiri, M. Miyashita, M. Noguchi, C. Hoffmann, T. Cruz, Polarization calibration of the solar optical telescope onboard Hinode. Sol. Phys. **249**, 233–261 (2008)

T. Kosugi, K. Matsuzaki, T. Sakao, T. Shimizu, Y. Sone, S. Tachikawa, T. Hashimoto, K. Minesugi, A. Ohnishi, T. Yamada, S. Tsuneta, H. Hara, K. Ichimoto, Y. Suematsu, M. Shimojo, T. Watanabe, J.M. Davis, L.D. Hill, J.K. Owens, A.M. Title, J.L. Culhane, L. Harra, G.A. Doschek, L. Golub, The Hinode (Solar-B) mission: An overview. Sol. Phys. **243**, 3–17 (2007)

M. Kubo, T. Shimizu, B.W. Lites, The evolution of vector magnetic fields in an emerging flux region. Astrophys. J. **595**, 465–482 (2003)

J.L. Leroy, Observation of prominence magnetic fields, in *Dynamics and Structures of Quiescent Solar Prominences*, ed. by E.R. Priest (Kluwer, Dordrecht, 1989), pp. 77–113

B.W. Lites, Magnetic flux ropes in the solar photosphere: the vector magnetic field under active region filaments. Astrophys. J. **622**, 1275–1291 (2005)

B.W. Lites, B.C. Low, V. Martínez Pillet, P. Seagraves, A. Skumanich, Z.A. Frank, R.A. Shine, S. Tsuneta, The possible ascent of a closed magnetic system through the photosphere. Astrophys. J. **446**, 877–894 (1995)

B.W. Lites, A. Skumanich, V. Martínez Pillet, Vector magnetic fields of emerging solar flux. I. Properties at the site of emergence. Astron. Astrophys. **333**, 1053–1068 (1998)

B.W. Lites, M. Kubo, H. Socas-Navarro, T. Berger, Z. Frank, R. Shine, T. Tarbell, A. Title, K. Ichimoto, Y. Katsukawa, S. Tsuneta, Y. Suematsu, T. Shimizu, S. Nagata, The horizontal magnetic flux of the quiet-sun internetwork as observed with the Hinode spectro-polarimeter. Astrophys. J. **672**, 1237–1253 (2008)

A. López Ariste, G. Aulanier, B. Schmieder, A. Sainz Dalda, First observation of bald patches in a filament channel and at a barb endpoint. Astron. Astrophys. **456**, 725–735 (2006)

B.C. Low, Coronal mass ejections, magnetic flux ropes, and solar magnetism. J. Geophys. Res. **106**, 25141–25164 (2001)

V. Mártinez Pillet, B.W. Lites, A. Skumanich, D. Degenhardt, Evidence for supersonic downflows in the photosphere of a delta sunspot. Astrophys. J. **425**, L113–L115 (1994)

T.J. Okamoto, S. Tsuneta, B.W. Lites, M. Kubo, T. Yokoyama, T.E. Berger, K. Ichimoto, Y. Katsukawa, S. Nagata, K. Shibata, T. Shimizu, R.A. Shine, Y. Suematsu, T.D. Tarbell, A.M. Title, Emergence of a helical flux rope under an active region prominence. Astrophys. J. **673**, L215–L218 (2008)

A.A. Pevtsov, R.C. Canfield, T.R. Metcalf, Latitudinal variation of helicity of photospheric magnetic fields. Astrophys. J. **440**, L109–L112 (1995)

D.M. Rust, A. Kumar, Helical magnetic fields in filaments. Sol. Phys. **155**, 69–97 (1994)

D.M. Rust, A. Kumar, Evidence for helically kinked magnetic flux ropes in solar eruptions. Astrophys. J. **464**, L199–L202 (1996)

T. Shimizu, S. Nagata, S. Tsuneta, T. Tarbell, C. Edwards, R. Shine, C. Hoffmann, E. Thomas, S. Sour, R. Rehse, O. Ito, Y. Kashiwagi, M. Tabata, K. Kodeki, M. Nagase, K. Matsuzaki, K. Kobayashi, K. Ichimoto, Y. Suematsu, Image stabilization system for Hinode (Solar-B) solar optical telescope. Sol. Phys. **249**, 221–232 (2008)

H.C. Spruit, A.M. Title, A.A. van Ballegooijen, Is there a weak mixed polarity background field? Theoretical arguments. Sol. Phys. **110**, 115–128 (1987)

L.H. Strous, Dynamics in solar active regions: patterns in magnetic-flux emergence. Ph.D. thesis, Utrecht University, The Netherlands, 1994

L.H. Strous, G. Scharmer, T.D. Tarbell, A.M. Title, C. Zwaan, Phenomena in an emerging active region. I. Horizontal dynamics. Astron. Astrophys. **306**, 947–959 (1996)

Y. Suematsu, S. Tsuneta, K. Ichimoto, T. Shimizu, M. Otsubo, Y. Katsukawa, M. Nakagiri, M. Noguchi, T. Tamura, Y. Kato, H. Hara, I. Mikami, H. Saito, T. Matsushita, N. Kawaguchi, T. Nakaoji, K. Nagae, S. Shimada, N. Takeyama, T. Yamamuro, The solar optical telescope of Solar-B (Hinode): The optical telescope assembly. Sol. Phys. **249**, 197–220 (2008)

S. Tsuneta, Y. Suematsu, K. Ichimoto, T. Shimizu, M. Otsubo, S. Nagata, Y. Katsukawa, A. Title, T. Tarbell, R. Shine, B. Rosenberg, C. Hoffmann, B. Jurcevich, M. Levay, B. Lites, D. Elmore, T. Matsushita,

N. Kawaguchi, I. Mikami, S. Shimada, L. Hill, J. Owens, The solar optical telescope for the Hinode mission: An overview. Sol. Phys. **249**, 167–196 (2008)

Y. Uchida, A. McAllister, K.T. Strong, Y. Ogawara, T. Shimizu, R. Matsumoto, H.S. Hudson, Continual expansion of the active-region corona observed by the YOHKOH soft X-ray telescope. Publ. Astron. Soc. Jpn. **44**, L155–L160 (1992)

A.A. van Ballegooijen, D.H. Mackay, Model for the coupled evolution of subsurface and coronal magnetic fields in solar active regions. Astrophys. J. **659**, 1713–1725 (2007)

Sunspots: From Small-Scale Inhomogeneities Towards a Global Theory

Rolf Schlichenmaier

Originally published in the journal Space Science Reviews, Volume 144, Nos 1–4, 213–228.
DOI: 10.1007/s11214-008-9465-6

Abstract The penumbra of a sunspot is a fascinating phenomenon featuring complex velocity and magnetic fields. It challenges both our understanding of radiative magnetoconvection and our means to measure and derive the actual geometry of the magnetic and velocity fields. In this contribution we attempt to summarize the present state-of-the-art from an observational and a theoretical perspective.

We describe spectro-polarimetric measurements which reveal that the penumbra is inhomogeneous, changing the modulus and the direction of the velocity, and the strength and the inclination of the magnetic field with depth, i.e., along the line-of-sight, and on spatial scales below 0.5 arcsec. Yet, many details of the small-scale geometry of the fields are still unclear such that the small scale inhomogeneities await a consistent explanation.

A simple model which relies on magnetic flux tubes evolving in a penumbral "background" reproduces some properties of sunspot inhomogeneities, like its filamentation, its strong (Evershed-) outflows, and its uncombed geometry, but it encounters some problems in explaining the penumbral heat transport. Another model approach, which can explain the heat transport and long bright filaments, but fails to explain the Evershed flow, relies on elongated convective cells, either field-free as in the gappy penumbra or filled with horizontal magnetic field as in Danielson's convective rolls. Such simplified models fail to give a consistent picture of all observational aspects, and it is clear that we need a more sophisticated description of the penumbra, that must result from simulations of radiative magnetoconvection in inclined magnetic fields. First results of such simulations are discussed. The understanding of the small-scales will then be the key to understand the global structure and the large-scale stability of sunspots.

Keywords Sunspots · MHD

R. Schlichenmaier (✉)
Kiepenheuer-Institut für Sonnenphysik, Schöneckstr. 6, 79104 Freiburg, Germany
e-mail: schliche@kis.uni-freiburg.de

M.J. Thompson et al. (eds.), *The Origin and Dynamics of Solar Magnetism*.
DOI: 10.1007/978-1-4419-0239-9_13

1 Introduction

Magnetic fields on the Sun exist in a large variety of phenomena and interact in various ways with the plasma and the radiation. In the convection zone large and small scale magnetic fields are generated. These magnetic fields are partially transported into the outer layers of the Sun, i.e., into the chromosphere and the corona. The most prominent example of a magnetic phenomenon is a sunspot as seen in the photosphere. A typical sunspot has a lifetime of a few weeks and has a size of about 30 granules. The magnetic field strength spans from 1000 to 3000 gauss in the deep photosphere, summing up to a magnetic flux of some 10^{22} Mx.

The magnetic field of a sunspot extends into the interior as well as into the outer layers of the Sun. The most detailed information of sunspots is obtained in the photosphere. The topology of the magnetic field above and beneath the photosphere is poorly understood. In particular our knowledge of the magnetic field extension into the interior presents a theoretical challenge. Direct measurements of the sub-photospheric structure are impossible, but at least for the larger scales, indirect methods are being explored in the framework of local helioseismology (cf. Gizon 2008).

Time Scales: Although the sunspot is a coherent phenomenon on large spatial and temporal scales, it seems crucial to realize that it is not static, but finds a dynamical equilibrium: A variety of small-scale features evolve on a dynamic time scale to produce a large-scale coherent structure on long time scales. This "fine structure" is complex and is seen in white light images in form of umbral dots, light bridges, bright and dark penumbral filaments, penumbral grains, dark-cored bright filaments, penumbral twists, and other features. This intensity fine structure corresponds to a fine structure of the velocity field and the magnetic field, which will be described below. The dynamic fine structure forms a globally stable sunspot and it is the goal of sunspot physics to understand how an ensemble of short-lived features with small scales is organized to form a coherent large and long-living sunspot.

2 Energy Transport in Umbra and Penumbra

The coolness of sunspots relative to the surrounding quiet Sun is readily explained by the tension of the magnetic field which tends to suppress convective motions. It is more difficult to understand why sunspots are as hot as they are: Neither radiative transport nor heat conduction can account for the surface brightness of sunspots. Hence convection cannot be fully suppressed and the energy must be transported by convective flows. Indeed, the fine structure manifests the inhomogeneities of the magnetic and velocity field and testifies that the energy transport in sunspots happens on small spatial scales by the motion of plasma. Yet, the crucial question is about the interaction between convective flows, the magnetic field, and the radiation. Are the flows non-magnetic or magnetic? What is their intrinsic spatial scale? Do coherent up- and downflows exist, similar to the granulation in the quiet Sun?

Jelly Fish and Field-Free Gaps: Parker (1979) has introduced the jelly fish model in which the sub-photospheric magnetic field separates into individual bundles of field lines, resulting in gaps free of magnetic field. The gaps between these bundles open up into very deep layers, being connected to the quiet Sun convection. Within these cracks, the field-free plasma would convect and transport heat upwards. An umbral dot would correspond to the peak of a field-free gap. More recently, Spruit and Scharmer (2006) suggested that such field-free gaps in the inclined magnetic field of the penumbra may result in elongated bright

filaments, instead of in point-like dots, thereby proposing an explanation for the brightness of the penumbra. The surplus brightness of the penumbra relative to the umbra would then be due to the fact that the convective cell can become larger in the more inclined and weaker magnetic field as in the less inclined (more vertical) and stronger field of the umbra.

Stability of Sunspots and Monolithic Models: Sunspots are stable relative to the dynamical time, i.e., Alfvén waves are estimated to travel across a spot in about 1 h, while the life time is in order of weeks. How can it be that all this dynamic fine structure constitutes a spot which is stable? The question of stability can be addressed if one assumes a "simple" vertical magnetohydrostatic magnetic flux tube that fans out with heigth. In such models the heat transport is attributed to (magneto-) convection, but is parametrized by a reduced mixing length parameter (Jahn 1989; Deinzer 1965). The dynamic fine structure is ignored and only their averaged effect on the stratification for umbra and penumbra is accounted for. The configuration is in magneto-static equilibrium together with a hydrostatic equilibrium vertically and with a total pressure balance between the umbra, penumbra, and quiet Sun horizontally (see e.g. Jahn and Schmidt 1994; Pizzo 1990). This configuration can be stable against the interchange instability (Meyer et al. 1977), at least in the first 5 Mm or so beneath the photosphere (Jahn 1997). In these upper layers of the convection zone the inclination of the interface between spot and surrounding is so large that buoyancy forces make the spot to float on the granulation. In deeper layers, beyond 5 Mm, the inclination of the outermost magnetic field line, i.e., the magnetopause, is small relative to the vertical. There, interchange (fluting) instability is no longer suppressed by buoyancy effects, and the magnetic configuration of a monolithic sunspot is unstable. Indeed, it has been proposed that the magnetic field strength progressively weakens in these deep layers shortly after the formation of a sunspot. The decreasing field strength, the convective motions, and the interchange instability dynamically disrupt the sunspot magnetic field from the deeper roots (Schüssler and Rempel 2005). Hence, the magnetic field in the deeper layers may be dispersed, but the floating part of the sunspot is stable.

3 Inhomogeneities in Umbra and Penumbra

For an extensive review of the sunspot structure, we refer the reader to an instructive overview by Solanki (2003).

3.1 Umbral Dots

The umbra of a sunspot harbors dynamic inhomogeneities. They are observed as dot-like bright spots with typical sizes of half an arcsec or less, embedded in a more uniform and darker background. These umbral dots seem to be present in all sunspots, although their intensity varies a lot. In some spots they can be almost as bright as bright penumbral filaments, in other spots their intensity is much smaller. In the latter case, the dot-like intensity variations occurs in a background that also shows a lower intensity.

The Physics of Umbral Dots: Umbral dots are an obvious signature of convection, yet it is not so obvious to understand the type of convection that leads to umbral dots. In the field-free gap idea of Parker, the convection is confined by the strong surrounding magnetic field, such that the column of convection narrows upwards and only a small brightening is seen at the surface. Observationally, it is established that the magnetic field in umbral dots is weaker

than in the surroundings and that an upflow of at least a few hundred m/s is associated with them (Socas-Navarro et al. 2004; Rimmele 2004, 2008; Bharti et al. 2007). The latter two observations also establish the presence of dark lanes across umbral dots.

The most recent simulations of radiatively driven magneto-convection in strong vertical magnetic field (Schüssler and Vögler 2006) result in local convective cells which produce umbral dots as well as their dark lanes. These cells barely touch the photosphere, similar as in Parker's idea. The cells extend downward for a few Mm, in the first 1 Mm the cells have a weak magnetic field strength. The weak field strength is caused by magnetic flux expulsion, i.e. convection advects the magnetic field (as in Weiss 1964). In the simulations the magnetic field strength in the cells amounts to a few hundred gauss, but this number may decrease if magnetic dissipation is reduced in more advanced simulation runs. In any case, in deeper layers the magnetic field strength increases considerably. Hence the cells do not connect to field-free plasma in deeper layers. In this respect these new simulations change our model vision of umbral dots. Now, we may conceive that the umbra is an overall monolithic fully magnetic structure, in which the fine structure is a local disturbance. The dots are produced locally by magneto-convection processes, which are needed for the energy transport.

3.2 Penumbral Inhomogeneities

3.2.1 Morphological Description

The penumbra is a manifestation of small-scale structure. The variety of the penumbral intensity fine structure is described in detail in the contribution to this volume by Göran Scharmer. In essence, there are bright and dark filaments, as well as penumbral grains. It turns out that bright filaments have dark cores (Scharmer et al. 2002; Sütterlin et al. 2004) and that intensity twists exist along bright filaments (Ichimoto et al. 2007b) on spatial scales of about 0.2 arcsec. The challenge consists in measuring the spectroscopic and spectropolarimetric signatures of this fine structure in order to derive their thermodynamic properties as well as their velocity and magnetic field. Only recently, with the technological advance of adaptive optics and with observations from space, it has become possible to acquire such high spatial resolution data for exposure times as long as 5 sec or more. This is a necessity to collect enough photons to have high spatial, spectral, and polarimetric resolution.

At a spatial resolution of better than half an arcsec, it can be demonstrated that not only the intensity and velocity, but also the magnetic field consists of a filamentary structure (Title et al. 1993; Langhans et al. 2005; Tritschler et al. 2007; Ichimoto et al. 2007a, 2008). Actually, at a spatial resolution of better than half an arcsec, all physical quantities in the penumbra show small-scale variations and predominantly filamentary (radially elongated) features.

However, the penumbra looks fairly uniform at a spatial resolution worse than 1 arc sec. At this lower spatial resolution, i.e. in average, the penumbra is brighter than the umbra, but less bright than in the surrounding granulation. But even if the penumbra is less bright in average, the small scale peak-to-peak intensity variation in the penumbra is larger than in the granulation, and the spatial scales of the variations are smaller than in the granulation. The same is true for velocities in the penumbra. Line-of-sight velocities in the penumbra of more than 5 km/s have been derived from Doppler shifts of photospheric lines (e.g., Wiehr 1995) and radial flow channels with widths of less than half an arcsec are observed (e.g., Tritschler et al. 2004; Rimmele and Marino 2006).

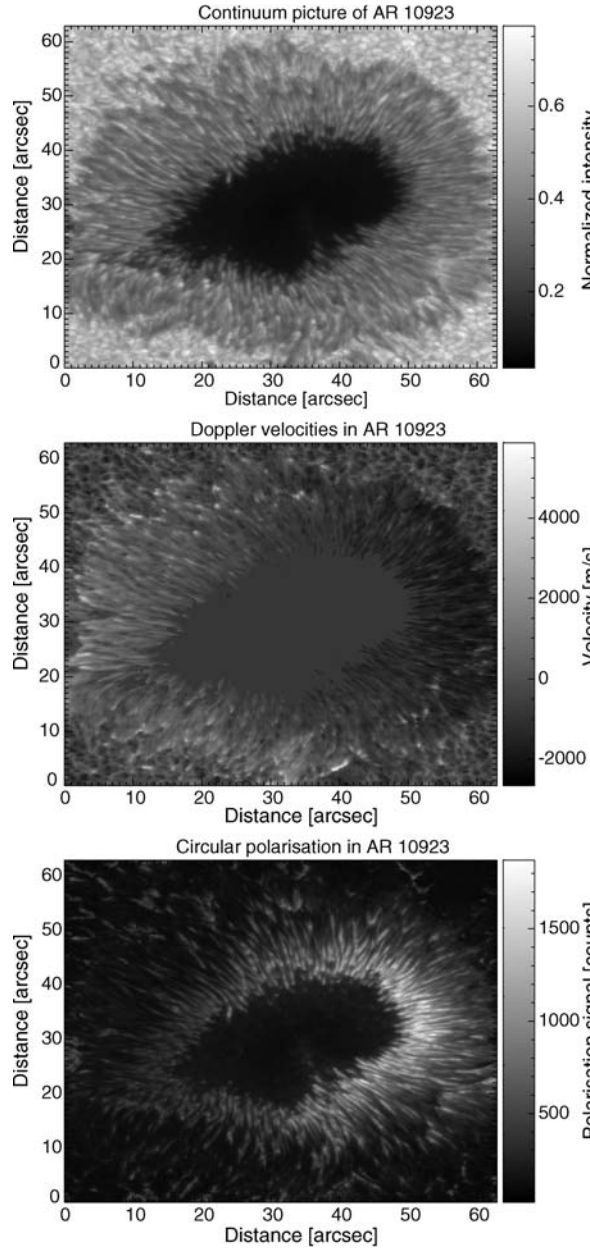


Fig. 1 Maps of intensity, LOS velocity, and circular polarization of sunspot (12 Nov 2006, $\theta = 30°$) from Fe I 630.2 nm taken with the spectropolarimeter SP attached to the SOT onboard Hinode

3.2.2 Evershed Flow, Uncombed Magnetic Field, and NCP

For understanding the nature of the penumbral fine structure, it is essential to know the topology of the velocity field and the magnetic field. The first attempt to measure the flow field was undertaken by Evershed in 1908 (see Evershed 1909) in order to test Hale's tornado theory of sunspots. Yet, instead of a circular flow, Evershed found a radial outflow of plasma, and until today we lack a consistent theory for sunspots. Before we discuss the progress in

modeling the characteristic feature of the penumbra, we discuss important observational aspects.

The Flow Field: With high spatial resolution, it is now established that the flow has a filamentary structure (Tritschler et al. 2004; Rimmele and Marino 2006). On average, the flow has a small upward component in the inner penumbra and a small downward component in the outer penumbra (Schlichenmaier and Schmidt 2000; Schmidt and Schlichenmaier 2000; Tritschler et al. 2004; Langhans et al. 2005). Recent observations have revealed that radially aligned up- and downflows exist on small scales next to each other (Sainz Dalda and Bellot Rubio 2008). Regarding the photospheric height at which the flow exists, there is convincing evidence that the flow is predominantly present in the very deep photosphere, i.e., beneath $\tau = 0.1$ (Maltby 1964; Schlichenmaier et al. 2004; Bellot Rubio et al. 2006). The flow velocities measured in the penumbra are substantially larger than what is measured in the granulation. Individual penumbral profiles exhibit line satellites that are Doppler shifted by up to 8 km/s (e.g., Wiehr 1995). From inversions, velocities well above 10 km/s have been found by del Toro Iniesta et al. (2001). Bellot Rubio et al. (2004) find an azimuthally averaged Evershed flow velocity of about 6.5 km/s, with local peaks of more than 10 km/s, based on two component inversions (see below). The small-scale flow field of dark cored bright filaments is discussed in the context of convective roll models (at the end of Sect. 4.1).

The Magnetic Field: Attempts to describe the magnetic field as being uniform along the line of sight are clearly inconsistent with the measured Stokes $Q(\lambda)$, $U(\lambda)$, and $V(\lambda)$ profiles (e.g., Westendorp Plaza et al. 2001a, 2001b). In particular, the penumbral V-profiles with 3 or more lobes cannot be explained by one component, even if unresolved Doppler-shifted components are assumed (Schlichenmaier and Collados 2002). Therefore, it was proposed that the magnetic field is interlocked or in other words uncombed (Solanki and Montavon 1993). In order to keep things as simple as possible, the magnetic field is assumed to have two components with different directions. Indeed, if the observed Stokes profiles with a spatial resolution of about 1 arcsec are interpreted with two components by means of inversions techniques, the fit to the observations is much better than with only one component (Bellot Rubio 2004; Bellot Rubio et al. 2003, 2004; Borrero et al. 2004, 2005; Beck 2008). Such inversions yield one less inclined magnetic component that is only slightly Doppler shifted, and a second magnetic field component that is somewhat weaker and more inclined, i.e., approximately horizontal. This second component carries the Evershed flow, with spatially averaged flow speeds of about 6.5 km/s.

These inversions also show that the magnetic field of the second component is aligned with the associated flow, pointing slightly upwards in the inner and slightly downward in the outer penumbra. The inclination of the first magnetic field component increases from some 30 degree at the umbral-penumbral boundary to some 60 degrees at the outer penumbral boundary. Inversions which are optimized to locate the width and height of the flow layer find that the flow is present in the very deep atmosphere, in the continuum forming layers (Bellot Rubio 2003; Borrero et al. 2006; Jurcak et al. 2007; Jurcak and Bellot Rubio 2008). Indeed, the width can hardly be determined, since the lower end of the flow layer is found to be beneath $\tau = 1$.

At 0.3 arcsec spatial resolution, spectropolarimetric measurements reveal that, at least in the inner penumbra, the more inclined magnetic component which carries the flow is associated with the dark cored bright filaments. Individual dark cores have a smaller degree of circular polarization than their lateral brightenings (Langhans et al. 2007). A thorough analysis shows that the latter statement is also true for the total polarization and that the

dark core magnetic field is weaker and more inclined than in the lateral brightenings (Bellot Rubio et al. 2007). Additionally, these studies confirm that the dark cores harbor strong Evershed flows.

The Magnetic Canopy: Outside the white-light boundary of the penumbra, the inclined magnetic field continues into the chromosphere, forming a magnetic chromospheric canopy in the surroundings of the sunspot, rising with distance to the spot up to a height of approximately 800 km (Solanki et al. 1992). In the canopy a radial outflow is present which is interpreted as the continuation of the Evershed flow (Solanki et al. 1992; Rezaei et al. 2006). However, it is estimated that only a few tenth of the flow mass is seen in the canopy. The rest of the penumbral Evershed flow must disappear within the penumbral downflow regions.

The Net Circular Polarization (NCP): The NCP, $\int V(\lambda)\,\mathrm{d}\lambda$, is a quantity that intimately links the flow and the magnetic field: NCP can only be non-zero, if and only if velocity gradients along the line of sight are present (e.g., Sanchez Almeida and Lites 1992). The magnitude and the size of the NCP depends on the gradient of the line of sight velocity, but also on the gradients in the magnetic field strength, inclination, and azimuth (Landolfi and Landi degl'Innocenti 1996; Müller et al. 2002, 2006; Borrero et al. 2008). A predominantly horizontal flow channel embedded in a less inclined background magnetic field successfully explains some symmetry properties of NCP maps of sunspots (Schlichenmaier et al. 2002) as well as some properties of the center to limb variation of NCP (Martínez Pillet 2000; Borrero et al. 2007).

Yet, some recent interpretations of NCP maps require that the flow component should be associated with stronger magnetic field (Tritschler et al. 2007; Ichimoto et al. 2008), rather than being associated with the same or weaker magnetic field in the flow channels, as we would expect from the models. Since there are also other indications for these stronger magnetic fields (e.g., Bellot Rubio 2003; Cabrera Solana et al. 2008; Borrero and Solanki 2008), the concept of embedded flow channels will need to be reviewed taking into account these new measurements.

Magnetized or Non-magnetized Flow: In terms of modeling the Evershed flow, it is crucial to know whether or not the flow is magnetized. While NCP can be generated by a field-free flow in a magnetized environment (e.g., Steiner 2000), the observed V profiles in certain locations in the penumbra show more than two lobes (e.g., Schlichenmaier and Collados 2002; Beck 2008). These additional lobes must be generated by an additional magnetic component: A non-magnetic component may produce a line asymmetry of Stokes-I and a non-zero NCP, but it cannot produce additional lobes in Stokes-V. For this it needs to be magnetized! And after all, the inversion results based on two components (see above) demonstrate that the Doppler-shifted "second" component is magnetized. Hence, we are convinced that any model for the penumbra needs to account for a magnetized Evershed flow.

4 Penumbral Models

The previous section stresses the point that the penumbra is a phenomenon of complex interaction of magneto-convective forces and radiation in a regime of inclined magnetic field of intermediate strength. One simplified view on this problem is to consider a separation

between convective plumes and a magnetic configuration as it is done in the field-free gap model. Another simplified view is by dealing with the problem in ideal MHD, in which the thin flux tube approximation is applicable. The latter perspective is taken in the siphon flow model and the dynamic extension of it, the moving tube model. Yet, for a full understanding it seems necessary to take into account dissipative magneto-convection driven by radiation. But we want to stress that simplified models often help to isolate the dominating physical processes, and to understand the essentials.

4.1 Convective Models

Originally proposed by Parker to explain the umbral dots, Spruit and Scharmer (2006) and Scharmer and Spruit (2006) extended the concept of the field-free gaps to explain the bright penumbral filaments and they realized that such a configuration may also produce the dark cores within bright filaments, caused by a subtle radiative effect at the top of the field-free gap. The idea of field-free gaps in the penumbra is that the inclined penumbral magnetic field produces bright elongations instead of dots. The gaps are supposed to be void of magnetic field and to be connected to the surrounding quiet Sun. Within the gaps, overturning convection transports ample amounts of heat which would account for the brightness of the penumbra. The convective flow field is directed upwards along the central lane of the filament and downward at the edges of the long sides of the filaments. Within the field-free gap there may exist a radial outflow that corresponds to the Evershed flow. The problem with this description is that the Evershed flow which is observed to be magnetized need to be non-magnetized in the field-free gap.

The field-free gap model is in many respects similar to the model of the convective rolls proposed in 1961 by Danielson (see also Grosser 1991 for a numerical investigation on this model). Convective rolls lie radially aligned next to each other. Two such rolls would form one filament as they rotate in opposite direction, producing an upflow in the central lane and a downflow at the lateral lanes. Danielson assumed that a horizontal magnetic field component would be associated with the rolls. This model has been discarded for two reasons: (1) There was no evidence for the corresponding convective flow field, and (2) a major fraction of the magnetic flux in the penumbra is directed upwards, and not horizontal. However, reason (1) depends on spatial resolution and the issue is not settled yet, as we cannot rule out the existence small amplitude vertical motions of a few hundred m/s. Reason (2) could be overcome by assuming that the rolls are separated by less inclined (more vertical) magnetic field lines, which constitute a more or less static background magnetic field. And, magnetized rolls interlaced by a static background field that is less inclined relative to the vertical would also meet the observational requirements of two magnetic components in the penumbra. In this respect, at least in principle, it is possible that the horizontal magnetic component carries an Evershed flow.

The problem here is that—up to now—there is only little support for downflows along the edges of bright filaments, although there is some indication for weak upflows within the dark cores (e.g., Zakharov et al. 2008). Rimmele (2008) does find a convective-roll like flow field in a filament that extends into the umbra for a sunspot close to disk center, while Bellot Rubio et al. (2005) did not find indications for up and down flows associated with a dark-cored bright filament at disk center. Zakharov et al. (2008) observe a very small downward velocity component. The latter authors argue that the downflow may be obscured by the upflows, and that convective rolls exist. Hence, the crucial question of vertical flows in the penumbra needs to be reconsidered. In order to minimize the effects of the horizontal radial outflow and of possible flows in azimuthal direction, sunspot observations at disk center are needed to learn about the presumably small vertical flow component.

4.2 Ideal Magneto-Convection

Another simplified view of the problem is by restricting oneself to ideal MHD. In the self-consistent magneto-static tripartite sunspot model of Jahn and Schmidt (1994) the surplus brightness of the penumbra relative to the umbra is produced by a heat transfer through the magnetopause, i.e. through the interface between the quiet Sun and the penumbra. This additional heat is thought to be distributed horizontally by interchange convection of magnetic flux tubes. The idea of dynamic magnetic flux tubes is compatible with the observationally finding of multiple magnetic components in the penumbra.

This motivated the study of the dynamics of a single thin magnetic flux tube as it evolves in a 2D static model background (Schlichenmaier et al. 1998a, 1998b). However, these studies did not confirm the concept of interchanging magnetic flux tubes which distribute heat horizontally. Instead, these studies created a new picture: The simulated tube lies along the magnetopause of the tripartite sunspot model and is taken to be a bundle of magnetic field lines with penumbral properties. Initially the tube is in magneto-static equilibrium. However, at a magnetopause that is sufficiently inclined, radiative heat exchange between the tube and the hotter quiet Sun triggers an instability: A thin magnetic flux tube that initially lies along the magnetopause, (a) feels the hotter quiet Sun, (b) heats up by radiation most effectively just beneath the photosphere, (c) expands, (d) rises through the subphotospheric convectively unstable stratification, and (e) develops an upflow along the tube, which brings hot subphotospheric plasma into the photosphere. (f) This hot upflow cools radiatively in the photosphere and streams radially outwards with supercritical velocity. The radiative cooling sustains the gas pressure gradient that drives the flow. (g) The outflow intrudes the convectively stable photosphere up to a height of some 50 to 100 km. The equilibrium height is determined by the balance of the diamagnetic force which pulls the conducting tube upwards toward decreasing magnetic field strength and the downward acting buoyancy which increases as the tube is being pulled up in a convectively stable stratification.

Weak Magnetic Field at Footpoint: The gas pressure gradient that drives the flow is caused by a surplus gas pressure building up inside the part of the tube that rises through the subphotospheric stratification. At the foot-point, i.e., the intersection of the tube with the transition layer from convectively unstable to stable, the gas pressure is high, and in order to balance the total pressure with the surroundings, the magnetic field strength is strongly decreased relative to the surroundings. In this sense the upflow foot-points can be considered as regions of weak magnetic field strength. In other words, the moving tube model is a magneto-convective mode which consists of a region of weak-field plasma that harbors hot up-flows and that travels inwards.

In principle the effect leading to the up and out flow works like an inverse convective collapse: In the classical convective collapse the plasma in the tube is cooled and a downflow occurs. Here, the heating of the plasma results in an upflow, and consequently the magnetic field strength in the tube decreases as the flow continues. In the photosphere, the gas pressure gradient is sustained by radiative cooling.

The moving tube scenario successfully explains a number of observational findings: (i) Penumbral grains are the photospheric footpoints of the tube, where the hot and bright plasma enters the photosphere. (ii) The upflow turns horizontally outwards in the photosphere and cools radiatively until it reaches temperature equilibrium. This determines the length of the penumbral grains. (iii) The footpoints migrate inward, as many observed penumbral grains do (e.g., Sobotka and Sütterlin 2001). (iv) The horizontal outflow corresponds to the Evershed flow. (v) The tube constitutes a flow channel being embedded in

a background magnetic field. This is in agreement with the uncombed penumbra, and produces realistic maps of NCP.

Magneto-Convective Overshoot: An interesting effect that can be studied with the idealized moving tube model, is related to overshooting (Schlichenmaier 2002, 2003). The upflow shoots into the convectively stable photosphere, and is turned horizontally by the magnetic curvature forces along the tube. The dominating forces here are the centrifugal force of the flow, $\kappa\rho v^2$ and the magnetic curvature force, $\kappa B^2/(4\pi)$, with κ being the curvature. In equilibrium v equals v_A, with v_A being the Alfvén velocity. During the evolution of the tube the velocity is roughly constant, but the magnetic field strength and hence v_A decreases leading to an overshoot, which creates an oscillation of the outflow around its equilibrium position such that the tube adopts a wave-like shape, i.e. the plasma first shoots up and then down, again passing the equilibrium position. Such a wave can be considered quasi-stationary, and the crest of such a wave can be compared with the properties of a Siphon flow (see below). Hence, the flow yields a serpentine shape, looking like a sea serpent, and evidence for such radially aligned up and down flows has been presented by Sainz Dalda and Bellot Rubio (2008). The amplitude of this wave increases as the magnetic field strength decreases, and eventually the downflow part dives in the sub-photosphere. There the stratification is convectively unstable and the magnetic flux tube experiences a dynamic evolution, that produces outward propagating waves. This scenario produces down-flows and makes the tube to disappear within the penumbra. Thereby it would solve a problem of the moving tube model: the out-flow would not extend into the surrounding canopy, but would disappear within the penumbra, as it is observed.

Serpentine Flow: Such a two-dimensional serpentine solution was criticized to be unstable in three dimensions (Thomas 2005), arguing that buoyancy forces make the wavy tube to fall over sideways. But this argument is not valid, since the influence of the upflow at the footpoint of the tube is not taken into account. At the footpoint the plasma is ejected upwards into the photosphere and due to conservation of momentum, the plasma overshoots and follows an up- and down wavy behavior. The fact that the density at the upper crest is larger than in the surroundings does not make the tube to fall over. As an analogy, one may think of a jet of water directed upwards with a garden hose. As long as the jet is pointing upwards with the hose (at the footpoint), the jet of water will not fall over. The jet of water will not fall over, even though the density of water is larger than that of the surrounding air. Since the footpoint of the up-flowing flux tube and its inclination is constrained, the boundary condition circumvents the wavy flow to fall over. Therefore the argument of Thomas (2005) is only true for a serpentine flow without a footpoint, and is not applicable here.

Siphon Flows: Siphon flow arches are stationary magnetic flux tube models, which were proposed to explain the Evershed flow (e.g., Meyer and Schmidt 1968; Thomas 1988; Degenhardt 1991; Thomas and Montesinos 1991). This class of models makes the ad hoc assumption of different magnetic field strengths at the two foot-points of a magnetic arch, which is responsible for a gas pressure gradient along the tube driving the flow. In the dynamic sea-serpent solutions (see above) a quasi-stationary solution exists (Schlichenmaier 2003). This solution corresponds to one (out of four) particular Siphon flow solution: a flow with a supercritical flow speed along the arch.

Heat Transport: Temporal measurements of the intensity evolution rule out the existence of interchange convection (Solanki and Rüedi 2003), and also, the numerical work of the

moving tube model did not confirm the concept of interchange convection of magnetic flux tubes as the heating mechanism for the surplus brightness of the penumbra: A crucial result of the numerical investigation is that a tube rises and develops an upflow, but the upflow does not stop nor does the tube sink back down to the magnetopause. Hence, instead of interchange convection the moving tube simulations suggests that the heating occurs in form of upflow channels along magnetic field lines. Ruiz Cobo and Bellot Rubio (2008) demonstrate that such an up-flow is capable to account for the brightness of the penumbra and that such up-flows can produce dark-cored bright filaments with a length of up to 3 Mm. Yet, even if such up flows can transport enough heat to account for the brightness of the penumbra, Schlichenmaier and Solanki (2003) have shown, that downflows within the penumbra are obligatory: There is not enough space for the magnetic flux associated with the up flows, such that down flows must remove the magnetic flux from the photosphere. In this respect, the overshoot scenario (serpentine flow) may help: the hot up-flow cools and the cool down-flow heats up in the hot sub-photosphere, and re-enters the photosphere as a hot upflow. Hence, the moving tube scenario encounters problems in accounting for sufficient heat transport, but there are ways to solve the heat transport problem with channeled flows. And these channeled flows are driven by radiative cooling.

4.3 Non-ideal Magneto-Hydrodynamics

Before we continue with well accepted model descriptions of the penumbra, we also want to mention an off-track approach by Kuhn and Morgan (2006) who argue that in the photosphere of a cool spot a large fraction of the plasma is neutral. A simplified consideration with two fluid components, one neutral and one ionized, yields an outward plasma flow driven by osmotic pressure. Whether or not such an effect and such a non-magnetic flow is realized on the Sun is not known, but it does not explain the observed Evershed flow, since the observed flow is magnetized.

4.4 Radiative Magneto-Convection

A better understanding of the sunspot penumbra is expected from numerical simulations of radiative magneto-convection in inclined magnetic fields. First results of such simulations (Heinemann et al. 2007; Scharmer et al. 2008; Rempel et al. 2008) consider 3D boxes solving for the full set of MHD equations including the (grey) radiative transport. These simulations consider a slice through the diameter of a round sunspot, including the umbra, the penumbra, and the surrounding quiet Sun. Assuming that the penumbral filament width is very small relative to the radius, the slice has a rectangular geometry with periodic boundary conditions in the horizontal directions. These simulations are still not able to produce a mature penumbra, but they succeed in reproducing single elongated filaments with lengths of up to a few Mm which resemble in many ways what is observed as thin light bridges and penumbral filaments of the inner penumbra.

The heating of these filaments does not occur by a single hot upflow channel, but rather in a form of a vertically elongated convective roll: a central lane of upflow, associated with two adjacent lanes of donwnflow. One convective cell has a vertical extension of some 500 km, while its lateral thickness is little less than 500 km. The vertical component of the magnetic field seems to become expelled by the convective flow, such that the convective cell is associated with more horizontal magnetic field. In this sense, these simulation results are a revival of the convective rolls proposed by Danielson in 1961. But in contrast to the Danielson rolls, the filaments in the simulations are interlaced with less inclined stronger magnetic field than in the filaments, and the rolls are elongated in depth.

The simulated penumbral filaments resemble light bridges with an inner upflow and two lateral downflows forming two apparent rolls as observed e.g. by Rimmele (2008). As the magnetic field becomes more inclined relative to the vertical, the upflow has an increasing horizontal outflow component. This horizontal outflow component increases with heigth, culminating in the photosphere. This horizontal flow is already present in the umbral dot simulations of Schüssler and Vögler (2006), but in an environment of more inclined magnetic fields this horizontal flow component becomes stronger. However, at this stage, the horizontal velocities in the simulations are only a little larger than the vertical velocities of the convective roll, while the state-of-the-art observations retrieve a horizontal velocity that is roughly a factor of 20 (10 km/s compared to <0.5 km/s) larger than the vertical velocity. In that sense the simulation fail to reproduce the Evershed effect, but there are indications that the simulated horizontal velocity component increases with more inclined magnetic field, and future simulations that may exhibit a fully developed penumbra and more inclined magnetic fields are expected to develop stronger horizontal flows, thereby reproducing the Evershed flow.

In the present MHD simulations the energy transport in umbral dots, light bridges, and filaments in the inner penumbra is accomplished by a magneto-convective mode, which may be characterized as convective elongated cells. Yet, these simulations do not exhibit a mature penumbra and the associated Evershed flow. It remains to be seen whether this magnetoconvective mode is also capable to reproduce a mature penumbra, or whether another magnetoconvective mode exists in the outer penumbra.

5 Conclusions

To explain different aspects of the penumbral properties two “simple” model classes have been proposed for the penumbra: (a) The moving magnetic flux tube models assumes ideal MHD, in which flows channeled by magnetic fields account for the filamentation, the Evershed flow, and the line asymmetries. (b) The gappy penumbra and convective rolls, which assume elongated convection cells to account for the surplus brightness of the penumbra. Neither of these simple models can account for all observational aspects: The moving tube scenario has problems to reproduce the overall down-flow in the outer penumbra and to account for all of the energy transport, while the elongated convective cells fail to produce an Evershed flow. The recent 3D box simulations of the full set of MHD equations show that magneto-convective heat transport may take place in elongated pancakes similar to what was proposed by Danielson. These simulations produce elongated convective cells which are associated with horizontal magnetic fields with weaker strengths, but in contrast to Danielson’s proposal the horizontal rolls with horizontal magnetic fields are embedded in stronger and more vertical magnetic field. This latter magnetic field component would form a more or less static background.

Considering the state-of-the-art of observational results and theoretical modeling, we conclude that there is evidence for both, a channeled flow with velocities of more than 5 km/s producing the Evershed flow, while the new 3D simulations suggest the existence of elongated convective rolls with much smaller up- and downflow velocities in the order 0.5 km/s. It may well be that the penumbra is a superposition of both, channeled flows above convective rolls, but at this point, we do not know. More advanced simulations will ultimately produce a fully developed penumbra, and by then it will be possible to understand how an ensemble of dynamic filaments is capable to form a stable penumbra, and how the penumbra is heated. Finally, the questions of sunspot stability and how sunspots form, evolve, and decay can be addressed.

Yet, at this point, the models are not fully consistent with observational facts. In particular it remains to be seen whether the flow pattern of convective rolls can be measured, and whether the observed penumbral line asymmetries in the Stokes parameters including the NCP can be reproduced by such models. Spectropolarimetric measurements need to have a spatial resolution of better than 0.1 arcsec to be comparable to the models.

Acknowledgements I am grateful to Oskar Steiner for many fruitful discussions, I like to thank Morten Franz for preparing Fig. 1, and Luis Bellot Rubio and Wolfgang Schmidt for valuable comments on the manuscript.

References

C. Beck, A 3D sunspot model derived from an inversion of spectropolarimetric observations and its implications for the penumbral heating. Astron. Astrophys. **480**, 825–838 (2008)

L.R. Bellot Rubio, *The Fine Structure of the Penumbra: from Observations to Realistic Physical Models*. ASP Conference Series, vol. 307 (2003), p. 301

L.R. Bellot Rubio, Sunspots as seen in polarized light. Rev. Mod. Astron. **17**, 21–50 (2004)

L.R. Bellot Rubio, H. Balthasar, M. Collados, R. Schlichenmaier, Field-aligned Evershed flows in the photosphere of a sunspot penumbra. Astron. Astrophys. **403**, L47–L50 (2003)

L.R. Bellot Rubio, H. Balthasar, M. Collados, Two magnetic components in sunspot penumbrae. Astron. Astrophys. **427**, 319–334 (2004)

L.R. Bellot Rubio, K. Langhans, R. Schlichenmaier, Multi-line spectroscopy of dark-cored penumbral filaments. Astron. Astrophys. **443**, L7–L10 (2005)

L.R. Bellot Rubio, R. Schlichenmaier, A. Tritschler, Two-dimensional spectroscopy of a sunspot. III. Thermal and kinematic structure of the penumbra at 0.5 arcsec resolution. Astron. Astrophys. **453**, 1117–1127 (2006)

L.R. Bellot Rubio, S. Tsuneta, K. Ichimoto, Y. Katsukawa, B.W. Lites, S. Nagata, T. Shimizu, R.A. Shine, Y. Suematsu, T.D. Tarbell, A.M. Title, J.C. del Toro Iniesta, Vector spectropolarimetry of dark-cored penumbral filaments with Hinode. Astrophys. J. Lett. **668**, L91–L94 (2007)

L. Bharti, C. Joshi, S.N.A. Jaaffrey, Observations of dark lanes in umbral fine structure from the Hinode solar optical telescope: Evidence for magnetoconvection. Astrophys. J. Lett. **669**, L57–L60 (2007)

J.M. Borrero, S.K. Solanki, L.R. Bellot Rubio, A. Lagg, S.K. Mathew, On the fine structure of sunspot penumbrae. I. A quantitative comparison of two semiempirical models with implications for the Evershed effect. Astron. Astrophys. **422**, 1093–1104 (2004)

J.M. Borrero, A. Lagg, S.K. Solanki, M. Collados, On the fine structure of sunspot penumbrae. II. The nature of the Evershed flow. Astron. Astrophys. **436**, 333–345 (2005)

J.M. Borrero, S.K. Solanki, A. Lagg, H. Socas-Navarro, B. Lites, On the fine structure of sunspot penumbrae. III. The vertical extension of penumbral filaments. Astron. Astrophys. **450**, 383–393 (2006)

J.M. Borrero, L.R. Bellot Rubio, D.A.N. Müller, Flux tubes as the origin of net circular polarization in sunspot penumbrae. Astrophys. J. **666**, L133–L136 (2007)

J.M. Borrero, B.W. Lites, S.K. Solanki, Evidence of magnetic field wrapping around penumbral filaments. Astron. Astrophys. **481**, L13–L16 (2008)

J.M. Borrero, S.K. Solanki, Are there field-free gaps near $\tau = 1$ in sunspot penumbrae? Astrophys. J. **687**, 668–677 (2008)

D. Cabrera Solana, L.R. Bellot Rubio, J.M. Borrero, J.-C. del Toro Iniesta, Temporal evolution of the Evershed flow in sunspots. II. Physical properties and nature of Evershed clouds. Astron. Astrophys. **477**, 273–283 (2008)

R.E. Danielson, The structure of sunspot penumbras. II. Theoretical. Astrophys. J. **134**, 289 (1961)

D. Degenhardt, Stationary fiphon flows in thin magnetic flux tubes. II. Astron. Astrophys. **248**, 637 (1991)

W. Deinzer, On the magneto-hydrostatic theory of sunspots. Astrophys. J. **141**, 548–563 (1965)

J.C. del Toro Iniesta, L.R. Bellot Rubio, M. Collados, Cold, supersonic Evershed downflows in a sunspot. Astrophys. J. Lett. **549**, L139–L142 (2001)

J. Evershed, Radial movement in sun-spots. Mon. Not. R. Astron. Soc. **69**, 454–457 (1909)

L. Gizon, Space. Sci. Rev. (2008, this issue)

H. Grosser, Zur Entstehung der Penumbra-Filamentierung von Sonnenflecken durch die Wirkung von Konvektionsrollen. PhD thesis, Universität Göttingen, 1991

T. Heinemann, Å. Nordlund, G.B. Scharmer, H.C. Spruit, MHD simulations of penumbra fine structure. Astrophys. J. **669**, 1390–1394 (2007)

K. Ichimoto, R.A. Shine, B. Lites, M. Kubo, T. Shimizu, Y. Suematsu, S. Tsuneta, Y. Katsukawa, T.D. Tarbell, A.M. Title, S. Nagata, T. Yokoyama, M. Shimojo, Fine-scale structures of the Evershed effect observed by the solar optical telescope aboard Hinode. Publ. Astron. Soc. Jpn. **59**, 593 (2007a)

K. Ichimoto, Y. Suematsu, S. Tsuneta, Y. Katsukawa, T. Shimizu, R.A. Shine, T.D. Tarbell, A.M. Title, B.W. Lites, M. Kubo, S. Nagata, Twisting motions of sunspot penumbral filaments. Science **318**, 1597 (2007b)

K. Ichimoto, S. Tsuneta, Y. Suematsu, Y. Katsukawa, T. Shimizu, B.W. Lites, M. Kubo, T.D. Tarbell, R.A. Shine, A.M. Title, S. Nagata, Net circular polarization of sunspots in high spatial resolution. Astron. Astrophys. **481**, L9–L12 (2008)

K. Jahn, Current sheet as a diagnostic for the subphotospheric structure of a SPOT. Astron. Astrophys. **222**, 264–292 (1989)

K. Jahn, in *Physical Models of Sunspots*, ed. by B. Schmieder, J.C. del Toro Iniesta, M. Vazquez. ASP Conf. Ser., vol. 118: 1st Advances in Solar Physics Euroconference. Advances in Physics of Sunspots, 1997, p. 122

K. Jahn, H.U. Schmidt, Thick penumbra in a magnetostatic sunspot model. Astron. Astrophys. **290**, 295–317 (1994)

J. Jurcak, L.R. Bellot Rubio, Penumbral models in the light of Hinode spectropolarimetric observations. Astron. Astrophys. **481**, L17–L20 (2008)

J. Jurcak, L.R. Bellot Rubio, K. Ichimoto, Y. Katsukawa, B.W. Lites, S. Nagata, T. Shimizu, Y. Suematsu, T.D. Tarbell, A.M. Title, S. Tsuneta, The analysis of penumbral fine structure using an advanced inversion technique. Publ. Astron. Soc. Jpn. **59**, S601–S606 (2007)

J.R. Kuhn, H. Morgan, *Osmotically Driven Neutral Sunspot Winds*, *Solar MHD Theory and Observations: A High Spatial Resolution Perspective*, ed. by J. Leibacher, R.F. Stein, H. Uitenbroek. Astronomical Society of the Pacific Conference Series, vol. 354, December 2006, pp. 230–+

M. Landolfi, E. Landi degl'Innocenti, Net circular polarization in magnetic spectral lines produced by velocity gradients: Some analytical results. Sol. Phys. **164**, 191–202 (1996)

K. Langhans, G.B. Scharmer, D. Kiselman, M.G. Löfdahl, T.E. Berger, Inclination of magnetic fields and flows in sunspot penumbrae. Astron. Astrophys. **436**, 1087–1101 (2005)

K. Langhans, G.B. Scharmer, D. Kiselman, M.G. Löfdahl, Observations of dark-cored filaments in sunspot penumbrae. Astron. Astrophys. **464**, 763–774 (2007)

V. Martínez Pillet, Spectral signature of uncombed penumbral magnetic fields. Astron. Astrophys. **361**, 734–742 (2000)

D.A.N. Müller, R. Schlichenmaier, O. Steiner, M. Stix, Spectral signatures of magnetic flux tubes in sunspot penumbrae. Astron. Astrophys. **393**, 305–319 (2002)

D.A.N. Müller, R. Schlichenmaier, G. Fritz, C. Beck, The multi-component field topology of sunspot penumbrae. A diagnostic tool for spectropolarimetric measurements. Astron. Astrophys. **460**, 925–933 (2006)

P. Maltby, On the velocity field in sunspots. Astrophys. Norvegica **8**, 205 (1964)

F. Meyer, H.U. Schmidt, Magnetisch ausgerichtete Strömungen zwischen Sonnenflecken. Zeitschrift für angewandte Mathematik und Mechanik **48**, T218 (1968)

F. Meyer, N.O. Weiss, H.U. Schmidt, The stability of sunspots. Mon. Not. R. Astron. Soc. **179**, 741 (1977)

E.N. Parker, Sunspots and the physics of magnetic flux tubes. IX Umbral dots and longitudinal overstability. Astrophys. J. **234**, 333–347 (1979)

V.J. Pizzo, Numerical modeling of solar magnetostatic structures bounded by current sheets. Astrophys. J. **365**, 764–777 (1990)

M. Rempel, M. Schüssler, M. Knölker, Radiative MHD simulation of sunspot structure. Astrophys. J. (2008, accepted)

R. Rezaei, R. Schlichenmaier, C. Beck, L.R. Bellot Rubio, The flow field in the sunspot canopy. Astron. Astrophys. **454**, 975–982 (2006)

T. Rimmele, Plasma flows observed in magnetic flux concentrations and sunspot fine structure using adaptive optics. Astrophys. J. **604**, 906–923 (2004)

T. Rimmele, On the relation between umbral dots, dark-cored filaments, and light bridges. Astrophys. J. **672**, 684–695 (2008)

T. Rimmele, J. Marino, The Evershed flow: Flow geometry and its temporal evolution. Astrophys. J. **646**, 593–604 (2006)

B. Ruiz Cobo, L.R. Bellot Rubio, Heat transfer in sunspot penumbrae. Origin of dark-cored penumbral filaments. Astron. Astrophys. **488**, 749–756 (2008)

A. Sainz Dalda, L.R. Bellot Rubio, Detection of sea-serpent field lines in sunspot penumbrae. Astron. Astrophys. **481**, L21–L24 (2008)
J. Sanchez Almeida, B.W. Lites, Observation and Interpretation of the asymmetric Stokes Q, U, and V line profiles in sunspots. Astrophys. J. **398**, 359–374 (1992)
G.B. Scharmer, H.C. Spruit, Magnetostatic penumbra models with field-free gaps. Astron. Astrophys. **460**, 605–615 (2006)
G.B. Scharmer, B.V. Gudiksen, D. Kiselman, M.G. Löfdahl, L.H.M.R. van der Voort, Dark cores in sunspot penumbral filaments. Nature **420**, 151 (2002)
G.B. Scharmer, T. Heinemann, A. Nordlund, Convection and the origin of Evershed flows in sunspot penumbrae. Astrophys. J. **677**, L149–L152 (2008)
R. Schlichenmaier, Penumbral fine structure: Theoretical understanding. Astron. Nachr. **323**, 303–308 (2002)
R. Schlichenmaier, *The Sunspot Penumbra: New Developments*, ed. by A. Pevtsov and H. Uitenbroek. ASP Conf. Ser., vol. 286: Current theoretical models and future high resolution solar observations, 2003, pp. 211–226
R. Schlichenmaier, M. Collados, Spectropolarimetry in a sunspot penumbra. Spatial dependence of stokes asymmetries in Fe I 1564.8 nm. Astron. Astrophys. **381**, 668–682 (2002)
R. Schlichenmaier, W. Schmidt, Flow geometry in a sunspot penumbra. Astron. Astrophys. **358**, 1122–1132 (2000)
R. Schlichenmaier, S.K. Solanki, On the heat transport in a sunspot penumbra. Astron. Astrophys. **411**, 257–262 (2003)
R. Schlichenmaier, K. Jahn, H.U. Schmidt, A dynamical model for the penumbral fine structure and the Evershed effect in sunspots. Astrophys. J. Lett. **493**, L121 (1998a)
R. Schlichenmaier, K. Jahn, H.U. Schmidt, Magnetic flux tubes evolving in sunspots. A model for the penumbral fine structure and the Evershed flow. Astron. Astrophys. **337**, 897–910 (1998b)
R. Schlichenmaier, D.A.N. Müller, O. Steiner, M. Stix, Net circular polarization of sunspot penumbrae. Symmetry breaking through anomalous dispersion. Astron. Astrophys. **381**, L77–L80 (2002)
R. Schlichenmaier, L.R. Bellot Rubio, A. Tritschler, Two-dimensional spectroscopy of a sunspot. II. Penumbral line asymmetries. Astron. Astrophys. **415**, 731–737 (2004)
W. Schmidt, R. Schlichenmaier, Small-scale flow field in a sunspot penumbra. Astron. Astrophys. **364**, 829–834 (2000)
M. Schüssler, M. Rempel, The dynamical disconnection of sunspots from their magnetic roots. Astron. Astrophys. **441**, 337–346 (2005)
M. Schüssler, A. Vögler, Magnetoconvection in a Sunspot Umbra. Astrophys. J. Lett. **641**, L73–L76 (2006)
M. Sobotka, P. Sütterlin, Fine structure in sunspots. IV. Penumbral grains in speckle reconstructed images. Astron. Astrophys. **380**, 714–715 (2001)
H. Socas-Navarro, V. Martinez Pillet, M. Sobotka, M. Vazquez, The thermal and magnetic structure of umbral dots form the inversion of high-resolution full Stokes Observations. Astrophys. J. **614**, 448–456 (2004)
S.K. Solanki, Sunspots: An overview. Astron. Astrophys. Rev. **11**, 153–286 (2003)
S.K. Solanki, C.A.P. Montavon, Uncombed fields as the source of the broad-band circular polarization of sunspots. Astron. Astrophys. **275**, 283 (1993)
S.K. Solanki, I. Rüedi, Spatial and temporal fluctuations in sunspots derived from MDI data. Astron. Astrophys. **411**, 249–256 (2003)
S.K. Solanki, I. Rueedi, W. Livingston, Infrared lines as probes of solar magnetic features. V – The magnetic structure of a simple sunspot and its canopy. Astron. Astrophys. **263**, 312 (1992)
H.C. Spruit, G.B. Scharmer, Fine structure, magnetic field and heating of sunspot penumbrae. Astron. Astrophys. **447**, 343–354 (2006)
O. Steiner, The formation of asymmetric Stokes V profiles in the presence of a magnetopause. Sol. Phys. **196**, 245–268 (2000)
P. Sütterlin, L.R. Bellot Rubio, R. Schlichenmaier, Asymmetrical appearance of dark-cored filaments in sunspot penumbrae. Astron. Astrophys. **424**, 1049–1053 (2004)
J.H. Thomas, Siphon flows in isolated magnetic flux tubes. Astrophys. J. **333**, 407–419 (1988)
J.H. Thomas, Flows along penumbral flux tubes in sunspots. Instability of super-Alfvénic, serpentine solutions. Astron. Astrophys. **440**, L29–L32 (2005)
J.H. Thomas, B. Montesinos, Siphon flows n isolated magnetic flux tubes. IV Critical flows with standing tube shocks. Astrophys. J. **375**, 404–413 (1991)
A.M. Title, Z. A Frank, R.A. Shine, T.D. Tarbell, K.P. Topka, G. Scharmer, W. Schmidt, On the magnetic and velocity field geometry of simple sunspots. Astrophys. J. **403**, 780–796 (1993)
A. Tritschler, R. Schlichenmaier, L.R. Bellot Rubio, the KAOS Team, Two-dimensional spectroscopy of a sunspot. I. Properties of the penumbral fine structure. Astron. Astrophys. **415**, 717–729 (2004)
A. Tritschler, D.A.N. Müller, R. Schlichenmaier, H.J. Hagenaar, Fine structure of the net circular polarization in a sunspot penumbra. Astrophys. J. Lett. **671**, L85–L88 (2007)

E. Wiehr, The origin of the Evershed asymmetry. Astron. Astrophys. **298**, L17–L20 (1995)
N.O. Weiss, Magnetic flux tubes and convection in the Sun. Mon. Not. R. Astron. Soc. **128**, 225 (1964)
C. Westendorp Plaza, J.C. del Toro Iniesta, B. Ruiz Cobo, V. Martínez Pillet, Optical tomography of a sunspot. III. Velocity stratification and the Evershed effect. Astrophys. J. **547**, 1148–1158 (2001a)
C. Westendorp Plaza, J.C. del Toro Iniesta, B. Ruiz Cobo, V. Martínez Pillet, B.W. Lites, A. Skumanich, Optical tomography of a sunspot. II. Vector magnetic field and temperature stratification. Astrophys. J. **547**, 1130–1147 (2001b)
V. Zakharov, J. Hirzberger, T.L. Riethmüller, S.K. Solanki, P. Kobel, Evidence of convective rolls in a sunspot penumbra. Astron. Astrophys. **488**(2), L17–L20 (2008)

Recent Evidence for Convection in Sunspot Penumbrae

Göran B. Scharmer

Originally published in the journal Space Science Reviews, Volume 144, Nos 1–4, 229–247.
DOI: 10.1007/s11214-008-9483-4

Abstract Whereas penumbral models during the last 15 years have been successful in explaining Evershed flows and magnetic field inclination variations in terms of flux tubes, the lack of contact between these models and a convective process needed to explain the penumbral radiative heat flux has been disturbing. We report on recent observational and theoretical evidence that challenge flux tube interpretations and conclude that the origin of penumbral filamentary structure is overturning convection.

Keywords Sunspots · Magnetic field

1 Introduction

Sunspot magnetic fields and dynamics have been studied scientifically for 100 years. Despite considerable progress during the last decade, a theoretical framework that explains sunspot fine structure, dynamics, magnetic fields and energy balance in a consistent manner is only now beginning to emerge. This situation can partly be attributed to the small horizontal scales associated with sunspot fine structure and the relatively poor spatial resolution achieved with spectropolarimetric observations. In addition, realistic numerical 3D MHD simulations of sunspots have only recently become possible.

During the last few years, there has been a remarkable improvement in the quality and diversity of observational data relevant to the understanding of sunspot fine structure, dynamics, magnetic fields and energy balance. In particular, high-spatial resolution observations from the Swedish 1-m Solar Telescope (SST) and the Solar Optical Telescope (SOT) on Hinode reveal new sunspot structure and flow patterns at odds with prevailing interpretations in terms of flux tube models. In addition, theoretical arguments as well as recent 3D MHD simulations of sunspot fine structure underline problems of these interpretations and lead to the conclusion that the origin of penumbral fine structure is overturning convection.

G.B. Scharmer (✉)
Institute for Solar Physics, Royal Swedish Academy of Sciences, AlbaNova University Center, 10691 Stockholm, Sweden
e-mail: scharmer@astro.su.se

M.J. Thompson et al. (eds.), *The Origin and Dynamics of Solar Magnetism*.
DOI: 10.1007/978-1-4419-0239-9_14

In the present review, we describe recent progress in our understanding of penumbral fine structure and put that in context with existing models. Rather than attempting to summarize the extensive literature on penumbrae, we discuss selected key papers and attempt to describe their interconnections and to critically review conclusions drawn. We also point out connections between observed penumbral fine structure and magnetic flux concentrations *outside* sunspots, such as faculae. We hope to convince the reader that the new picture emerging is one of improved consistency as regards observations and theory of sunspot penumbrae in particular, but also with respect to umbral dots, light bridges and faculae.

2 Overview of Established Models

For the past 15 years, the predominant paradigm of penumbral filaments has been based on nearly horizontal flux tubes, portrayed and modeled as radially aligned cylinders, embedded in a more vertical magnetic field. These flux tube interpretations have their roots in the work of Meyer and Schmidt (1968) who proposed to explain Evershed flows as siphon flows, originating from a difference in magnetic field strength between the two foot points of a flux tube. In the 70's, flux tubes and clusters of flux tubes were also established in models of magnetic flux concentrations, surrounded by field-free gas, with scales ranging from less than 100 km to that of a large sunspot.

2.1 Embedded Flux Tubes

Of particular importance in the current sunspot literature is the *uncombed penumbra* model, proposed by Solanki and Montavon (1993) to explain the strongly asymmetric Stokes V profiles observed on the limb side penumbra for sunspots away from disk center. This model addressed an apparent problem of very strong line-of-sight (LOS) *gradients* in the inclination angle of the penumbral magnetic field, inferred from Stokes data by Sanchez Almeida and Lites (1992). The large LOS inclination gradients derived from this data were (incorrectly, see Sect. 4) interpreted to imply volume currents and associated curvature forces strong enough to completely disrupt static force balance in the spot (Sanchez Almeida and Lites 1992; Solanki et al. 1993). The uncombed penumbra model avoids this problem, at least partly, by *postulating* the existence of discrete flux tubes, within which the magnetic field is assumed to be homogeneous and therefore current-free. The uncombed model thus 'replaces' smooth inclination gradients and (assumed) large volume currents with discontinuous changes at the boundary of the flux tube and an associated current sheet. Solanki and Montavon demonstrated that a nearly horizontal flux tube, with a strong flow parallel to its magnetic field, can explain the observed net circular polarization (NCP) resulting from Stokes V asymmetries of this configuration. Moreover, if such a flux tube is located *entirely* above the photosphere, both its upper and lower boundaries contribute with the same sign to the asymmetry of its Stokes V profile, thus enhancing the NCP.

Various implementations of flux tube models with polarized radiative transfer were later developed by e.g. Martínez Pillet (2000), Schlichenmaier and Collados (2002), Borrero et al. (2007), Borrero (2007), Tritschler et al. (2007), Bellot Rubio et al. (2003, 2004). These calculations demonstrate consistency between the calculated azimuthal variation of NCP and measurements made at low spatial resolution in visible and near infrared spectral lines. Two-component inversions interpreted within the context of embedded flux tube models and applied to low spatial resolution (0.6–1 arc sec) Stokes data by Borrero et al. (2004, 2005, 2006) similarly were shown to be largely consistent with the assumed inversion (flux tube) model.

2.2 Siphon Flow and Dynamic Flux Tube Models

The first flux tube model proposed to explain Evershed flows in penumbrae is the siphon flow model of Meyer and Schmidt (1968). In this model, a difference in field strength between the two footpoints of a flux tube leads to a difference in gas pressure, driving a flow in the direction of the footpoint with the highest field strength. This work was later followed up by Degenhardt (1989, 1991) and Thomas (1988) and the model further refined in a series of papers by Montesinos and Thomas (1989, 1993, 1997), Thomas and Montesinos (1990, 1991). Given the free parameters of the calculations, the siphon flow model of Montesinos and Thomas (1997) allow consistency with the discovery that the Evershed flow connects to patches of opposite magnetic polarity at deep layers near the outer boundary of a sunspot (Westendorp Plaza et al. 1997).

Siphon flow models allow an interpretation of Evershed flows as steady flows, with the properties of the flow determined by assumed conditions at the footpoints of the flux tube and of the surrounding atmosphere. The mechanism that *produces* the magnetic field strength difference between the footpoints, needed to generate a gas pressure gradient to drive the flow, is not explained by such models. Jahn and Schmidt (1994) proposed the concept of interchange convection of magnetic flux tubes (or rather, sheets) as an explanation of the penumbral heat flux. To investigate this proposal, Schlichenmaier et al. (1998a, 1998b) developed a simplified 1D, one-component numerical model of such a flux tube and studied its time evolution. In this model, a flux tube initially in contact with the magnetopause (the outer boundary of the sunspot) is heated radiatively by the external field-free gas. Its subsequent evolution is driven by the buoyancy of the flux tube and the superadiabatic stratification of the surrounding penumbra atmosphere, assumed to have properties unaffected by the flux tube. At the surface, radiative cooling of the tube causes it to loose buoyancy such that its upper part settles at a height of about 100 km above the photosphere. A gas pressure gradient, driving the Evershed flow, develops along the tube from downstream radiative cooling.

Later simulations by Schlichenmaier (2002, 2003) with reduced numerical viscosities show a similar initial behavior of the flux tube. However, near the surface, the flux tube subsequently develops standing waves downstream from the footpoint with downflows diving down into the convectively unstable layers beneath the surface. The crests of this oscillating flux tube remain visible above the surface and show an inward migration in the inner part of the umbra and an outward migration in the outer penumbra and outside the penumbra. This behavior is similar to that of observed penumbral grains in the inner and outer penumbra and moving magnetic features outside the penumbra. The discovery of small-scale bipolar magnetic features propagating from the mid penumbra to outside the penumbra, where they become moving magnetic features, is consistent with the 'sea serpent' behavior of Schlichenmaier's moving flux tubes (Sainz Dalda and Bellot Rubio 2008).

A problem, investigated by several authors, e.g., Solanki et al. (1993), Rezaei et al. (2006), is the large radial mass flux of the Evershed flow inside the penumbra. Only part of this flow appears to continue in the magnetic canopy above the quiet sun photosphere outside the spot. To explain this, most of the Evershed flow must submerge close to the outer boundary of the penumbra. The moving tube model simulations show such downflows within the penumbra. However, Thomas (2005, 2006) objected that the undulations seen in the 'sea serpents' of Schlichenmaier (2002) should occur preferentially in the *horizontal* plane and hence can explain neither moving grains nor convective downflows.

The question of a heating mechanism to explain the penumbral radiative heat flux was investigated by Schlichenmaier and Solanki (2003). Based on estimates of the radiative cool-

ing time and the time span of successive emergences of flux tubes, they concluded that interchange convection cannot provide the needed energy flux. Weiss et al. (2004) also argued against interchange convection on the basis that long loops of magnetic field connecting to a distant active region cannot possibly interchange with horizontal fields carrying Evershed flows.

The conclusion of Schlichenmaier and Solanki (2003) was that upflows along the magnetic flux tubes can explain the penumbral brightness, but only if the flux tube submerges again within a distance of 1000–2000 km from their footpoint. The upflow in a narrow tube cannot supply the radiative energy losses over a distance corresponding to the entire radial extent of a penumbra unless it submerges and is re-heated. Such re-heating does not solve the energy flux problem, however, since it relies on a (convective) mechanism to transport the heat to the bottom of the flux tube. Nevertheless, the discovery of field lines returning to the penumbra and associated downflows (Westendorp Plaza et al. 1997) was considered as support for this explanation (Schlichenmaier and Solanki 2003).

2.3 Convection and Downward Pumping of Magnetic Flux

The siphon flow model of Montesinos and Thomas is unrelated to any convection process operating in the penumbra. This model represents a stationary solution that cannot explain time dependent behavior such as moving penumbral grains (Thomas 2006). These grains are instead interpreted as originating from a moving convective pattern in the brighter parts of the penumbra (Weiss 2002, 2006a, 2006b). Whereas the moving tube simulations show localized downflows inside and outside the penumbra, the arched flux tubes of the siphon flow model require a mechanism to submerge and hold down their outer parts to sustain equilibrium (Montesinos and Thomas 1997). Thomas et al. (2002a) and Weiss et al. (2004) proposed that this submergence of the flux tubes occurs as the result of downward pumping by convection *outside* the sunspot. They even took this proposal one step further and proposed that this downward pumping is the *origin* of the filamentary structure of the penumbra. In this view, the salient features of penumbrae: their filamentary structures, the strong variations in magnetic field inclination across filaments and the Evershed flows, are to a large extent explained by what happens *outside* the sunspot. Magnetic fields in bright and dark filaments are distinct and cannot be interchanged (Thomas and Weiss 2004). To support this, Thomas and Weiss (2004), Weiss et al. (2004), Weiss (2006b) refer to X-ray observations and TRACE images showing loops extending over great distances across the Sun. We believe that their description and connection to the interlocking comb structure of the penumbra is misleading. Virtually all information about large fluctuations in the magnetic field inclination within the penumbra comes from spectral lines formed within a few hundred km above the photosphere. The images referred to (Sams et al. 1992) do not have the spatial resolution needed to separate X-ray loops (interpreted to outline field lines) from bright and dark filaments. As far as the author knows, there are no observations that allow us to conclude that azimuthal variations in the magnetic field inclination associated with filamentary structures correspond to field lines that are widely separated also far away from the penumbral photosphere. We have argued (Spruit and Scharmer 2006) that the strong inferred variations in the magnetic field inclination within the first one or two hundred km above the penumbral photosphere, seen even in the inner penumbra, cannot be explained by a mechanism operating outside the sunspot. Instead, these strong variations suggest a *local* mechanism at work. The bright and dark filaments are not distinct and they can interchange. Furthermore, Evershed flows are associated with field lines that only locally and during a limited time are nearly horizontal.

3 Limitations and Problems of Flux Tube Interpretations

The success of the uncombed penumbra model (Solanki and Montavon 1993) in explaining observed Stokes spectra and net circular polarization (NCP) (Borrero et al. 2007, Borrero 2007, Tritschler et al. 2007, Bellot Rubio et al. 2003, 2004), is unquestionable. Furthermore, the moving tube simulations of Schlichenmaier (2002) make excellent contact with the uncombed penumbra model. It demonstrates consistency with observed strong upflows in bright grains (Rimmele and Marino 2006) and the behavior of bright grains and moving magnetic features (Sainz Dalda and Bellot Rubio 2008). It is hardly surprising that the ability of these models to explain azimuthal and line-of-sight (LOS) gradients of the penumbral magnetic field and Evershed flows was deemed successful.

However, a fundamental problem remains. Whereas the embedded flux tubes simulated by Schlichenmaier are consistent with observations in many respects, such flux tubes present problems in explaining penumbral heating (Spruit and Scharmer 2006; Scharmer and Spruit 2006). As discussed above, a horizontal flux tube is likely to heat the penumbra over a radial distance not much more than 1000 km (Solanki and Montavon 1993; Schlichenmaier 2003), which is typical of a penumbral grain rather than a penumbral filament. Even over such a short distance, radiative cooling of the flow leads to significant temperature and brightness gradients along the flux tube unless there is a separate source of heating below the flow channel (Schlichenmaier et al. 1999). Spruit and Scharmer therefore argued that the presence of flux tubes covering a large fraction of the penumbral surface would constitute a *hindrance* for heating of the penumbra. They also pointed out that the existence of *elevated* flux tubes extending up to a few hundred km above the penumbral photosphere correspond to unlikely perturbations in a magnetic field so dominant already at this height that it must be expected to be nearly potential. Furthermore, the moving tube simulations represent a highly idealized model that cannot not be expected to be more than a coarse representation of reality. In particular:

- The existence of the flux tube is an *assumption* in the model.
- The model is 1-dimensional and the flux tube assumed to be 'thin' (see below).
- The simulations correspond to a 1-component model with the properties of the *background* atmosphere unaffected by the evolution of the flux tube.
- Only a single flux tube is simulated. The influence of neighboring flux tubes is not accounted for.
- The curvature forces of the surrounding magnetic field are ignored and its influence is reduced to a scalar magnetic pressure, similar to a gas pressure.

The same objections apply to the siphon flow models discussed in previous sections.

In view of these short-comings, it is remarkable that the moving tube simulations appear to capture important properties of penumbral dynamics. This is further discussed in Sect. 4.4.

Magnetostatic flux tube models including forces from a surrounding potential magnetic field (Borrero 2007) demonstrate the difficulties of embedded flux tube configurations. Prescribing a specific (circular) cross section for the flux tube corresponds to an overconstrained problem such that not only the gas pressure but also the temperature and density within the flux tube are given by force balance alone. There is thus no room for an energy equation with this type of models. Furthermore, equilibrium is not possible with a purely potential magnetic field inside the flux tube. In the models shown there is an azimuthal component, corresponding to a volume current aligned with the flux tube, in addition to the radial field component. The bottom part of the flux tube is nearly evacuated whereas the top part is denser than the surroundings in order to balance the magnetic forces at the top and bottom,

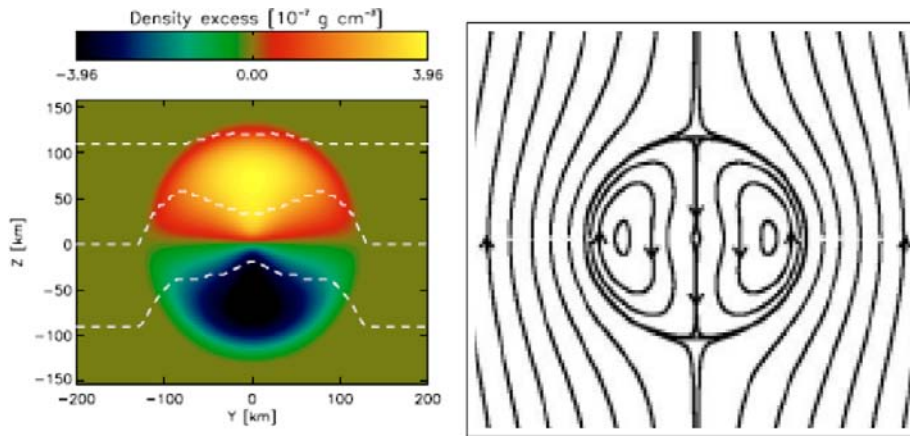


Fig. 1 *Left*: Magnetostatic flux tube model of Borrero (2007) with *dashed curves* showing (top to bottom) optical depths τ of 0.1,1 and 10. Note that only the upper part of the flux tube is located above the photosphere and that nearly the entire flux tube is located below $\tau = 0.1$. *Right*: Transverse field lines of a similar (but not identical) model (Borrero et al. 2007). Note the similarity of the magnetic field configuration *above* the flux tube in this model and that of the convective gap models, shown in Fig. 4

stretching and flattening the flux tube. These problems originate from the surrounding magnetic field wrapping around the flux tube and cannot be resolved by making the flux tube thinner.

In attempt to understand the temperature structure and energy balance of penumbral flux tubes, Ruiz Cobo and Bellot Rubio (2008) developed a model for a flux tube with a weak magnetic field aligned with a homogeneous magnetic field along the flux tube axis. An objection against this model is that it suffers from a lack of consistency as regards force balance, which is implemented in a way that is equivalent to ignoring the vector properties of the surrounding magnetic field. The origin of this problem is the same as that of the models of Borrero (2007): Prescribing the shape of the flux tube cross section is in general incompatible with either force balance or an energy equation.

3.1 Ambiguities of Interpretations Based on Inversions

A major obstacle to understanding penumbra fine structure has been the lack of adequate spatial resolution in observed polarized and unpolarized spectra. In spite of successful adaptive optics systems operating on major solar telescopes, it has not been possible to reach the diffraction limit with the long integration times needed for such data with adequate signal-to-noise. (However, by combining many *short* exposure frames and using image restoration techniques, filter-based systems allow near diffraction limited spectropolarimetry (van Noort and Rouppe van der Voort 2008; Scharmer et al. 2008a).) The exception is observations in the near infrared, for example those made with the Tenerife Infrared Polarimeter, TIP (Martínez Pillet et al. 1999) on the German Vacuum Tower Telescope (VTT) at wavelengths around 1.5 μm. At that wavelength, the diffraction limited resolution is about 0.6 arc sec with the VTT.

At a spatial resolution of 0.6 arc sec or worse, penumbral fine structure is not adequately resolved. To compensate for this and in order to test the validity of the embedded flux tube (uncombed penumbra) model, various observers, e.g., Borrero et al. (2004,

2006), Bellot Rubio et al. (2003, 2004) have implemented two-component inversion techniques to interpret Stokes spectra. These investigations show that it is indeed possible to fit the data with the assumed (highly idealized) representations of flux tubes. Forward calculations of Stokes spectra (Martínez Pillet 2000) and unpolarized spectra (Rimmele 1995; Bellot Rubio et al. 2006) based on flux tube models also demonstrate consistency. In some cases, it was demonstrated also that the observed data could equally well be reproduced with flux tube representations and models with smooth gradients (Martínez Pillet 2000; Rouppe van der Voort 2002; Borrero et al. 2004; Bellot Rubio et al. 2006). This ambiguity is a consequence of the width of the radiative transfer response function, smearing out the effects of discontinuities in the observed (Stokes) spectra. The interpretations of Stokes spectra clearly show compatibility with flux tube interpretations. However, the simplicity of the implemented inversion models and the use of two components to represent observational data of penumbral fine structure at inadequate spatial resolution adds to these uncertainties to the extent that we are justified in questioning whether a description in terms of embedded flux tubes is an adequate representation of penumbra fine structure.

4 Convective Origin of Penumbral Filaments

An alternative explanation to understanding penumbra fine structure was proposed by Spruit and Scharmer (2006). The filamentary structure is explained by convection in radially aligned (nearly) field-free gaps just below the visible surface. Such intrusions unavoidably lead to strong variations in the magnetic field strength and inclination above the gaps, but these variations are fully consistent with even a simple *potential* magnetic field configuration. The model explains dark cores in bright filaments (Scharmer et al. 2002), seen in the inner and mid penumbra (c.f., Figs. 2 and 3) as an indicator of strong field strength variations across filaments, leading to a strongly varying Wilson depression. The Evershed flow is in this model identical to the horizontal flow component of the convection (Scharmer et al. 2008b). We explain this model in more details in the following.

The large inclination gradients are a consequence of magnetic fields being divergence free ($\nabla \cdot \mathbf{B} = 0$): field lines cannot disappear at the top of a gap (or flux tube) but must bend around it. To estimate the characteristic vertical scale H for these inclination variations, we can assume a potential magnetic field (Spruit and Scharmer 2006). For filaments separated by a distance L, this gives $H \approx L/2\pi$. With a typical separation of $1''$ between filaments, this corresponds to a vertical height scale of 120 km. Simple magnetostatic models for such configurations, based on identical temperature variations with height for the two components, show distinct differences between the inner penumbra, where the magnetic field is more vertical and stronger than in the outer penumbra (Scharmer and Spruit 2006). In the *inner* penumbra, the magnetic field is cusp-shaped above the gaps and associated with a large ($\approx$200–300 km) Wilson depression relative to that of the gaps. Even when the temperature is the same inside and outside the gaps, the strong Wilson depression leads to an observed brightness that is lower above the gaps than between the gaps. The *dark-cored filaments* discovered with the Swedish 1-m Solar Telescope (SST) (Scharmer et al. 2002), seen in the inner and mid penumbra, are thus explained by a *combination of increased opacity associated with a strongly reduced field strength and an overall drop of temperature with height* (Spruit and Scharmer 2006). In the *outer* penumbra, the Wilson depression is only on the order of 50 km in these models. With such a small Wilson depression, intensity variations from 'global' vertical temperature gradients cannot be expected to completely dominate over *local* horizontal and vertical gradients associated with details of the heating and cooling of

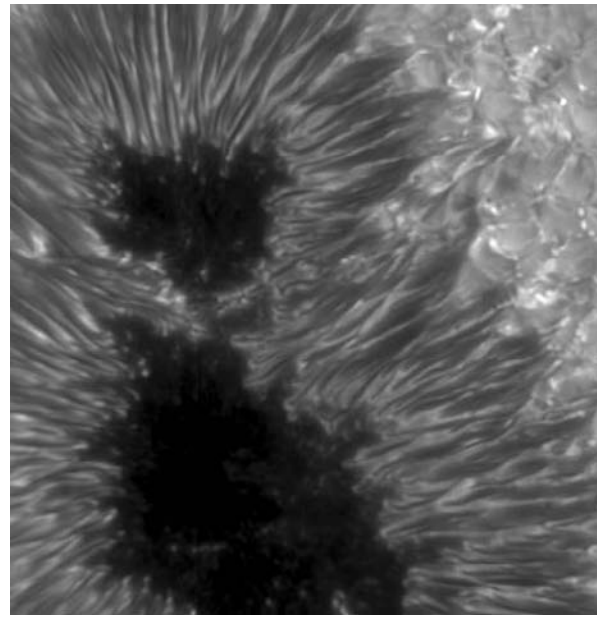


Fig. 2 Sunspot located at a heliocentric distance of 20 deg, observed with the SST on 12 Sep 2006 (van Noort and Rouppe van der Voort 2008). The image shows the Stokes I intensity, averaged over the blue and red wings of the 6302 iron line. Note the dark cores, clearly visible in the *inner* penumbra

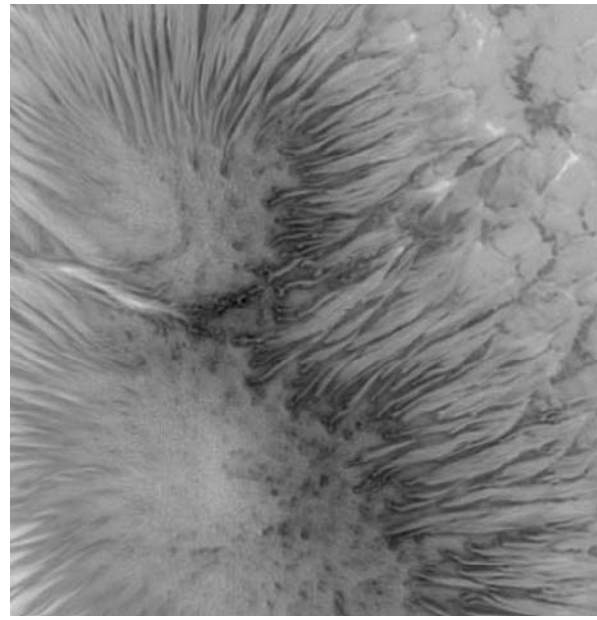


Fig. 3 The same sunspot as shown in Fig. 2 (van Noort and Rouppe van der Voort 2008). The image shows the difference between Stokes V (circularly polarized light), recorded in the blue and red wings of the 6302 iron line. This serves as a proxy for the line-of-sight component of the magnetic field vector. Note the strongly reduced polarization signal at locations of the dark cores, suggesting strongly reduced field strength at these locations

convecting gas. In the *outer* penumbra we therefore do not expect the same kind of relation between filament brightness and field strength as for the inner penumbra. The magnetostatic models are therefore in qualitative agreement with the absence of dark cores in the outer penumbra.

The potential magnetic field configurations associated with two magnetostatic models are shown in Fig. 4. The right figure shows the calculated field lines for a weak (1000 G) nearly

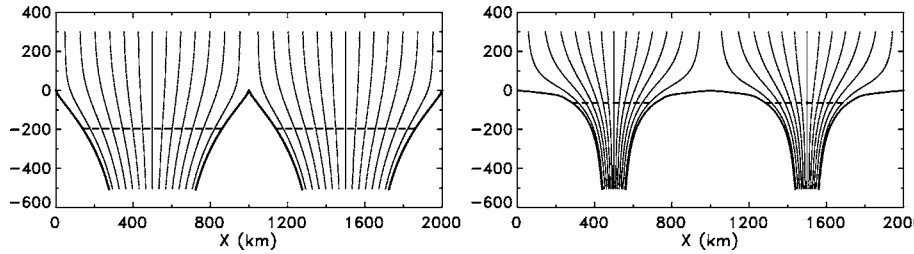


Fig. 4 Magnetostatic convective gap models (Scharmer and Spruit 2006) for the inner (*left*) and outer (*right*) penumbra. The *dashed curves* show the height at which the gas pressure in the magnetic component equals the photospheric gas pressure in the field-free component. This serves as proxy for the $\tau = 1$ surface and leads to Wilson depressions on the order of 200 km for the inner and 50 km for the outer penumbra. Note the similarity of the magnetic field configuration for the outer penumbra above the 'photosphere' in this model and the flux tube model, shown in Fig. 1

horizontal (average inclination 75 deg) magnetic field and the lower figure for a stronger (1800 G) and more vertical (average inclination 45 deg). Also shown are the shapes of the gap and the height at which the gas pressure between the gaps is equal to that of the field-free component at $z = 0$ (horizontal dashed lines). This serves as a proxy for the height at which the continuum optical depth is equal to unity. As is clear from the figure, the magnetic field configurations *above* the gaps are associated with strong gradients. The simple potential field model thus explains large magnetic field inclination variations *above* the penumbral photosphere without invoking forces in these layers. The associated current sheet is located at and below the photosphere, where the gas pressure is much higher than a few hundred km above the photosphere. This is in contrast to the uncombed model (Solanki and Montavon 1993), where the strong gradient in the magnetic field is a direct consequence of a *local* perturbation in the form av an embedded flux tube located *above* the photosphere. The current sheets associated with such flux tubes are difficult to combine with magnetostatic equilibrium because of the lower gas pressure at these heights (Spruit and Scharmer 2006). Indeed, most of flux tube in the magnetostatic model of Borrero (2007) is buried below the photosphere and only about 130 km protrudes above the surrounding photosphere. It seems very difficult, if not impossible, to construct similar models for flux tubes located entirely above the penumbral photosphere.

The convective gap model thus explains strong magnetic field gradients above the penumbral photosphere as a necessary consequence of potential fields. This model eliminates the problems of large curvature forces discussed by Sanchez Almeida and Lites (1992), Solanki et al. (1994), constituting a corner-stone argument in favor of the uncombed penumbra model (Solanki and Montavon 1993). The convective gap model also predicts configurations for the inner and outer penumbra that are quite different. In the inner penumbra, the field is cusp-shaped and associated with a large Wilson depression, large field strength fluctuations but relatively small fluctuations in inclination. In the outer penumbra, the magnetic field is spine-like (Lites et al. 1993), with a small Wilson depression, and small field strength fluctuations above the photosphere but with large inclination variations. These qualitative differences between the inner and outer penumbra are in good agreement with observations (Scharmer and Spruit 2006).

The overturning convective flow patterns associated with the gaps are predicted to be upward in the middle and downward along the boundaries to the magnetic components (Scharmer and Spruit 2006). Added to this flow pattern is a radially outward (Evershed) flow, explained by Scharmer et al. (2008b) on the basis of 3D MHD simulations (Heinemann

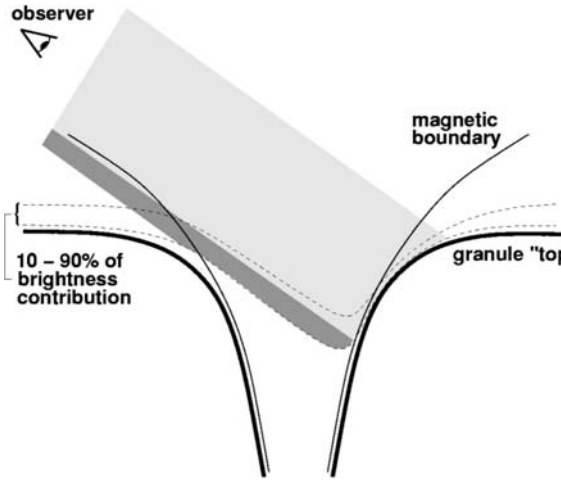


Fig. 5 Schematic drawing explaining the brightness of faculae observed near the limb (Keller et al. 2004). The low gas pressure associated with the strong magnetic field of the flux concentration makes it essentially transparent such that we can see the hot granular wall of the surrounding denser field-free gas. The optical depth unity surface is strongly tilted with respect to the horizontal by the shape of the flux tube magnetic field such that we see *deeper in to* the hot convecting granules close to the limb than at disk center, explaining the brightness. A similar situation occurs with penumbral filaments associated with strong variations in field strength (see Fig. 4). Another similarity between flux tubes and faculae is that the radiative cooling of the *surrounding* field–free gas through the facula leads to a convective flow pattern that is downward adjacent to the facula and that this downflow is observable on the limb-side but hidden from view on the center-side of the facula. In a similar way, strong Wilson depressions associated with narrow penumbral filaments lead to obscuration of the limb-sides of the filaments already for small heliocentric distance. Convection in such filaments are also affected by radiative cooling through the surrounding magnetic gas such that we expect upflows in the middle of the filaments and downflows at their sides

et al. 2007) as being identical to the horizontal component of this convection. In our model, the *dark* cores of the penumbral filaments correspond to locations of convective *upflows*, in contradiction with what we expect from field-free convection. As explained above, the strong fluctuations in field strength across the filamentary structures in the inner penumbra lead to a correlation between brightness and field strength, such that we see deeper in to the hotter gas where the field strength is high. The mechanism for producing the dark penumbral cores is directly related to the mechanism that produces bright faculae, proposed initially by Spruit (1976) and well established by numerical simulations (Keller et al. 2004; Carlsson et al. 2004; Steiner 2005) and high-resolution SST observations (Lites et al. 2004). In the case of faculae, convection occurs in the surrounding darker and field-free photosphere, whereas convection is inhibited or strongly suppressed in the brighter and magnetic faculae.

The connection between dark–cored filaments and faculae can be carried further, c.f., Fig. 5. The brightness of faculae near the limb is explained by the $\tau = 1$ surface being strongly inclined to the horizontal at the limbside of faculae (Spruit 1976; Keller et al. 2004; Carlsson et al. 2004). Looking through the nearly transparent gas within the strong magnetic field, we see deeper into the *surrounding* hot convecting gas close to the limb than at sun center. In a similar way, we see deeper into the convecting parts of the penumbral filaments when viewing sunspots well away from disk center and at ± 90 degrees from the symmetry line (Scharmer and Spruit 2006). The convective gap model leads us to interpret the absence of observational evidence for penumbral convection as not only due to lack of spatial resolution: there is also a difficulty of seeing deep enough into the filaments to be able to observe

the convection for sunspots close to disk center. In addition, there is the confusion from the brightness—field strength correlation already discussed above.

4.1 Limitations of the Convective Gap Model

The magnetostatic gap models discussed above (Scharmer and Spruit 2006) predict a gradual transition from cusp-shaped magnetic fields in the inner penumbra to spine-like (Lites et al. 1993) magnetic field configurations in the outer penumbra, in good agreement with observations. However, these simple models assume a perfectly field-free gap and do not include an energy equation, nor are the forces associated with the convective flows included. Details of flows and their interactions with the magnetic field cannot be explained with this simple model. The explanation of dark cores relies on opacity effects that are obvious only for the model corresponding to the inner penumbra. We expect these results to be relatively robust. In the outer penumbra, predictions about filamentary brightness are more difficult without an energy equation. Furthermore, observations show strong Evershed flows in the outer penumbra where the magnetic field is weaker. This combines to making the kinetic energy density $\rho v^2/2$ of comparable magnitude to the magnetic energy density $B^2/2\mu_0$ such that we expect relatively strong effects from the flow on the magnetic field. Whereas the magnetostatic gap models show good overall agreement with observed properties of penumbral magnetic fields, we cannot expect detailed agreement between the gap model and observations also in the outer penumbra. Only more accurate models and numerical simulations can provide this.

4.2 Support From Observations

The interpretation of light bridges as essentially field-free gaps dividing the umbra of a sunspot in two parts (Leka 1997; Jurčák et al. 2006) does not seem controversial. 3D MHD simulations of such structures (Nordlund 2006; Heinemann 2006) reproduce observed dark lanes running along the center of such structures (Lites et al. 2004) and demonstrate that the origin of this dark structure is the same as proposed for the convective gap model. Of considerable importance is therefore that dark-cored light bridge structures occasionally show smooth *transitions* to dark-cored penumbral filaments (Langhans 2006; Scharmer et al. 2007), strongly suggesting similar origin. Upflows in light-bridge dark lanes and the dark cores of penumbral filaments (Rimmele 2008) suggest a common interpretation in terms of convection, but with evidence for horizontal flows at greater heights also fitting a flux tube interpretation. Connections of dark-cored penumbral filaments to peripheral umbral dots and dark cores in light bridges have been reported also by Bharti et al. (2007b).

Several recent papers report evidence for convection in umbral dots (Bharti et al. 2007a; Rimmele 2008; Riethmüller et al. 2008). Although this provides no direct evidence for penumbral convection, the direct *connection* of peripheral umbral dots to dark-cored filaments (Langhans 2006; Langhans et al. 2007; Rimmele 2008) provides 'circumstantial' evidence for this interpretation. We caution, however, that the umbral dots observed are much larger than those simulated by Schüssler and Vögler (2006) and in some cases resemble granular intrusions.

In a paper of fundamental importance, Ichimoto et al. (2007) find strong evidence for overturning penumbral convection by analyzing continuum images and spectra observed with SOT on Hinode. Space-time plots created along lines crossing filaments in the *inner* penumbra, at directions ± 90 deg from the line connecting the center of the solar disk and the center of the sunspot, displayed twisted ropelike structures. The 'twisting' motion was

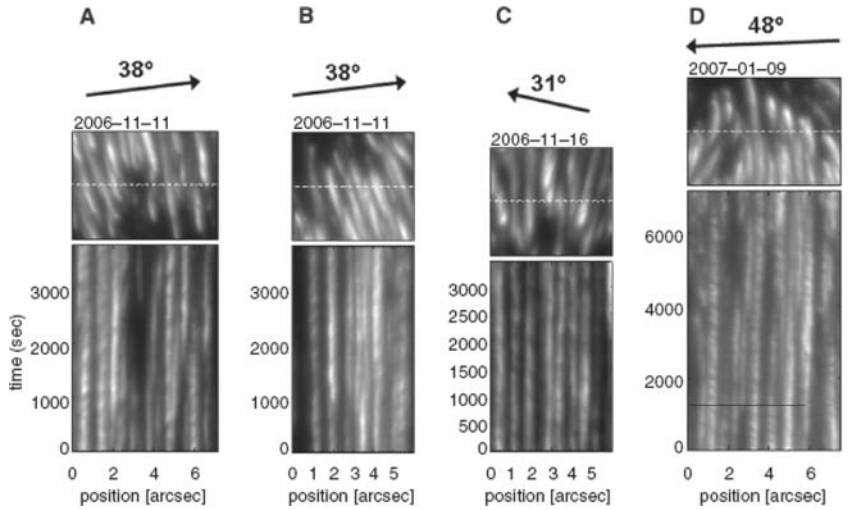


Fig. 6 Examples of space–time slices across filaments in the inner penumbra observed with SOT/Hinode (Ichimoto et al. 2007). The disk center direction is indicated with an *arrow* and labeled with the corresponding heliocentric distance. The *upper image of each pair* shows the surroundings of the cut across the filaments (*dashed line*) analysed in the space–time slices (*lower image*)

consistently in the direction toward sun center for both sides of the spot and irrespective of whether the spot was East or West of the meridian (c.f., Fig. 6). The apparent twist observed is therefore neither an actual twist nor a helical motion of individual filaments but must be a viewing angle effect. The interpretation (Ichimoto et al. 2007) is of upflows of overturning convection, viewed from the side. With the limb-side part of these filaments hidden from view, such flows will always appear to be in the direction of sun center direction for spots observed away from disk center. This is in perfect agreement with the predictions of the magnetostatic gap model for the *inner* penumbra (Scharmer and Spruit 2006). Here the Wilson depression between the gaps is predicted to be so large that the limbside part of the filaments is invisible already for disk center distances in the range 21–35 deg. The spot observed by Ichimoto et al. was located at a disk center distance of 31–48 deg. The flow pattern observed is consistent with that predicted by Scharmer and Spruit (2006).

Based on high-resolution SST data, Zakharov et al. (2008) recently inferred similar evidence of convective flows for a sunspot located 40 deg from disk center. Surprisingly, this is interpreted in terms of convective rolls (Danielson 1961). However, the horizontal flow toward the *center* of the filament at the *bottom* of such a roll, needed to verify its existence, is unobservable. Furthermore, the life times of penumbral filaments are on the order of 1 hour or more (Langhans et al. 2007). To sustain the radiative output over such a long time, the convective upflow must persist to depths much larger than a few hundred km. Finally, the observations of Ichimoto et al. (2007) were made in the *inner* penumbra, where the magnetic field has a strong vertical component, whereas roll-like convection is expected to be primarily associated with more horizontal magnetic field.

4.3 Support from 3D MHD Simulations

In contrast to what is to what is the case for faculae (Keller et al. 2004; Carlsson et al. 2004) and umbral dots (Schüssler and Vögler 2006), realistic simulations of entire sunspots have

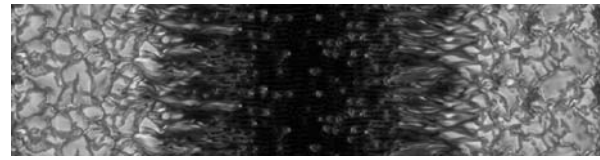


Fig. 7 Synthetic continuum image calculated from 3D MHD simulations, showing a sunspot and its surrounding photosphere (Rempel et al. 2008). In the umbral part of the spot is seen umbral dots with dark cores, in the penumbral part filaments that reach lengths of up to 2–3 Mm

not yet been feasible. This is partly due to the difficulties of thermally relaxing such a deep structure and maintaining its stability but mostly due to the huge range of scales associated with a fully developed sunspot. The first attempts to carry out 3D MHD simulations with radiative energy transfer of fine structure in a small sunspot were carried out by Heinemann et al. (2007) and further discussed by Scharmer et al. (2008b). Recently, simulations using a similar setup and grid separation, but with a much larger computational box, were carried out by Rempel et al. (2008). A synthetic continuum image calculated from these simulations is shown in Fig. 7. The approach taken in both simulations is to reduce the computational effort by using a rectangular computational box containing only a small 'azimuthal' slice of a sunspot. The overall results of these simulations obtained with two independent codes are quite similar, although differing strongly in the length of the penumbral filaments:

- The origin of filamentary structures is overturning convection and the dark cores are caused by a locally elevated $\tau = 1$ surface, supporting the convective gap model (Spruit and Scharmer 2006).
- The convection occurs in deep gaps, up to about 2 Mm (Rempel et al. 2008) with strongly reduced field strength.
- The simulations show horizontal outflows, similar to Evershed flows but with smaller velocities, peaking near optical depth unity and associated with locally strongly inclined fields.
- The bright heads of the penumbra filaments show inward propagation and strong upflows.
- The simulations show moving magnetic features (MMF's) and moat flow in the surrounding photosphere.

For overturning convection to be efficient, the gas needs to stay near the surface for a significant amount of time in order to give it time to cool radiatively. At the same time, it needs to move away from its upflow point in order to allow more gas to flow up. *Horizontal* flows are thus essential components of overturning convection. The Evershed flow is identified as being *identical* to the horizontal flow component of this penumbral convection (Scharmer et al. 2008b).

The simulations reproduce fundamental properties of observed penumbrae. This gives us confidence in concluding that the simulations constitute good representations of penumbral dynamics and energy balance. However, neither of these penumbra simulations show nearly horizontal magnetic fields in the outer penumbra, as observed, so something is missing in the simulations. Other features that do not appear realistic are elongated structures in the outer penumbra that appear intermediate to granules and filaments and a less distinct boundary to the umbra than observed. A complex question concerns the field strengths in the gaps. In the simulations these are on the order of 700–1000 G. In the simulations of Heinemann et al. (2007), this was interpreted as a consequence of numerical diffusivities, allowing overturning convective flows to cross field lines. Similar processes take place in the simulations of Rempel et al. (Rempel, private communication). The existence of strong magnetic fields

associated with overturning convection in penumbrae thus rely on turbulent magnetic diffusion at small scales. In contrast, umbral dots in the simulations of Rempel et al. (2008) as well as those of Schüssler and Vögler (2006) appear as nearly field–free plumes near the surface.

The simulations of Rempel et al. show some indication of flow patterns that are reminiscent of roll convection (Danielson 1961), but this interpretation is uncertain. The convective energy transport is mainly provided by *deep* upflow and downflow plumes. An intriguing result of both simulations is that the convective gaps do not extend to the field-free atmosphere below the sunspot, suggesting that sunspots radiate energy that is contained within the magnetic field initially. A difference between the simulations of Rempel et al. and Heinemann et al. is that the former simulations show radial *inflows* adjacent to the outflows near the surface whereas the latter simulations show evidence of such return flows only deeper below the surface.

A perhaps relevant result for understanding penumbral magnetic fields comes from simulations of quiet sun emerging flux. In these simulations, the expulsion of magnetic flux takes place in the horizontal direction by horizontal flows but also in the vertical direction by overshooting convection (Steiner et al. 2008). The weak horizontal fields of such emerging flux are expelled to heights of about 500 km, where they are difficult to observe in photospheric lines. Such expelled magnetic fields are limited to a height at which the gas pressure below the expelled magnetic field is roughly equal to $B^2/2\mu_0$. If a similar expulsion mechanism operates in the penumbra, where the field strength is much higher, any expelled horizontal field must be located much closer to the penumbral photosphere than for the emerging flux discussed by Steiner et al. (2008). Such magnetized gas would then show strong polarization signatures reminiscent of flux tubes.

4.4 Connections to Flux Tube Models

In spite of its 1D representation and the failure to confirm interchange convection, the moving tube model (Schlichenmaier et al. 1998a, 1998b; Schlichenmaier 2002) has connections to convective processes. The upflow *within* the flux tube is driven by the superadiabatic stratification of the external atmosphere, similar to that of a field-free convective upflow. The flux tube evolution is similarly driven by the superadiabatic stratification of the external atmosphere, even though the complete cycle of inward/outward movement and heating/cooling of the flux tubes does not take place in the simulations: a flux tube initially located along the magnetopause moves toward the umbra and remains there. These 1D simulations cannot show convective flow patterns. However, the 'sea serpent' behavior (Schlichenmaier 2002), with upflows and downflows along the length of the tube, can be interpreted as a 1D representation of 3D convection. In the 3D penumbra simulations (Heinemann et al. 2007; Rempel et al. 2008), the upflows occur at the centers of the gaps and the downflows on either side of the upflows, displaced both the in the azimuthal and radially outward directions. In the moving tube model, such upflows and downflows can only be spatially separated in the radial direction. Both models have in common a convective upflow and radiative cooling driving an outflow away from the center of the spot. In both moving tube and 3D penumbra simulations, the outflow peaks in a thin layer near $\tau = 1$, which is where the gas cools most efficiently (Scharmer et al. 2008b).

In the convective gap model, the strong magnetic field gradients above the gap are explained as a perturbation of a nearly potential magnetic field above the penumbra, introduced by the nearly field-free gap (Spruit and Scharmer 2006; Scharmer and Spruit 2006). This leads to a magnetic field that is cusp-shaped in the inner penumbra and locally nearly

horizontal in the outer penumbra. Adding a weak horizontal magnetic inside the gap will not change this configuration significantly. Such a configuration is in its *upper* parts quite similar to a flux tube but deeper down, these two types of structures are very different.

Support for the existence of embedded flux tubes based on a magnetostatic model is claimed from calculations of net circular polarization (NCP) (Borrero et al. 2007). However, the $\tau = 1$ surface of these models intersect the symmetry axis of the flux tube well above the center of the flux tube and the $\tau = 0.1$ surface (typical of the line formation height) cuts through the top of the flux tube (c.f., Fig. 1). These calculations clearly are sensitive only to the upper part of the flux tube, where its magnetic field is similar to that of the convective gap model.

A similar ambiguity concerns the origin of the dark cores of penumbral filaments, explained by convective gap models (Spruit and Scharmer 2006; Scharmer and Spruit 2006). Also flux tube models with *weaker* field in the flux tube than in the surroundings produce opacity effects resulting in dark-cored structures (Ruiz Cobo and Bellot Rubio 2008). This is interpreted as support for flux tube models by the authors. However, also for this model, only the top of the flux tube is visible above $\tau = 1$, so this configuration is similar to a convective gap model in its observable parts.

Interpretations of highly resolved Stokes spectra SOT/Hinode show wrapping around structures that can similarly be interpreted either as flux tubes or as convecting gaps (Borrero et al. 2008). These and other recent high-resolution Stokes data do not provide evidence for flow channels and flux tubes *elevated* above the photosphere, as discussed in some papers, e.g., Solanki and Montavon (1993), Rimmele (1995), Borrero et al. (2006). Based on observations of penumbra magnetic fields, interpretations in terms of convective gaps or flux tubes partly buried below the $\tau = 1$ surface thus are inherently ambiguous. This also serves as a reminder of the difficulties of interpreting (inadequately resolved) observations in terms of unknown underlying physics.

While this ambiguity, in our opinion, undermines arguments for the very existence of embedded flux tubes, it primarily suggests that magnetic field measurements are not likely to show a clear distinction between flux tube and convective gap models, at least in the outer parts of the penumbra. The distinction between the models may need to be based primarily on measurements of the velocity field, which is horizontal and along a flux tube but with added vertical and azimuthal flow components in the convective gap models. It is this diagnostics that so far provides the strongest observational evidence in favor of the convecting gap model (Ichimoto et al. 2007). The observational evidence accumulated so far is however too scarce to be conclusive. Zakharov et al. (2008) also reported evidence for convective upflows in penumbral filaments, but clearly observations of the entire sequence of upflows, horizontal flows and downflows are needed to fully reveal the nature of convection in penumbrae.

A highly controversial issue is whether the penumbral convection is essentially field-free or associated with kG strengths. Bellot Rubio et al. (2007), based on SOT/Hinode data and Milne–Eddington (ME) inversions, found only small variations in field strength across dark–cored penumbral filaments. Scharmer et al. (2008a), based on SST observations and ME inversions, demonstrated that an improvement of the spatial resolution from $0.3''$ to $0.15''$ increases measured field strength variations over dark cores by approximately a factor of two. Zakharov et al. (2008), based on SST data and ME inversions, found locally weaker fields by approximately a factor two, associated with convecting filaments. Jurcák et al. (2007) using SOT/Hinode data and inversions allowing for gradients and a Gaussian perturbation in the magnetic field to represent flux tubes, concluded that the field strength is reduced by only 600 G at the centers of bright filaments (where dark cores should be located) in the inner penumbra and furthermore that this reduction in field strength occurs only

close to the photosphere and disappears already at $\log \tau = -0.5$. Based on observations and inversions, there is so far no support for the assumption that these gaps are nearly field-free. This conclusion refers to observational data from the layers *above* the photosphere whereas the nature of penumbral convection in deeper layers can ultimately only be determined from 3D MHD simulations.

5 Conclusions

We believe that there is now strong evidence to support the conclusion that penumbral fine structure should be interpreted as the result of overturning convection, as proposed by Spruit and Scharmer (2006). Evidence for this conclusion comes from recent SOT/Hinode and SST observations (Ichimoto et al. 2007; Zakharov et al. 2008), showing *vertical* flows of the right magnitude to explain the penumbral radiative heat flux. Recent numerical 3D MHD simulations (Heinemann et al. 2007; Rempel et al. 2008) reproduce fundamental properties of observed penumbrae and confirm the convective origin of penumbral filaments. The simulations show that the nature of this convection takes place in gaps with up to 2 Mm depth and that any roll-like convection (Danielson 1961; Thomas and Weiss 2004; Rempel et al. 2008), if present, is of small importance (Rempel et al. 2008). The Evershed flow is interpreted to be identical to the horizontal flow component of this convection (Scharmer et al. 2008b). Such horizontal flows are necessary in order to cool hot upflows by radiation.

Neither observations nor simulations lead to the conclusion that this convection is nearly field-free, as suggested (Spruit and Scharmer 2006). However, inferred field strengths from spectropolarimetric data are obviously limited to layers above the photosphere, whereas simulations rely on numerical diffusivities to prevent instabilities at scales corresponding to the grid separation. Other uncertainties relate to the outer parts of penumbrae where observations show nearly horizontal field and even field lines dipping down into the photosphere (Borrero and Solanki 2008). Simulations do not show structures of this type. It appears likely that downward pumping of magnetic field by convection outside the sunspot plays a role in the outer penumbra, as proposed earlier (Thomas et al. 2002b), but we disagree strongly with the conclusion that this explains the origin of the filamentary structure of the penumbra. Downward pumping by the convection *inside* the penumbra also must take place and this probably explains why observations (Westendorp Plaza et al. 1997) show evidence of return flux well inside the outer penumbral boundary (Scharmer et al. 2008b).

Penumbral filaments have been successfully interpreted in terms of embedded flux tubes during a period of 15 years. While we conclude that this interpretation is misleading in terms of underlying physics, there are several reasons why this model has been so successful. We have shown that the opening of radially aligned gaps with nearly field-free convecting gas leads to a magnetic field that is much more horizontal over the gaps, giving the illusion of a flux tube (Spruit and Scharmer 2006). There are also other arguments for expecting nearly horizontal fields in the penumbral atmosphere: Horizontal cooling flows are most efficient near optical depth unity and if this gas is magnetized, it will aid in producing nearly horizontal magnetic fields. Also, emerging flux simulations relevant to the quiet sun suggests that convection can lead to flux expulsion in the vertical direction in addition to the horizontal direction and that this explains quiet sun horizontal magnetic fields above the photosphere. Clear evidence of such 'vertical' flux expulsion is however not seen in the penumbral part of the sunspot simulated by Rempel et al. (2008).

We emphasize the connections of the moving tube model to convective processes and to the radiative cooling of such flows near the photosphere in both types of models. The

similarity of the magnetic field above convecting gaps and flux tubes add to the difficulties of correctly interpreting observations and distinguishing between models.

In the embedded flux tube models, the Evershed flow is at center stage and the mechanism for heating the penumbra remains obscure. The new view of penumbral fine structure as caused by overturning convection implies that the main driver of penumbra fine structure is the energy flux below the surface and that the Evershed flow is 'only' a consequence of this convection (Scharmer et al. 2008b).

We believe that siphon flow models (Montesinos and Thomas 1997) are of little relevance for understanding penumbrae. These are linked to the idea that there are two distinct families of field lines: those associated with dark filaments and Evershed flows and those associated with bright filaments connecting to distant magnetic regions (Weiss et al. 2004; Thomas and Weiss 2004). As far as we know, there is no observational support for this 'static' picture of penumbral magnetic fields. Observations suggest life times for penumbral filaments on the order of one hour associated with flow channels opening and closing continuously (Rimmele and Marino 2006). 3D MHD Simulations (Rempel et al. 2008; Heinemann et al. 2007) confirm the transient nature of azimuthal variations in field strength and inclination. Theoretical arguments and models (Spruit and Scharmer 2006; Scharmer and Spruit 2006) as well as simulations clearly lead to the conclusion that the large variations in inclination across filaments are *local* perturbations, caused by *penumbral* convection and vanishing a few hundred km above the penumbral photosphere.

As regards further progress in this rapidly evolving field, we expect that even more realistic 3D MHD simulations in the near future will further improve our understanding of penumbrae, in particular as regards their outermost parts. Observed unpolarized and polarized spectra at the highest possible spatial resolution are needed. Of particular importance is such spectra giving information about the layer immediately above the photosphere. Emphasis should be given to analyzing data at ± 90 deg from the symmetry axis of sunspots located away from disk center, as was done by Ichimoto et al. (2007). This is in part to allow analysis of flows perpendicular to the radial direction of the filaments, but also in order to see as deep into these structures as possible. Analysis of such data need to account for the pronounced 3D nature of these filaments, caused by strong azimuthal variations in the Wilson depression, as well as strong LOS variations in the magnetic field and flow velocity. A dilemma here is that the use of inversion techniques with many nodes along the LOS raises questions of uniqueness and difficulties in comparing the results of such inversions with simulations. Existing 3D MHD simulation data allow inversion techniques to be tested with synthetic Stokes spectra from penumbral atmospheres, as done already with simulations of small-scale flux concentrations outside sunspots (Khomenko and Collados 2007; Orozco Suárez et al. 2007). The effect of assuming e.g. hydrostatic equilibrium can be evaluated quantitatively.

Presently used inversion techniques process polarized spectra pixel by pixel without constraining, for example, the magnetic field to be divergence-free. The requirement of divergence-free magnetic fields is crucial in forcing field lines to bend around convecting gaps (and flux tubes), leading to strong gradients in field strength and inclination. With spectropolarimetric observations approaching a spatial resolution of 100 km (Scharmer et al. 2008a), which is similar to the equivalent LOS resolution achieved with inversion techniques using a small number of nodes, it is reasonable to enforce magnetic fields constrained by $\mathrm{div}(\mathbf{B}) = 0$. Stray-light corrections are with most inversion techniques implemented in an ad-hoc manner pixel by pixel whereas a physical stray-light implementation would employ a point spread function that does not vary, or varies slowly, across the FOV. Micro- and macro-turbulence parameters are used as fudge parameters to compensate spatial smearing

of unresolved structures. With improved spatial resolution, modeled LOS velocity gradients should eliminate the need for such parameters in the inversions. We expect future inversion techniques to develop as 'global' techniques, in the sense of fitting model parameters for a large number of connected pixels simultaneously. This will allow constraints, such as $\mathrm{div}(\mathbf{B}) = 0$, and physical straylight models to be incorporated in a consistent manner to further enhance the usefulness of inversion techniques for inferring the physical state of the atmospheres above sunspots and other magnetic structures.

Acknowledgements The author gratefully acknowledges discussions with Henk Spruit, Åke Nordlund, Rolf Schlichenmaier, Matthias Rempel, Manfred Schüssler and Oskar Steiner. The author also thanks Rolf Schlichenmaier for valuable comments on the manuscript.

References

L.R. Bellot Rubio, H. Balthasar, M. Collados, R. Schlichenmaier, Astron. Astrophys. **403**, L47 (2003)
L.R. Bellot Rubio, H. Balthasar, M. Collados, Astron. Astrophys. **427**, 319 (2004)
L.R. Bellot Rubio, R. Schlichenmaier, A. Tritschler, Astron. Astrophys. **453**, 1117 (2006)
L.R. Bellot Rubio, S. Tsuneta, K. Ichimoto et al., Astrophys. J. Lett. **668**, L91 (2007)
L. Bharti, R. Jain, S.N.A. Jaaffrey, Astrophys. J. Lett. **665**, L79 (2007a)
L. Bharti, C. Joshi, S.N.A. Jaaffrey, Astrophys. J. Lett. **669**, L57 (2007b)
J.M. Borrero, Astron. Astrophys. **471**, 967 (2007)
J.M. Borrero, S.K. Solanki, ArXiv e-prints, 806 (2008)
J.M. Borrero, S.K. Solanki, L.R. Bellot Rubio, A. Lagg, S.K. Mathew, Astron. Astrophys. **422**, 1093 (2004)
J.M. Borrero, A. Lagg, S.K. Solanki, M. Collados, Astron. Astrophys. **436**, 333 (2005)
J.M. Borrero, S.K. Solanki, A. Lagg, H. Socas-Navarro, B. Lites, Astron. Astrophys. **450**, 383 (2006)
J.M. Borrero, L.R. Bellot Rubio, D.A.N. Müller, Astrophys. J. Lett. **666**, L133 (2007)
J.M. Borrero, B.W. Lites, S.K. Solanki, Astron. Astrophys. **481**, L13 (2008)
M. Carlsson, R.F. Stein, Å. Nordlund, G.B. Scharmer, Astrophys. J. Lett. **610**, L137 (2004)
R.E. Danielson, Astrophys. J. **134**, 289 (1961)
D. Degenhardt, Astron. Astrophys. **222**, 297 (1989)
D. Degenhardt, Astron. Astrophys. **248**, 637 (1991)
T. Heinemann, M.S. thesis, Univ. Copenhagen, 2006
T. Heinemann, Å. Nordlund, G.B. Scharmer, H.C. Spruit, Astrophys. J. **669**, 1390 (2007)
K. Ichimoto, Y. Suematsu, S. Tsuneta et al., Science **318**, 1597 (2007)
K. Jahn, H.U. Schmidt, Astron. Astrophys. **290**, 295 (1994)
J. Jurčák, V. Martínez Pillet, M. Sobotka, Astron. Astrophys. **453**, 1079 (2006)
J. Jurcák, L. Bellot Rubio, K. Ichimoto, et al., Publ. Astron. Soc. Jpn. **59**, 601 (2007)
C.U. Keller, M. Schüssler, A. Vögler, V. Zakharov, Astrophys. J. Lett. **607**, L59 (2004)
E. Khomenko, M. Collados, Memorie della Societa Astronomica Italiana **78**, 166 (2007)
K. Langhans, in *Astronomical Society of the Pacific Conference Series*, ed. by R. Casini, B.W. Lites. Astronomical Society of the Pacific Conference Series, vol. 358 (2006), pp. 3–+
K. Langhans, G.B. Scharmer, D. Kiselman, M.G. Löfdahl, Astron. Astrophys. **464**, 763 (2007)
K.D. Leka, Astrophys. J. **484**, 900 (1997)
B.W. Lites, D.F. Elmore, P. Seagraves, A.P. Skumanich, Astrophys. J. **418**, 928 (1993)
B.W. Lites, G.B. Scharmer, T.E. Berger, A.M. Title, Sol. Phys. **221**, 65 (2004)
V. Martínez Pillet, Astron. Astrophys. **361**, 734 (2000)
V. Martínez Pillet, M. Collados, L.R. Bellot Rubio, et al., in *Astronomische Gesellschaft Abstract Series*. ed. by R.E. Schielicke. Astronomische Gesellschaft Abstract Series, vol. 15 (1999), pp. 89–+
F. Meyer, H.U. Schmidt, Mitteilungen der Astronomischen Gesellschaft Hamburg **25**, 194 (1968)
B. Montesinos, J.H. Thomas, Astrophys. J. **337**, 977 (1989)
B. Montesinos, J.H. Thomas, Astrophys. J. **402**, 314 (1993)
B. Montesinos, J.H. Thomas, Nature **390**, 485 (1997)
Å. Nordlund, in *Solar MHD Theory and Observations: A High Spatial Resolution Perspective*, ed. by J. Leibacher, R.F. Stein, H. Uitenbroek. Astronomical Society of the Pacific Conference Series, vol. 354 (2006), pp. 353–+
D. Orozco Suárez, L.R. Bellot Rubio, J.C. del Toro Iniesta, Astrophys. J. Lett. **662**, L31 (2007)
M. Rempel, M. Schuessler, M. Knoelker (2008). ArXiv e-prints

R. Rezaei, R. Schlichenmaier, C. Beck, L.R. Bellot Rubio, Astron. Astrophys. **454**, 975 (2006)
T.L. Riethmüller, S.K. Solanki, A. Lagg, Astrophys. J. Lett. **678**, L157 (2008)
T.R. Rimmele, Astron. Astrophys. **298**, 260 (1995)
T. Rimmele, Astrophys. J. **672**, 684 (2008)
T. Rimmele, J. Marino, Astrophys. J. **646**, 593 (2006)
L.H.M. Rouppe van der Voort, Astron. Astrophys. **389**, 1020 (2002)
B. Ruiz Cobo, L.R. Bellot Rubio, Astron. Astrophys. **488**, 749 (2008)
A. Sainz Dalda, L.R. Bellot Rubio, Astron. Astrophys. **481**, L21 (2008)
B.J. Sams, III, L. Golub, N.O. Weiss, Astrophys. J., **399**, 313 (1992)
J. Sanchez Almeida, B.W. Lites, Astrophys. J. **398**, 359 (1992)
G.B. Scharmer, H.C. Spruit, Astron. Astrophys. **460**, 605 (2006)
G.B. Scharmer, B.V. Gudiksen, D. Kiselman, M.G. Löfdahl, L.H.M. Rouppe van der Voort, Nature **420**, 151 (2002)
G.B. Scharmer, K. Langhans, D. Kiselman, M.G. Löfdahl, in *New Solar Physics with Solar-B Mission*, ed. by K. Shibata, S. Nagata, T. Sakurai, Astronomical Society of the Pacific Conference Series, vol. 369, pp. 71–+ (2007)
G.B. Scharmer, G. Narayan, T. Hillberg et al. (2008a). ArXiv e-prints
G.B. Scharmer, Å. Nordlund, T. Heinemann, Astrophys. J. Lett. **677**, L149 (2008b)
R. Schlichenmaier, Astronomische Nachrichten **323**, 303 (2002)
R. Schlichenmaier, in *Current Theoretical Models and Future High Resolution Solar Observations: Preparing for ATST*, ed. by A.A. Pevtsov, H. Uitenbroek. Astronomical Society of the Pacific Conference Series, vol. 286 (2003), pp. 211–+
R. Schlichenmaier, M. Collados, Astron. Astrophys. **381**, 668 (2002)
R. Schlichenmaier, J.H.M.J. Bruls, M. Schüssler, Astron. Astrophys. **349**, 961 (1999)
R. Schlichenmaier, K. Jahn, H.U. Schmidt, Astrophys. J. Lett. **493**, L121+ (1998a)
R. Schlichenmaier, K. Jahn, H.U. Schmidt, Astron. Astrophys. **337**, 897 (1998b)
R. Schlichenmaier, S.K. Solanki, Astron. Astrophys. **411**, 257 (2003)
M. Schüssler, A. Vögler, Astrophys. J. Lett. **641**, L73 (2006)
S.K. Solanki, C.A.P. Montavon, Astron. Astrophys. **275**, 283 (1993)
S.K. Solanki, U. Walther, W. Livingston, Astron. Astrophys. **277**, 639 (1993)
S.K. Solanki, C.A.P. Montavon, W. Livingston, Astron. Astrophys. **283**, 221 (1994)
H.C. Spruit, Sol. Phys. **50**, 269 (1976)
H.C. Spruit, G.B. Scharmer, Astron. Astrophys. **447**, 343 (2006)
O. Steiner, Astron. Astrophys. **430**, 691 (2005)
O. Steiner, R. Rezaei, W. Schaffenberger, S. Wedemeyer-Böhm, Astrophys. J. Lett. **680**, L85 (2008)
J.H. Thomas, Astrophys. J. **333**, 407 (1988)
J.H. Thomas, Astron. Astrophys. **440**, L29 (2005)
J.H. Thomas, in *Solar MHD Theory and Observations: A High Spatial Resolution Perspective*, ed. by J. Leibacher, R.F. Stein, H. Uitenbroek. Astronomical Society of the Pacific Conference Series, vol. 354 (2006), pp. 224–+
J.H. Thomas, B. Montesinos, Astrophys. J. **359**, 550 (1990)
J.H. Thomas, B. Montesinos, Astrophys. J. **375**, 404 (1991)
J.H. Thomas, N.O. Weiss, Annu. Rev. Astron. Astrophys. **42**, 517 (2004)
J.H. Thomas, N.O. Weiss, S.M. Tobias, N.H. Brummell, Nature **420**, 390 (2002a)
J.H. Thomas, N.O. Weiss, S.M. Tobias, N.H. Brummell, Astronomische Nachrichten **323**, 383 (2002b)
A. Tritschler, D.A.N. Müller, R. Schlichenmaier, H.J. Hagenaar, Astrophys. J. Lett. **671**, L85 (2007)
M.J. van Noort, L.H.M. Rouppe van der Voort, Astron. Astrophys. **489**, 429 (2008)
N.O. Weiss, Astronomische Nachrichten **323**, 371 (2002)
N.O. Weiss, Space Sci. Rev. **124**, 13 (2006a)
N.O. Weiss, in *Solar MHD Theory and Observations: A High Spatial Resolution Perspective*, ed. by J. Leibacher, R.F. Stein, H. Uitenbroek. Astronomical Society of the Pacific Conference Series, vol. 354 (2006), pp. 213–+
N.O. Weiss, J.H. Thomas, N.H. Brummell, S.M. Tobias, Astrophys. J. **600**, 1073 (2004)
C. Westendorp Plaza, J.C. del Toro Iniesta, B. Ruiz Cobo et al., Nature **389**, 47 (1997)
V. Zakharov, J. Hirzberger, T.L. Riethmüller, S.K. Solanki, P. Kobel, Astron. Astrophys. **488**, L17 (2008)

Helioseismology of Sunspots: A Case Study of NOAA Region 9787

L. Gizon · H. Schunker · C.S. Baldner · S. Basu · A.C. Birch · R.S. Bogart · D.C. Braun · R. Cameron · T.L. Duvall Jr. · S.M. Hanasoge · J. Jackiewicz · M. Roth · T. Stahn · M.J. Thompson · S. Zharkov

Originally published in the journal Space Science Reviews, Volume 144, Nos 1–4, 249–273.
DOI: 10.1007/s11214-008-9466-5

Abstract Various methods of helioseismology are used to study the subsurface properties of the sunspot in NOAA Active Region 9787. This sunspot was chosen because it is axisymmetric, shows little evolution during 20–28 January 2002, and was observed continuously by the MDI/SOHO instrument. AR 9787 is visible on helioseismic maps of the farside of the Sun from 15 January, i.e. days before it crossed the East limb.

Oscillations have reduced amplitudes in the sunspot at all frequencies, whereas a region of enhanced acoustic power above 5.5 mHz (above the quiet-Sun acoustic cutoff) is seen outside the sunspot and the plage region. This enhanced acoustic power has been suggested to be caused by the conversion of acoustic waves into magneto-acoustic waves that are refracted back into the interior and re-emerge as acoustic waves in the quiet Sun. Observations show that the sunspot absorbs a significant fraction of the incoming p and f modes around 3 mHz. A numerical simulation of MHD wave propagation through a simple model of AR 9787 confirmed that wave absorption is likely to be due to the partial conversion of incoming waves into magneto-acoustic waves that propagate down the sunspot.

L. Gizon (✉) · H. Schunker · R. Cameron · J. Jackiewicz · M. Roth · T. Stahn
Max-Planck-Institut für Sonnensystemforschung, 37191 Katlenburg-Lindau, Germany
e-mail: gizon@mps.mpg.de

C.S. Baldner · S. Basu
Department of Astronomy, Yale University, PO Box 208101, New Haven, CT 06520, USA

A.C. Birch · D.C. Braun
Colorado Research Associates, A Division of NorthWest Research Associates, Inc., 3380 Mitchell Lane, Boulder, CO 80301-5410, USA

R.S. Bogart · S.M. Hanasoge
Hansen Experimental Physics Laboratory, Stanford University, Stanford, CA 94305, USA

T.L. Duvall Jr.
Laboratory for Solar Physics, NASA/Goddard Space Flight Center, Greenbelt, MD 20771, USA

M.J. Thompson · S. Zharkov
School of Mathematics and Statistics, University of Sheffield, Houndsfield Road, Sheffield, S3 7RH, UK

M.J. Thompson et al. (eds.), *The Origin and Dynamics of Solar Magnetism*.
DOI: 10.1007/978-1-4419-0239-9_15

Wave travel times and mode frequencies are affected by the sunspot. In most cases, wave packets that propagate through the sunspot have reduced travel times. At short travel distances, however, the sign of the travel-time shifts appears to depend sensitively on how the data are processed and, in particular, on filtering in frequency-wavenumber space. We carry out two linear inversions for wave speed: one using travel-times and phase-speed filters and the other one using mode frequencies from ring analysis. These two inversions give subsurface wave-speed profiles with opposite signs and different amplitudes.

The travel-time measurements also imply different subsurface flow patterns in the surface layer depending on the filtering procedure that is used. Current sensitivity kernels are unable to reconcile these measurements, perhaps because they rely on imperfect models of the power spectrum of solar oscillations. We present a linear inversion for flows of ridge-filtered travel times. This inversion shows a horizontal outflow in the upper 4 Mm that is consistent with the moat flow deduced from the surface motion of moving magnetic features.

From this study of AR 9787, we conclude that we are currently unable to provide a unified description of the subsurface structure and dynamics of the sunspot.

Keywords Sun · Sunspots · Helioseismology

1 Introduction

One of the main goals of solar physics is to understand the physical processes responsible for solar magnetism and activity. This requires the study of magnetic flux tubes, their transport and dynamics in the convection zone, and their emergence at the solar surface in the form of sunspots and active regions. The overall nature of sunspots is still a matter of debate. Many open questions remain concerning their structure and, above all, their formation and stability. How can regions of such intense magnetic flux come into existence and remain stable over several days, weeks, and sometimes months? Another common question is whether sunspots are monolithic magnetic flux tubes or have a spaghetti-like structure (Parker 1979). Schüssler and Rempel (2005) proposed a scenario whereby a sunspot expands rapidly below the surface during the early stages of its formation, leading to a disconnection from its magnetic roots. This disconnection may allow a transition to a spaghetti-like subsurface structure. As to the stability of sunspots, it may be due to the presence of surface and subsurface collar flows (Parker 1979). Other questions concern the energetics of sunspots, the flow of heat through and around sunspots, and the nature of magnetoconvection at kilogauss fields. What is known about sunspots has been summarized by, e.g., Thomas and Weiss (1992) and Solanki (2003). For a description of various magnetostatic sunspot models, we refer the reader to Jahn (1992) and Rempel et al. (2008).

In this paper we will discuss the potential of helioseismology to probe the subsurface structure of sunspots, with the hope of answering, one day, some of the questions listed above. Local helioseismology includes several methods of analysis, which have been described in some detail by Gizon and Birch (2005). All these methods rely on continuous time series of Doppler images of the Sun's surface. Fourier-Hankel analysis was developed to study the relationship between ingoing and outgoing waves around a sunspot (Braun et al. 1987; Braun 1995). Ring-diagram analysis consists of analysing the frequencies of solar acoustic waves over small patches of the solar surface (Hill 1988; Antia and Basu 2007). Time–distance helioseismology (TD) measures the travel times of wave packets moving through the solar interior (Duvall Jr. et al. 1993). Helioseismic holography (HH) uses the observed wave field at the solar surface to infer the wave field at different depths (Lindsey and Braun 1997). A summary of recent results is provided by Gizon (2006) and Thompson and Zharkov (2008).

There have been several studies of sunspots using helioseismology. Braun et al. (1987, 1992a) and Braun (1995) used Fourier-Hankel decomposition to measure wave absorption and scattering phase shifts caused by sunspots. The absorption is believed to be the result of a partial conversion of incoming p modes into slow magnetoacoustic waves (e.g., Spruit and Bogdan 1992; Cally 2000). Observational signatures of the mode conversion process have been discussed, for example, by Schunker and Cally (2006). Agreement between the observations of Braun (1995) and simplified sunspot models were reported by Fan et al. (1995), Cally et al. (2003), and Crouch et al. (2005) using a forward modeling approach. Time–distance helioseismology and helioseismic holography aim at making images of the solar interior from maps of travel times or phase shifts under the traditional assumption that the Sun is weakly inhomogeneous in the horizontal directions. TD and HH have been used to infer wave speed variations and flows in and around sunspots (e.g., Duvall Jr. et al. 1996; Jensen et al. 2001; Braun and Lindsey 2000; Gizon et al. 2000; Kosovichev et al. 2000; Zhao et al. 2001; Couvidat et al. 2006). Ring-diagram analysis has a coarser horizontal resolution and is used to study the subsurface structure of entire active regions (e.g., Basu et al. 2004; Antia and Basu 2007; Bogart et al. 2008). While these methods and their variants appear to be quite robust, it has not been demonstrated that they are consistent. For instance ring-diagram analysis has not been directly compared to time–distance or holography in the case of an isolated sunspot. This paper reports on a joint study of the sunspot in NOAA Active Region 9787.

2 Observations of NOAA Region 9787

2.1 MDI/SOHO Observations

NOAA Active Region 9787 was chosen from the MDI/SOHO data library because it hosts a large, round, isolated sunspot. A quick look at the data is given by Fig. 1. The data consist of nine days of MDI full disk Dopplergrams for each minute from 20 January to 28 January 2002. MDI also recorded the line-of-sight magnetic field every minute and intensity images every six hours. The images were remapped using Postel projection with a map scale of 0.12°. The centers of projection were chosen to track the motion of the sunspot (Carrington longitude $\phi \sim 133°$ and latitude $\lambda = -8.3°$). The re-mapping routine employs a cubic convolution interpolation. Missing data was linearly interpolated in time and a daily temporal mean was subtracted from each Dopplergram. Finally, we are left with one $512 \times 512 \times 1440$ data cube of Doppler velocity data for each day. These data sets are made available on the European Helio- and Asteroseismology Network (HELAS) web site at http://www.mps.mpg.de/projects/seismo/NA4/. All authors were invited to analyse the same data, thereby eliminating discrepancies in the data reduction methods.

Figure 1 shows a daily average of the MDI intensity continuum, magnetic field and Doppler velocity showing that there is little evolution of the sunspot during the period covered by the observations. The Dopplergrams show a ∼2 km/s Evershed outflow in the penumbra of the sunspot. The sunspot exhibits some amount of proper motion. Figure 2 shows the intensity profile of the sunspot averaged over nine days and over azimuthal angle, after correcting for the proper motion of the sunspot. The umbral and penumbral boundaries are at radii 9 Mm and 20 Mm respectively.

The sunspot is surrounded by a region of horizontal outflow called the moat flow. In order to characterise the strength and extent of the moat, we measured the motion of the moving magnetic features (MMFs) from hourly averages of the magnetograms using a local

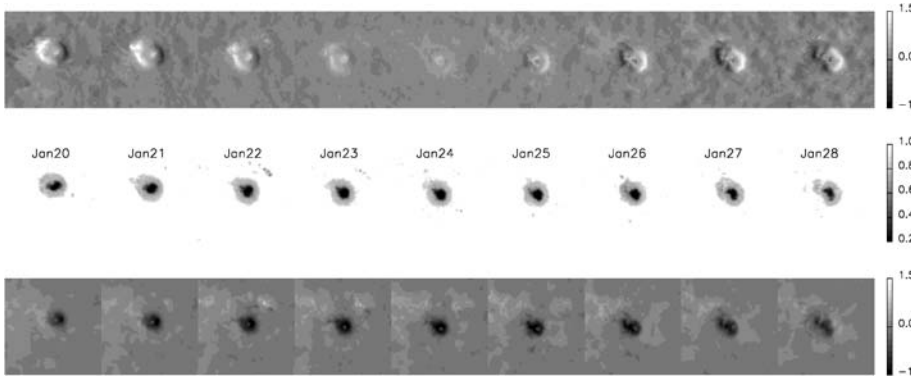


Fig. 1 Daily averages of the SOHO/MDI Doppler velocity (*top*), intensity (*middle*), and line-of-sight magnetic field (*bottom*) of the sunspot in Active Region 9787 during January 20–28, 2002. The Doppler velocity is in units of km/s, the magnetic field in units of kG. Each daily frame is a square with sides of length 200 Mm

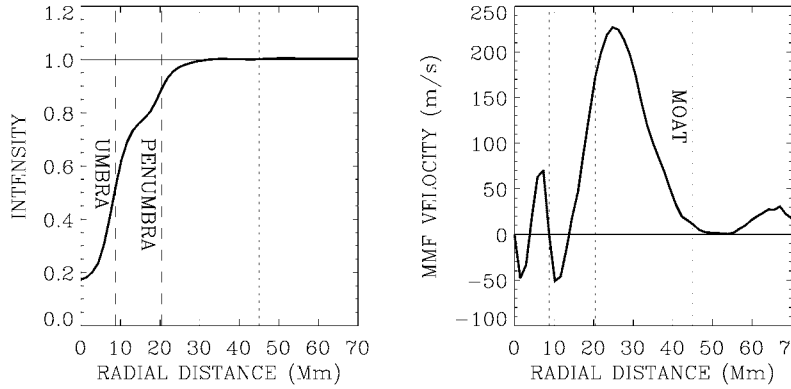


Fig. 2 (*Left*) Normalised intensity profile of the sunspot averaged over time and azimuth. The *vertical dashed lines* indicate the umbral and penumbral boundaries. The boundary of the moat is given by the *dotted line*. (*Right*) Velocity of moving magnetic features (MMFs) averaged over time and azimuth as a function of distance from the center of the sunspot. The MMFs track the moat flow and are moving outward from the outer penumbra up to a radius of about 45 Mm (*dashed line*). The *dotted vertical lines* indicate the boundaries of the umbra and the penumbra

correlation tracking method. The temporal and azimuthal averages of the MMF velocity is plotted in Fig. 2 as a function of distance from the center of the sunspot. The moat flow has a peak amplitude of 230 m/s and extends to about 45 Mm. The moat radius is about twice the penumbral radius, which is a standard value (Brickhouse and Labonte 1988).

The early development of Active Region 9787 can be traced using the helioseismic technique of farside imaging (Braun and Lindsey 2000; Lindsey and Braun 2000). Figure 3 shows four maps of the full Sun created by K. Oslund and P.H. Scherrer (2006). These maps were taken from the SOI website at http://soi.stanford.edu/data/full_farside/. The overlaid grid represents Carrington solar coordinates where vertical lines of longitude are separated by 60°. Active Region 9787 is located close to latitude $\lambda = -8.3°$ and longitude $\phi = 133°$, as shown by the red circles in Fig. 3. The active region is detected on the farside of the Sun in the top two panels of Fig. 3. We then see the active region ro-

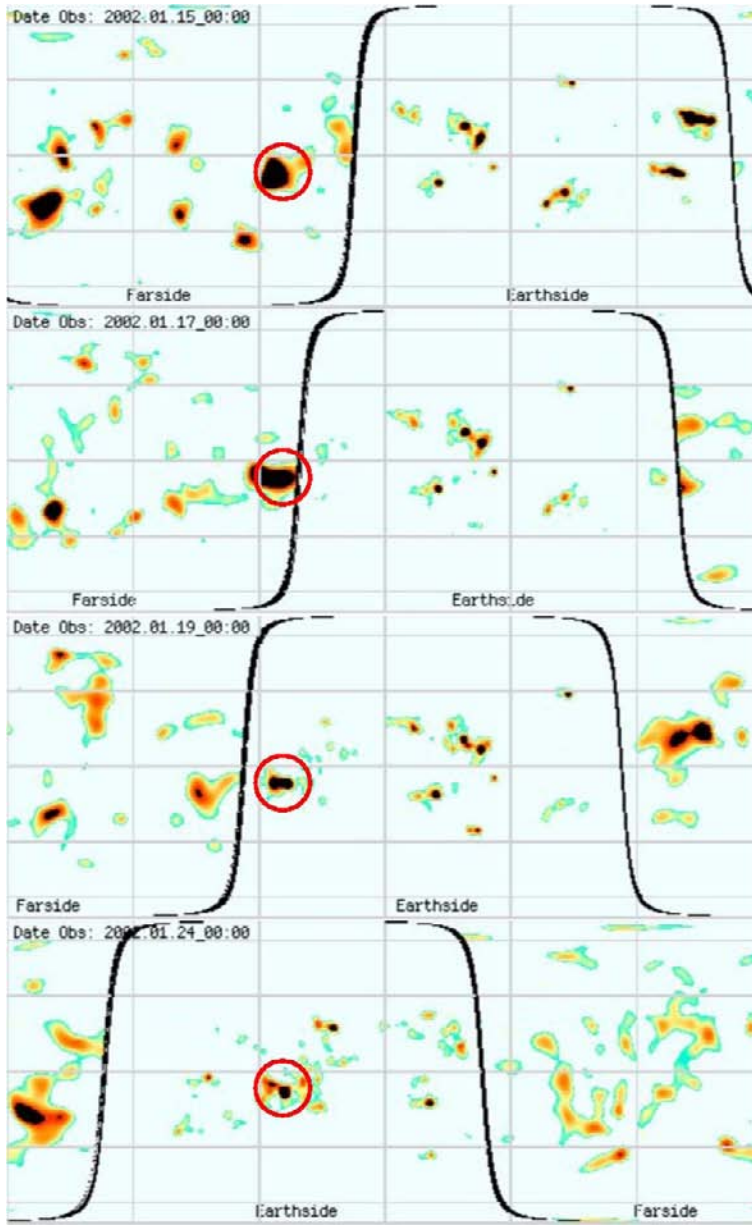


Fig. 3 Maps of the full Sun created using helioseismic waves to infer the presence of magnetic activity on the farside of the Sun. Strong shifts (*black*/*orange*) in the phase indicate an active region, the *light blue* represents the quiet Sun. The *horizontal lines* are lines of constant latitude, the *vertical lines* are lines of constant Carrington longitude separated by 60°. The dates from top to bottom are 2002 January 15, 17, 19, and 24. Active Region 9787 is located close to latitude $\lambda = -8.3°$ and longitude $\phi = 133°$ in all four maps, indicated by the *red circles*. (Courtesy of K. Oslund and P.H. Scherrer (2006))

tate past the East limb to the Earth side (third panel) and then within our observation period (24th January 2002) in the bottom panel. Additional information is available online at http://news-service.stanford.edu/pr/2006/pr-sun-031506.html.

2.2 Oscillatory Power and Acoustic Halos

There are two main properties of the acoustic power in and around active regions that are well documented. One is the power reduction in strong magnetic field regions, particularly sunspots, and the second is the enhancement of power in the higher frequencies (5–6 mHz) in the nearby photosphere, known as the acoustic halo (Braun et al. 1992b; Donea et al. 2000). Here we quantify these properties for AR 9787.

We calculate the temporal Fourier transform of the Doppler images for each day of observation. We divide this into 0.5 mHz bandwidths and calculate the power averaged over each of these frequency bandwidths. For all frequency bands the acoustic power suppression is greater than 80% in the umbra compared to the quiet Sun. We observe enhanced acoustic power at higher frequencies (5–6 mHz) in regions outside the sunspot and strong plage (Fig. 4).

Previous analysis of MDI acoustic power maps by Ladenkov et al. (2002) showed a modest excess of power around a sunspot in the higher frequencies, but which also appeared to be directly related to the location of plage, rather than the sunspot itself. Hindman and Brown (1998) also find that the high frequency velocity signal in an active region is higher (up to 60%) in pixels with moderate magnetic field strengths between 50 to 250 G. The fact that AR 9787 has an extended plage region offers an opportunity to analyse the plage far from the associated sunspot. In essence, this study bolsters the previous analysis of Hindman and Brown (1998) and Ladenkov et al. (2002).

When the sunspot is located close to the limb we find a significant enhancement of power in the umbra at high frequencies, to a level almost as high as in the quiet Sun. If this power is real and not an artifact of observing conditions at the limbs, then it could be due to magnetoacoustic waves with a component of motion perpendicular to the field lines in the umbra.

From Fig. 4, a region of particularly strong power close to the south-east side of the sunspot can be seen. This enhanced power is associated with the strong plage region just to

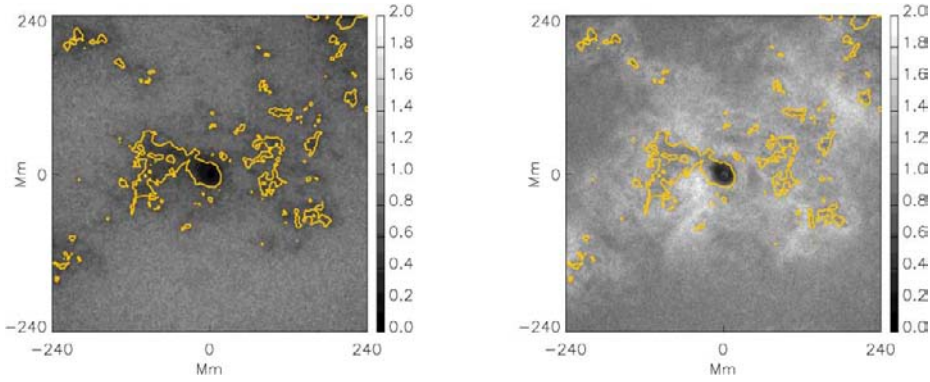


Fig. 4 Acoustic power averaged over all days and frequency bands 3–3.5 mHz (*left*) and 5.5–6 mHz (*right*). Overplotted contour is for $B = 100$ G, outlining regions of plage. The excess power outside the plage regions is clearly seen in the high-frequency maps (*right*). The power is normalised to unity in the quiet Sun (*grey scale*)

the North. Braun (1995) finds evidence of an acoustic deficit immediately outside a small (8 Mm) sunspot, extending out to 35 Mm and appearing to be well defined by the location of surrounding plage. The sunspot in AR 9787 shows little evidence for a well defined, axisymmetric acoustic halo, leading to the suggestion that the enhanced acoustic power is associated with the strong surrounding plage regions, rather than the sunspot itself.

The results of recent numerical work by Hanasoge (2008) seem to reproduce the high frequency halo surrounding a small model sunspot. The halo appears between 5 and 6 mHz, close to the acoustic cut-off. In agreement with the suggestion of Donea et al. (2000), the upwardly propagating fast mode waves may be reflected in magnetic regions due to the rapidly increasing Alfvén velocity. These waves later re-emerge in the region surrounding the sunspot causing the observed enhanced power.

2.3 Wave Absorption

A useful analysis procedure for studying the interaction of p modes with sunspots is the decomposition of solar oscillations, observed in an annular region around the sunspot, into inward and outward propagating waves. Fourier-Hankel spectral decomposition has been used to identify p-mode absorption in sunspots and active regions by comparing the amplitudes of the outward and inward moving waves (Braun et al. 1988; Bogdan et al. 1993; Braun 1995).

Here we use a sunspot-centered spherical-polar coordinate system (θ, ϕ) with the sunspot axis at $\theta = 0$. The annular region is defined by the inner and outer circles at $\theta_{\min} = 2.5^\circ$ and $\theta_{\max} = 11.25^\circ$, which correspond to distances between 30 and 137 Mm. The Doppler signal $\psi(\theta, \phi, t)$ in the annular region is decomposed into components of the form

$$\mathrm{e}^{i(m\phi + 2\pi\nu t)} \left[A_m(l, \nu) H_m^{(1)}(l\theta) + B_m(l, \nu) H_m^{(2)}(l\theta) \right], \tag{1}$$

where m is the azimuthal order, l is the harmonic degree, $H_m^{(1)}$ and $H_m^{(2)}$ are Hankel functions of the first and second kinds, t is time, ν is temporal frequency, and A_m and B_m are the complex amplitudes of the incoming and outgoing waves respectively. The range in l is between 70 and 1500. The boundaries of the annulus, $\theta_{\min}$ and $\theta_{\max}$, were selected such as to resolve the low-order p mode ridges with a resolution in l of approximately 40. The numerical procedure needed to compute the wave amplitudes $A_m(l, \nu)$ and $B_m(l, \nu)$ is described by Braun et al. (1988).

For each value of l we measure the mode amplitudes for the azimuthal order $m = 0$. The power spectra of the incoming and outgoing modes are displayed in Fig. 5. The outgoing p-mode power appears to be significantly reduced compared to the incoming p-mode power.

To measure this more quantitatively we determine an absorption coefficient for the f and p_n ridges in a frequency band between 2.9 and 3.1 mHz. The absorption coefficient is defined as

$$\alpha_n = \frac{\int \mathrm{d}l \, \mathrm{d}\nu \; W_n (P_{\mathrm{in}} - P_{\mathrm{out}})}{\int \mathrm{d}l \, \mathrm{d}\nu \; W_n P_{\mathrm{in}}}, \tag{2}$$

where W_n is a window function that selects the n-th ridge, and $P_{\mathrm{in}}(l, \nu)$ and $P_{\mathrm{out}}(l, \nu)$ are the power of the ingoing and outgoing waves. In this frequency band 2.9–3.1 mHz, the f, p_1, p_2, p_3, and p_4 ridges show absorption coefficients of 57%, 54%, 51% , 49%, and 50% respectively. These values are in agreement with those given in earlier studies (Braun et al. 1988; Braun 1995) and confirm that AR 9787 "absorbs" acoustic waves.

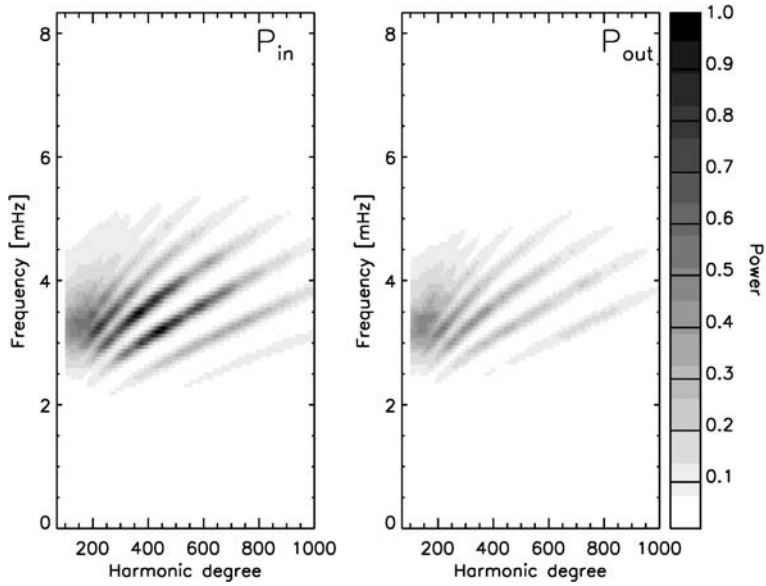


Fig. 5 Power spectra of inward (*left*) and outward (*right*) propagating waves as a function of harmonic degree and frequency. The reduction in power of the outgoing modes is noticeable

3 Travel Time Measurements

3.1 Phase-Speed Filtering versus Ridge Filtering

The principle of Helioseismic holography (HH) is to computationally regress the acoustic amplitudes observed at the surface into the solar interior (Lindsey and Braun 1997). To facilitate comparisons with results from time–distance analyses, we use *surface-focused* HH. In the "space–frequency" domain, i.e. where $\psi(\mathbf{r}, \nu)$ denotes the temporal Fourier transform of the observed Doppler velocities, the regressions in surface-focused HH are computed from

$$H_{\pm}^{P}(\mathbf{r}, \nu) = \int_{P} \mathrm{d}^2\mathbf{r}'\; G_{\pm}(\mathbf{r}, \mathbf{r}', \nu)\; \psi(\mathbf{r}', \nu). \tag{3}$$

H_+ and H_- are the egression and ingression which represent estimates of the amplitudes propagating into and out of the focal point at position $\mathbf{r}$ on the surface and ν is the temporal frequency. G_+ and G_- are Green's functions that express how a monochromatic point disturbance at a position $\mathbf{r}'$ on the surface propagates backward and forward in time, into the solar interior and back up to the focus. They are computed using the eikonal approximation (Lindsey and Braun 1997). The correlations,

$$C_{+}^{P}(\mathbf{r}) = \langle H_{+}^{P}(\mathbf{r}, \nu)\psi^{*}(\mathbf{r}, \nu)\rangle_{\Delta\nu}, \tag{4}$$

and

$$C_{-}^{P}(\mathbf{r}) = \langle \psi(\mathbf{r}, \nu) H_{-}^{P*}(\mathbf{r}, \nu)\rangle_{\Delta\nu}, \tag{5}$$

describe the egression and ingression control correlations respectively, which are directly comparable to center-annulus correlations used in time–distance (TD) helioseismology (e.g.,

Duvall Jr. et al. 1996; Braun 1997). The asterisk denotes complex conjugation, and the brackets indicate an average over a chosen positive frequency range $\Delta\nu$.

Surface-focused HH can be used to study flows by dividing the pupil P, over which the ingressions and egressions are computed, into four quadrants (labeled N, S, E, and W), each spanning 90° and oriented in the North, South, East and West directions respectively. We then compute the eight control correlations, $C_{\pm}^{N,S,E,W}$. Various combinations of these correlations are used to derive travel-time shifts due to the presence of flows or wave speed perturbations. In general, we compute travel-time shifts from various sums or differences of correlations such that, if C denotes some linear combination of correlations, the travel-time shift is

$$\delta\tau(\mathbf{r}) = \arg[C(\mathbf{r})]/2\pi\nu_0, \tag{6}$$

where ν_0 is the central frequency of the bandpass $\Delta\nu$. These represent travel-time shifts of the observed combination of waves relative to the travel times expected for the same ensemble of waves propagating in the solar model used to compute the Green's functions.

We present measurements of the mean travel-time shift ($\delta\tau_{\text{mean}}$), which represents the shift computed from the sum of all eight correlations:

$$\begin{aligned}\delta\tau_{\text{mean}}(\mathbf{r}) = \arg[&C_+^E(\mathbf{r}) + C_-^E(\mathbf{r}) + C_+^W(\mathbf{r}) + C_-^W(\mathbf{r}) \\ &+ C_+^N(\mathbf{r}) + C_-^N(\mathbf{r}) + C_+^S(\mathbf{r}) + C_-^S(\mathbf{r})]/2\pi\nu_0.\end{aligned} \tag{7}$$

We also present travel-time shifts sensitive to horizontal flows. For example, we define a "EW" travel-time asymmetry

$$\delta\tau_{\text{EW}}(\mathbf{r},\nu) = (\arg[C_-^E(\mathbf{r},\nu) + C_+^W(\mathbf{r},\nu)] - \arg[C_+^E(\mathbf{r},\nu) + C_-^W(\mathbf{r},\nu)])/2\pi\nu_0. \tag{8}$$

A similar travel-time shift can be measured for a North – South asymmetry. The sign of the travel-time perturbations is such that a perturbation in the background wave speed resulting in a faster propagation time will lead to a negative value of the mean travel-time shift ($\delta\tau_{\text{mean}}$) and a horizontal flow directed from East to West produces a negative value of the EW travel-time asymmetry ($\delta\tau_{\text{EW}}$).

Starting with the tracked, Postel-projected datacube described earlier (Sect. 2.1) we perform the following steps: 1) a temporal detrending by subtraction of a linear fit to each pixel signal in time, 2) removal of poor quality pixels, identified by a five-sigma deviation of any pixel from the linear trend, 3) Fourier transform of the data in time, 4) a correction for the amplitude suppression in magnetic regions (Rajaguru et al. 2006), 5) spatial Fourier transform of the data and multiplication by a chosen filter, 6) extraction of the desired frequency bandpass, 7) computation of Green's functions over the appropriate pupils, 8) computation of ingression and egression amplitudes by a 3D convolution of the data with the Green's functions, and 9) computation of the travel-time shift maps by (4)–(8).

In step 5 we have used two general kinds of filters: phase-speed filters and ridge filters. While not as commonly used as phase-speed filters, ridge filters have been used previously for f-mode studies (e.g., Duvall Jr. and Gizon 2000; Gizon and Birch 2002; Jackiewicz et al. 2007; Jackiewicz et al. 2007) and recently for p modes (Jackiewicz et al. 2008; Braun and Birch 2008). The phase-speed filters used here are the same set of 11 filters (hereafter denoted "TD1–TD11") listed by Gizon and Birch (2005) and Couvidat et al. (2006) and commonly used in time–distance analyses. The inner and outer radii of the corresponding pupil quadrants are chosen so that acoustic rays at a frequency of $\nu = 3.5$ mHz propagating from the focus to the edges of the pupil have phase speeds (denoted by w) which span the

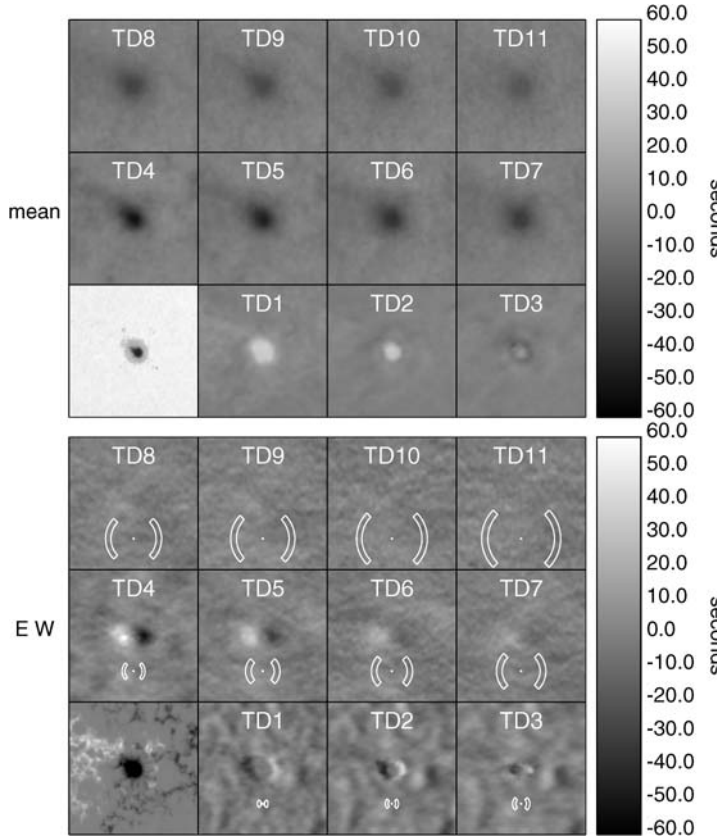


Fig. 6 Maps of travel-time shifts $\delta\tau_{\rm mean}$ (*top panels*) and $\delta\tau_{\rm EW}$ (*bottom panels*) using phase speed filters over a frequency bandpass of 2.5–5.5 mHz for AR 9787 observed over a 24 hour period on 2002 January 24. The labels TD1 through TD11 indicate the phase-speed filter used. Sizes of the East and West pupil quadrants used to measure $\delta\tau_{\rm EW}$ are shown in the bottom set of panels. The map in the lowest-left position of the top set of panels shows a MDI continuum intensity image while the map in the same position in the bottom set shows a line-of-sight magnetogram of the sunspot in AR 9787. The portion of the region shown here extends 219 Mm on each side. For the purpose of this figure (and the others in this section), some spatial smoothing has been applied to the travel-time maps

full width at half maximum (FWHM) of the squared filter. The EW pupil quadrants for the 11 filters are shown in the bottom set of panels in Fig. 6. All of the phase speed filters used also remove the contribution of the f mode (Braun and Birch 2008).

Figure 6 shows mean and EW travel-time shifts using a frequency bandpass (step 6) between 2.5–5.5 mHz. As is well known, there are distinct patterns of travel-time shifts associated with the use of phase-speed filters. In particular, the mean travel-time shift in the sunspot is positive for the smallest values of phase speed (and mean pupil diameter; e.g. TD1 – TD3) and then switches sign for larger phase speeds (TD4 and beyond). The EW travel-time shifts also undergo a similar change of sign. At the smallest (largest) phase speeds, the EW travel-time differences are consistent with inflow-like (outflow-like) perturbations centered on the sunspot. Remarkably, the switch in sign for both the mean shift and the EW differences occurs between filters TD3 and TD4.

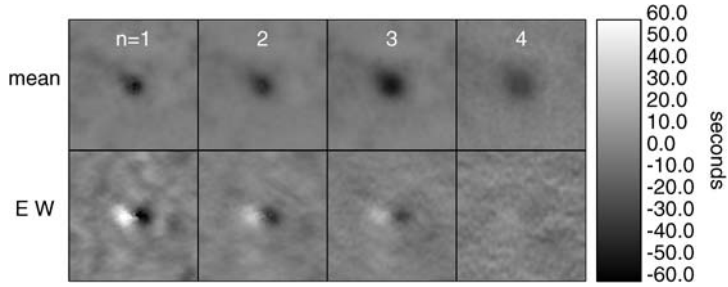


Fig. 7 Maps of travel-time shifts $\delta\tau_{\mathrm{mean}}$ (*top panels*) and $\delta\tau_{\mathrm{EW}}$ (*bottom panels*) using ridge filters over a frequency bandpass of 2.5–5.5 mHz for AR 9787 on 2002 January 24. The *columns* indicate different radial orders as indicated

Figure 7 shows mean and EW travel-time shifts using the same frequency bandpass (2.5–5.5 mHz) as shown in Fig. 6, but obtained with ridge filters isolating the $p_1 - p_4$ ridges. In contrast with the results obtained using phase-speed filters, ridge filters show values of $\delta\tau_{\mathrm{mean}}$ which are always negative within the sunspot, while $\delta\tau_{\mathrm{EW}}$ is consistent with an outflow-like perturbation. This is similar to results from time–distance analyses as well as previous HH analyses (Braun and Birch 2008).

Motivated by recent studies which show frequency variations of travel-time shifts observed in active regions (Braun and Birch 2006, 2008; Couvidat and Rajaguru 2007), we also employed narrow frequency bandpasses to both the phase-speed and ridge filters. Figure 8 shows mean and EW travel-time shifts using phase-speed filters TD1–TD4 in conjunction with 1-mHz wide frequency filters centered at 2, 3, 4, and 5 mHz. Positive mean travel-time shifts, and EW travel-time differences consistent with inflows, are observed primarily in frequency bandwidths that are centered below the p_1 ridge, shown by the solid line in Fig. 8. There is one instance (filter TD3 at 5 mHz, which lies immediately below the p_2 ridge) which also produces a positive mean shift and inflow-like signature. All other filters (including TD5–TD11 not shown) show negative mean travel-time shifts and outflow-like signatures.

Braun and Birch (2008) find that a condition for producing positive travel-time shifts such as in Fig. 8 appears to be a disproportionate contribution to the correlations of wave power from the low-frequency wing of the p_1 ridge relative to the high-frequency wing. We note that recent work by Moradi et al. (2008) is relevant to these issues.

3.2 Ridge and Off-Ridge Filtering

Following Braun and Birch (2006, 2008) and Thompson and Zharkov (2008), we study the sensitivity to filtering of mean travel-time perturbations measured in the vicinity of isolated sunspot AR 9787 relative to the surrounding quiet Sun, using a centre-to-annulus geometry and a skip distance of 11.7 Mm (Fig. 9). For each row a bandpass filter was used to select data within a 1 mHz frequency band with a 0.1 mHz Gaussian roll-off, centred (from bottom to top) at 2.5, 3.0, 3.5, 4.0, 4.5 and 5.0 mHz. These were combined for each column (left to right) with filters selecting the data from in between f and p_1 ridges, from the p_1 ridge, in between the p_1 and p_2 ridges, and the p_2 ridge. The filters were constructed as follows: at constant frequency we apply a filter that takes the value of unity at the horizontal wavenumber corresponding to either a particular ridge, e.g. p_1 (a "ridge filter"), or a midpoint between the adjacent ridges, e.g. p_1–p_2 (an "off-ridge filter"). On either side of this

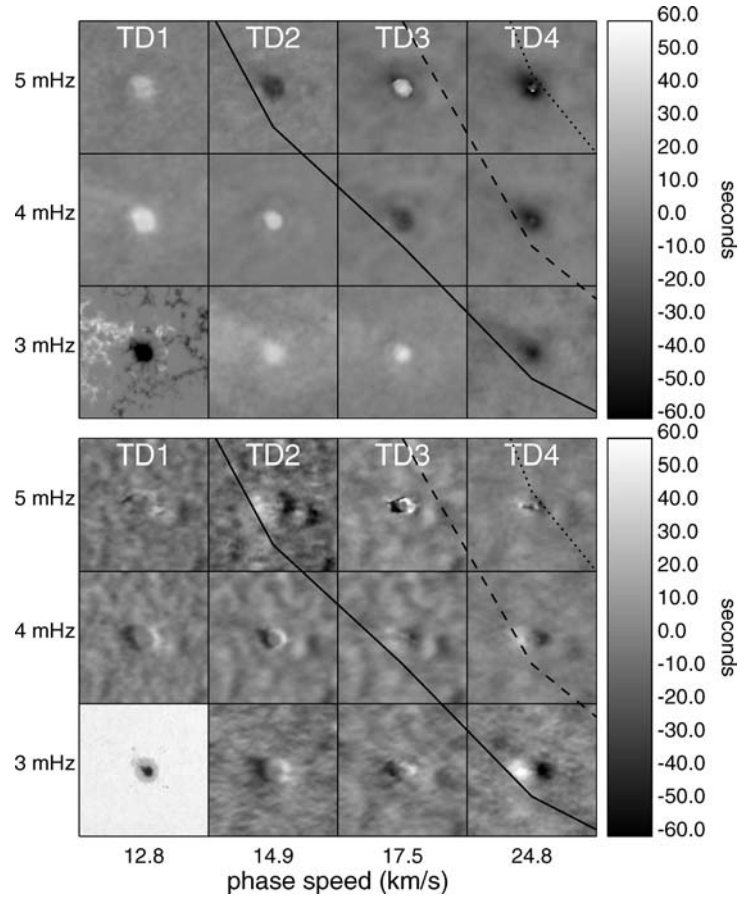


Fig. 8 Maps of travel-time shifts $\delta\tau_{\mathrm{mean}}$ (*top panels*) and $\delta\tau_{\mathrm{EW}}$ (*bottom panels*) using phase speed filters and 1 mHz-wide frequency bandpasses for AR 9787 observed on 2002 January 24. The columns of maps labeled TD1 through TD4 indicate the phase-speed filter used, while the rows indicate the frequency bandpass. The *solid jagged line* running diagonally through the panels connects the location of the p_1 ridge in the ν–w domain for each filter, with the centers of the maps assigned to values of frequency and phase speed as indicated on the left and bottom edges of the plot. The *dashed* and *dotted lines* indicate the locations of the p_2 and p_3 ridges respectively. The map in the lowest-left position of the top set of panels shows a MDI continuum intensity image while the map in the same position in the bottom set shows a line-of-sight magnetogram

centre line the filter has a Gaussian roll-off with half width at half maximum (HWHM) equal to 0.32 times the distance to the neighbouring ridge on that side for the ridge filter and with HWHM equal to 0.63 times the distance to the adjacent ridge in the case of the off-ridge filter. No phase-speed filter was applied.

The cross-correlation functions were estimated using center-to-annulus geometry with annuli taken to be one data pixel wide. Travel-time perturbations were measured using the definition of Gizon and Birch (2004).

In agreement with Braun and Birch (2006, 2008), as illustrated in Fig. 9, we observe a positive travel-time perturbation in the region beneath the p_1 ridge, but we also find such a signal between the p_1 and p_2 ridges, and find that on the p_1 ridge the positive perturbation

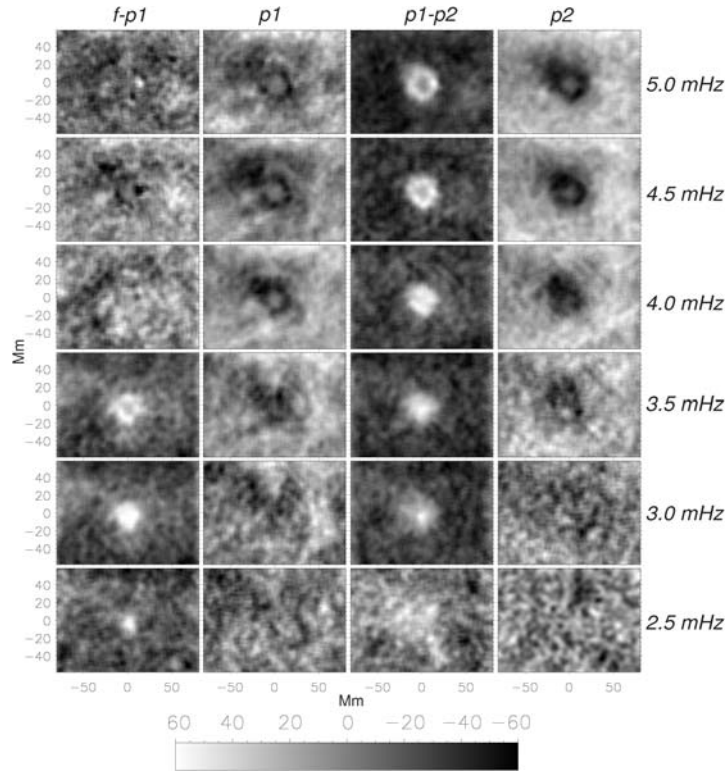


Fig. 9 Travel-time perturbations $\delta\tau_{\text{mean}}$ for isolated sunspot AR 9787 obtained for various filtering schemes. The *colorbar* is in units of seconds. The skip distance is equal to 11.64 Mm. From the left, the columns respectively have filters applied as follows: pass-filter centred between the f and p_1 ridges; pass-filter centred on the p_1 ridge; pass-filter centred between the p_1 and p_2 ridges; pass-filter centred on the p_2 ridge. From the bottom, the rows have bandpass filters centred on 2.5, 3.0, 3.5, 4.0, 4.5, and 5.0 mHz respectively. More details are given in the text

is absent. Our tentative conclusion is that the positive travel-time perturbation signal arises only in the regions between the p-mode ridges; and that the travel-time perturbations associated with the data on the ridges themselves are all consistently negative. Similar, though noisier, results were obtained for the Gabor-wavelet fitting travel-time definition. The data appear to suggest that measured travel-time perturbations are very sensitive to the part of the wave-propagation diagram selected during the filtering stage.

4 Moat Flow Inversion (Ridge Filters)

Here we invert travel times to obtain the flows around the sunspot in AR 9787. We restrict ourselves to ridge filtering and TD travel times. We then compare the inferred flows with the velocities of the moving magnetic features (MMFs) in the moat (Sect. 2.1).

The details of the travel-time inversion can be found in Jackiewicz et al. (2008). In summary, we measure center-to-quadrant travel-time differences using the method of Gizon and Birch (2004), after correcting for reduced power in magnetic regions. The travel times are obtained for the f ridge and the p_1 to p_4 ridges, separately. For all five ridges, the east–west

travel-time differences are consistent with an outflow from the center of the sunspot. Using the same definition of travel times, Born sensitivity kernels are computed (Birch and Gizon 2007). We input the travel times and the sensitivity functions, as well as the covariance of the travel times, into a three-dimensional subtractive optimally localised averages (SOLA) inversion procedure to infer the vector flows at several depths. The procedure also provides good estimations of the resolution and the noise levels, which are important for any interpretation. We note that neither the modeling nor the inversion takes into account the magnetic field.

We have inverted the middle seven days of data from the nine day set. We obtain flow maps for several depths, extending down to about 5 Mm beneath the surface. In Fig. 10 we compare the inversion results near the surface (left column) to the motion of the moving magnetic features (MMFs, right column). For both sets of maps, the flows are averaged over 7 days. Furthermore, they are approximately of the same horizontal resolution ($\sim$6 Mm). For this particular figure, we choose to study the inferred time–distance flows taken at a depth as near to the surface as we can achieve, about 1 Mm below.

The bottom row of Fig. 10 compares the radial velocities derived from TD helioseismology and MMF tracking. We see for each case quite clearly a strong outflow extending beyond the penumbra (20 Mm) of several hundred m/s, known as the moat flow. The overall features of the flows from both methods are quite similar, even the slight 'knob' on the northeast quadrant of the sunspot moat. A look in the quiet Sun reveals other similarities. The correlation coefficient between the two maps is about 0.65. The magnitude of the TD surface flows is about 20% less than that revealed by MMF tracking. This can be due to many factors, such as the implied depth at which we are comparing not being equal, inaccurate travel-time sensitivity kernels, or the magnetic field affecting the inversion results through the travel times, among others. The estimated noise in the time–distance maps is about 5 m/s. These results are consistent with a previous f-mode TD study of sunspot moat flows (Gizon et al. 2000).

Since the sunspot is very nearly circular, we may average the radial flows azimuthally about the sunspot center. In Fig. 11 we plot the azimuthally-averaged radial flows for several depths versus the distance from the center of the sunspot. Also shown for comparison is the averaged MMF velocity. The moat flow, which extends to about 45 Mm, is seen at all depths. The inferred flows get stronger with depth and begins to level off in strength at about 4.5 Mm below the surface. The MMF velocity is consistent with these flows at depths of between 1 and 2.6 Mm beneath the surface. The variations of the moat flow as a function of radial distance is similar for all cases.

5 Sound Speed versus Wave Speed

The sound speed, c, is the speed of sound. In this paper we make an important distinction between c and the speed at which waves are inferred to have travelled under the assumptions of the helioseismic inversions. Different inversions will include different physics. A common example is the assumption that a travel-time perturbation associated with a sunspot can be modelled purely by an equivalent small-amplitude sound-speed perturbation. The sound-speed inferred under such an assumption is at best a local wave speed, and need not even reflect the sign of the sound speed perturbation. For this reason, we use the notation $c_{\rm w}$ to denote the inferred wave speed.

Lin et al. (2008) studied the meaning of $c_{\rm w}$ in the case of ring-diagram analysis. They find that their inversions return perturbations in $c_{\rm w}^2 = \Gamma_1 P_{\rm tot}/\rho$, where $P_{\rm tot} = P + P_{\rm mag}$ is the

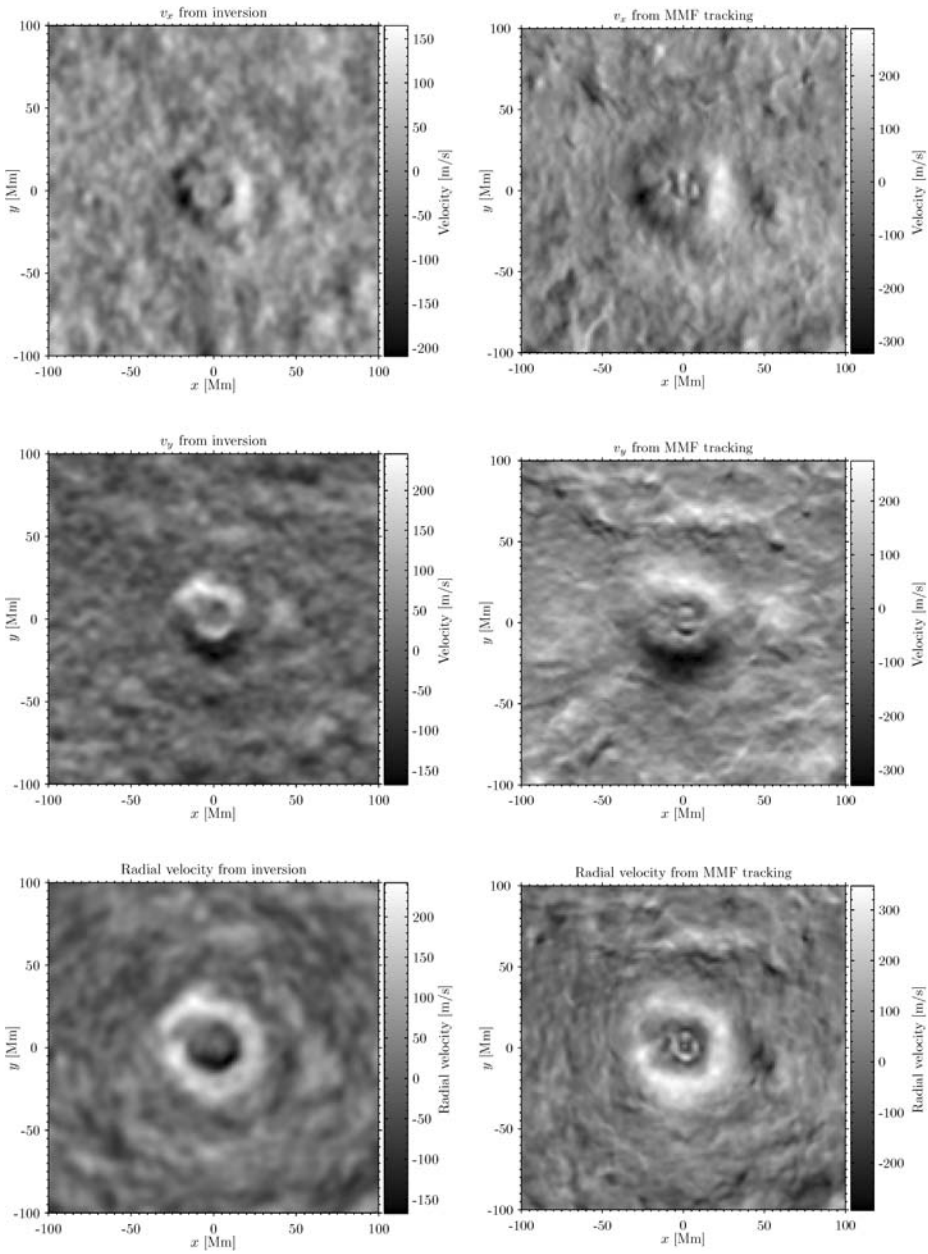


Fig. 10 Comparison of near-surface flows around the sunspot from time–distance inversions and MMF tracking. The time–distance analysis uses five ridge filters (f and p_1 to p_4). The *left column* shows flows obtained from inversions at a depth of 1 Mm beneath the surface and averaged over 7 days. The *right column* are the flows obtained from MMF tracking averaged over the same 7 days. Each set of flows was obtained with approximately the same resolution/smoothing per pixel. The *top row* is the x component of the velocity, and the *middle row* shows the y component of the velocity. The *bottom row* compares the radial velocity (from the center of the sunspot) of the two measurements. An outward moat flow is seen for each case

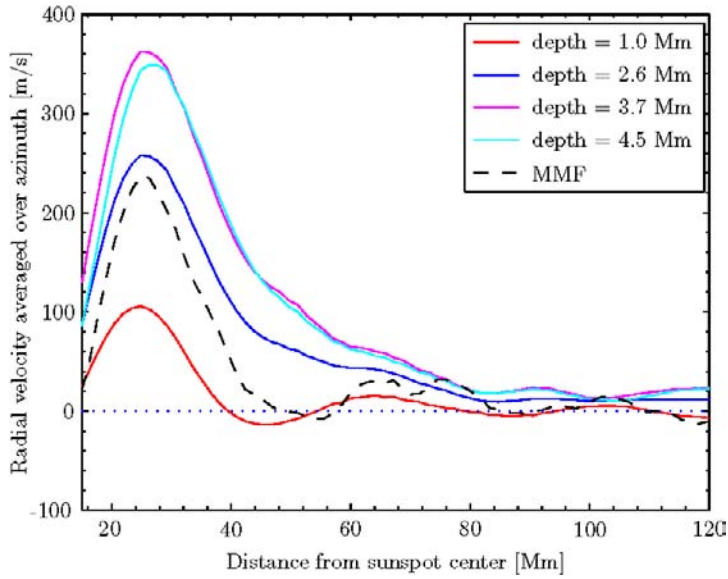


Fig. 11 Azimuthally-averaged radial flows from the sunspot center at different depths over 7 days, obtained from a time–distance inversion using ridge filtering. The radial velocity from the MMF tracking is shown by the *dashed line*. In this plot only the region from the edge of the penumbra outward is shown. The results within the sunspot (distance less than 20 Mm) cannot be reliably interpreted

sum of the gas and magnetic pressures. The inferred wave speed $c_{\rm w}$ has two components: the sound speed, c, and a magnetic component.

What we have said about sound speed versus wave speed, also applies to any chosen physical quantity versus inferred quantity. For example, the next section presents an inversion for wave slowness, $s_{\rm w} = 1/c_{\rm w}$, using traveltime sensitivity kernels for inverse sound speed $s = 1/c$.

6 Wave-Speed Inversion (Phase-Speed Filters)

The data was reduced using standard time–distance helioseismology techniques. First, the data was preprocessed by applying an amplitude modulation correction as described by Rajaguru et al. (2006), followed by applying a high-pass filter at 1.7 mHz in order to remove the supergranulation and a low-pass filter at 5.1 mHz to remove signal above the acoustic cut-off frequency. We then apply a Gaussian phase speed filter to select waves with horizontal phase speed w_i near the value corresponding to skip distance Δ_i given by ray theory. We define 12 such filters F_i corresponding to different distances Δ_i. These phase-speed filters are similar to the set of filters used in Sect. 3.1 and commonly used in time–distance analyses. Each filter is applied by pointwise multiplication of the Fourier transform $\psi(k_x, k_y, \omega)$ of the observed velocity data.

We use a centre-to-annulus geometry to compute the cross-covariance $C(\mathbf{r}, \Delta_i, t)$, where $\mathbf{r}$ is the center of the annulus and t is the time lag. The annulus width is 4.5 Mm. A reference cross-covariance function, $C^{\rm ref}$, is obtained by spatial averaging C over a quiet Sun area. Wave travel times are then extracted by fitting a Gabor wavelet to the positive- and negative-time branches of C (Kosovichev and Duvall Jr. 1997). The wavelet has five parameters:

the central frequency, the width and amplitude of the envelope, and the group and phase travel times. We denote by τ_+ and τ_- the measured phase travel times for the positive- and negative-time branches of C respectively. The reference travel times for the quiet Sun are similarly defined using $C^{\rm ref}$. The phase travel time perturbations, $\delta\tau_+$ and $\delta\tau_-$, are defined as the difference between the measured and reference travel times. As we are interested in wave-speed perturbations only, we consider mean travel-time perturbations, $\delta\tau_{\rm mean} = (\delta\tau_+ + \delta\tau_-)/2$.

For the forward problem we use sensitivity kernels estimated using the first-order Rytov approximation (Jensen and Pijpers 2003). These kernels, K^s, relate mean travel-time perturbations, $\delta\tau_{\rm mean}$, to inverse sound speed perturbations, $\delta s = \delta(1/c)$, of a quiet-Sun model. In the sunspot region, we have

$$\delta\tau_{\rm mean}(\mathbf{r}, \Delta_i) = \int_S \mathrm{d}^2\mathbf{r}' \int_{-d}^{0} \mathrm{d}z\; K^s(\mathbf{r} - \mathbf{r}', z; \Delta_i)\, \delta s_{\rm w}(\mathbf{r}', z), \tag{9}$$

where S is the area of the region, d is its depth. The quantity $\delta s_{\rm w} = s_{\rm w} - s$ is the equivalent change in the local wave slowness caused by the sunspot.

We invert for $N = 14$ layers in depth located at $[z_1, \ldots, z_N] = [0.36,\ 1.2,\ 2.1,\ 3.3,\ 4.7,\ 6.4,\ 8.6,\ 11.2,\ 14.3,\ 17.8,\ 21.8,\ 26.3,\ 31.4,\ 37.0]$ Mm. We use a multi-channel deconvolution algorithm (Jensen et al. 1998, 2001) enhanced by the addition of horizontal regularisation (Couvidat et al. 2006). The above equation is Fourier transformed with respect to two-dimensional position $\mathbf{r}$. For each wavevector $\mathbf{k}$, we define $d_i = \delta\tau(\mathbf{k}, \Delta_i)$, $G_{ij} = K(\mathbf{k}, z_j; \Delta_i)$, and $m_j = \delta s(\mathbf{k}, z_j)$, and the corresponding vector $\mathbf{d}$, matrix G, and vector $\mathbf{m}$. Then for each $\mathbf{k}$ we solve for the vector $\mathbf{m}$ that minimises

$$\|(\mathbf{d} - G\mathbf{m})\|^2 + \epsilon\|L\mathbf{m}\|^2, \tag{10}$$

where L is a regularisation operator and $\epsilon(\mathbf{k})$ is a positive regularisation parameter. In this work we apply more regularisation at larger depths, to which travel times are less sensitive, by setting $L = \mathrm{diag}(c_1, c_2, \ldots, c_N)$, where $c_j = c(z_j)$ is the sound speed in the j-th layer of the reference model. We regularise small horizontal scales by taking $\epsilon(\mathbf{k}) = 2 \times 10^3(1 + |\mathbf{k}|^2)^{100}$.

Figure 12 shows the result of the inversion, expressed in terms of the relative wave-speed perturbation $\delta c_{\rm w}/c$. We see a two-layer structure: a region of decreased wave speed (down by -13%) situated directly underneath the surface and a region of increased wave speed (up to 9%) starting from a depth of approximately 3 Mm. This is consistent with other time–distance inversions of travel times using phase-speed filters, e.g., those of Kosovichev et al. (2000).

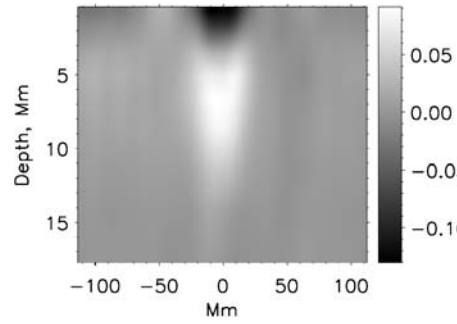


Fig. 12 Relative wave speed perturbation, $\delta c_{\rm w}/c$, obtained for AR 9787 using time–distance helioseismology and phase speed filtering

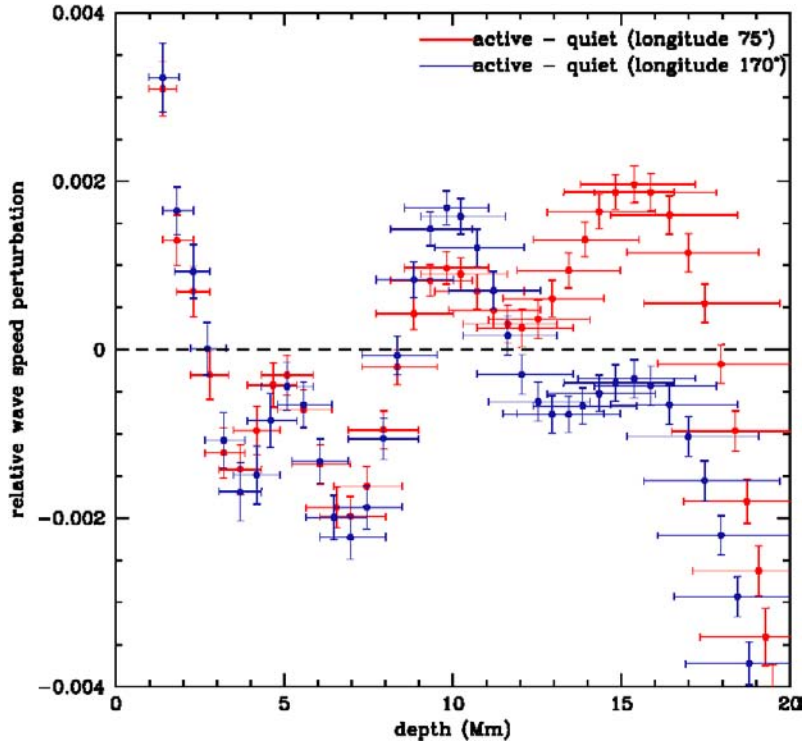


Fig. 13 SOLA inversions for the depth dependence of the relative difference in wave speed, $\delta c_{\mathrm{w}}/c$, between the region around AR 9787 and each of two comparison quiet regions at the same latitude but different longitudes. The inversions are based on fits to power spectra for 5.7 days of tracked data for each region. The differences are in the sense $\delta c_{\mathrm{w}} = c_{\mathrm{w,active}} - c_{\mathrm{w,quiet}}$. The active region is seen to have a positive wave-speed anomaly very near the surface relative to the quiet regions

7 Ring-Diagram Analysis

We have used ring diagrams to analyse the mean structure of the region containing AR 9787 compared with quiet-Sun structure. To do so, we use the techniques described in Basu et al. (2004). We invert the differences in the ring-diagram fit parameters between the spectra of the active region and those of suitable selected quiet-Sun regions. In this case, two quiet-Sun regions were chosen at the same latitude as that of the active region ($-7°$) and at Carrington longitudes 170° and 75° (the active region is at longitude 130°). Each region is independently tracked in a time interval of 5.7 days centered on its central meridian crossing, so any geometrical differences in the spectra due to foreshortening or geometric image distortion are very nearly cancelled out. The regions chosen for analysis are 15° in diameter, so the results apply to a spatial mean over these areas (with an unknown weighting function). The average results from the two comparison regions is shown in Fig. 13. It is clear that there is a region of negative wave-speed anomaly under the active region between 3 Mm and 8 Mm in depth, with a turnover to positive wave-speed anomalies both above and below this region, and yet another turnover to negative anomalies at depths greater than about 17 Mm. This behaviour is typical of that seen for other active regions (Bogart et al. 2008), although the changes at the surface and deeper than 17 Mm are unusually pronounced in this case.

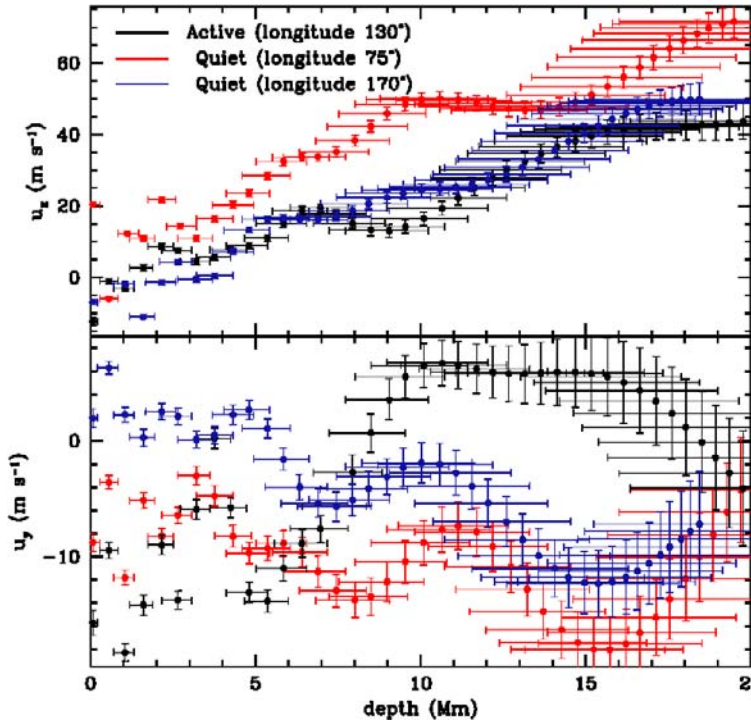


Fig. 14 SOLA inversions for the depth dependence of the mean zonal (v_x) and meridional (v_y) flows over each of the three regions analysed with ring-diagram analysis. The zonal rates are relative to the tracking velocity, which was 54 m/s less than the Carrington rate at this latitude

We can also infer the mean flow speeds at depth for the active regions and the comparison quiet regions directly from the fitted parameters to the ring-diagram spectra. These are shown in Fig. 14. There is no evident anomalous zonal flow through the active region; indeed, the zonal flow structure is remarkably similar to that of the preceding comparison region at longitude 170°. There does appear to be an anomaly in the structure of the meridional flow, however, with a substantial shear at depths greater than 7 Mm, the flow being poleward near the surface (the region is in the southern hemisphere) and equatorward at greater depths. It is especially marked if the mean meridional velocity at the active region's latitude is negative at depth, as the two comparison regions suggest, but this needs to be verified by averaging over more longitudes.

8 Numerical Forward Modeling

In various circumstances it has been shown that the cross-covariance is closely related to the Green's function. This allows us to characterise the interaction of arbitrary wavepackets with the sunspot from the MDI observations. The sunspot discussed in this paper, being observed over nine days and almost axisymmetric, is ideally suited to such a study. In preliminary work, Cameron et al. (2008) considered the cross-covariance between the Doppler signal averaged along a great circle 40 Mm from the centre of the sunspot and the Doppler signal at each point in a region surrounding the sunspot. The data had been f-mode ridge filtered.

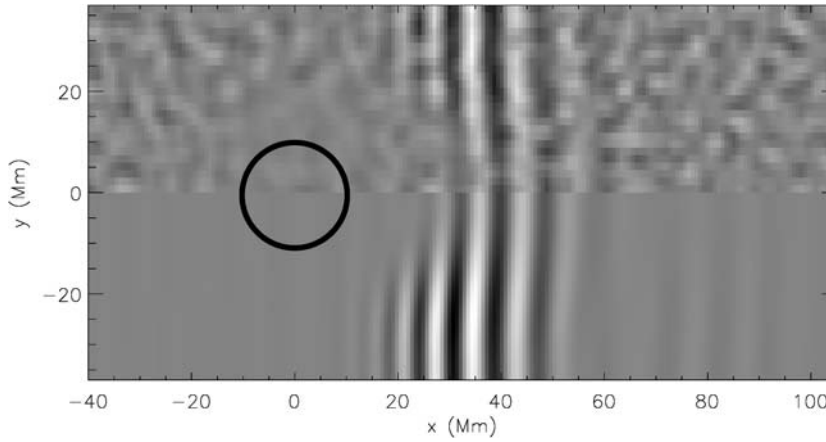


Fig. 15 Comparison between the observed cross-covariance and a numerical simulation. The *circle* indicates the location of the sunspot in AR 9787. (*Top half*) f-mode cross-covariance between the MDI Doppler signal averaged over a line at $x = -40$ Mm and the Doppler signal at each point. The correlation time-lag is 130 min, large enough for wave packets to traverse the sunspot. The cross-covariance is averaged over 9 days and uses the assumed azimuthal symmetry of the sunspot to reduce noise. (*Bottom half*) SLiM numerical simulation of an f-mode wave packet propagating in the $+x$ direction through a model of AR 9787 with a peak magnetic field of 3 kG. (From Cameron et al. 2008)

A numerical simulation was then performed of the propagation an f-mode plane wave packet beginning 40 Mm from a model sunspot. The background atmosphere is Model S of Christensen-Dalsgaard et al. (1996), stabilised with respect to convection. The sunspot model used was a simple self-similar model in the vein of Schlüter and Temesváry (1958). The half-width of the vertical magnetic field at the surface was taken to be 10 Mm (as in AR 9787) and field strengths of 2000, 2500 and 3000 G were considered. The SLiM code (Cameron et al. 2007) was used to perform the simulations.

The top half of Fig. 15 shows the observed cross-correlation with time-lag 130 min. The bottom half of the figure shows the results of the simulation at time 130 min. The match between the observations and simulation is quite good, in this case for a sunspot with a peak field strength of 3000 G. The match was not as good for peak field strengths of 2000 G and 2500 G. This then places a helioseismic constraint on the magnetic field of the spot. Whereas this constraint makes sense, it cannot be assessed directly using MDI magnetograms, which are not reliable in sunspot umbrae. The partial absorption of the waves (reduced cross-covariance amplitude) was explained in terms of mode conversion into slow magneto-acoustic waves that propagate down the sunspot, as predicted by, e.g., Cally (2000). Full details of this work are given in Cameron et al. (2008).

For the 3 kG sunspot model, Fig. 16 shows the relative change in the gas pressure and the square of the sound speed along the sunspot axis with respect to the quiet Sun model. Both perturbations become very small (less than 1%) deeper than a depth of about 4 Mm. The temperature, closely related to the squared sound speed, is reduced at all depths within the sunspot. At the surface the relative decrease in temperature is around 18%. The reduction in gas pressure is larger with a 36% decrease. The vertical dashed line indicates that the sound speed is equal to the Alfvén speed ($c = a$) at a depth of approximately 580 km. The $c = a$ level is where mode conversion is expected to occur. We also plot the depth at which the sound speed is equal to the surface quiet-Sun sound speed. This gives a rough indication of the Wilson depression, in this case a rather low 170 km.

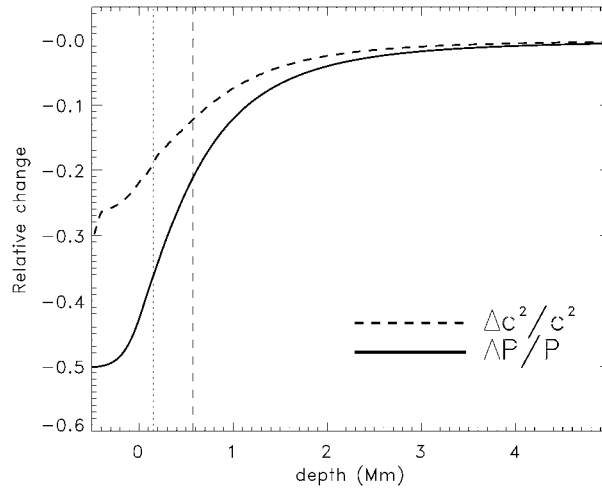


Fig. 16 Structure of the 3 kG sunspot model used in the simulation of Fig. 15. The *thick solid curve* shows the relative change in gas pressure, P, with respect to the quiet Sun, measured along the sunspot axis. The *thick dashed curve* shows the relative change in the squared sound speed, c^2, which is also an estimate of the relative change in temperature. The *vertical dashed line* indicates the depth at which the sound speed and the Alfvén velocity are equal. The *vertical dotted line* indicates the depth at which the sound speed is equal to its zero-depth quiet-Sun value

9 Discussion: The Elusive Structure of Sunspots

9.1 Problematic Travel Times

Here we use Born approximation based forward modeling to test if the flows estimated from inversions of the travel-times obtained using ridge filters (Sect. 4) are consistent with the travel-times measured using phase-speed filters (Sect. 3).

To carry out the forward modeling we employ sensitivity functions, kernels, computed using the method of Birch and Gizon (2007). The calculations account for both the phase-speed filters and pupil sizes used for the measurements described in Sect. 3. The resulting kernels, $\mathbf{K}^v$, relate three-dimensional steady flows, $\boldsymbol{v}$, to predictions for the EW travel-times differences,

$$\delta\tau_{\mathrm{EW}}(\boldsymbol{r}, \Delta) = \int \mathrm{d}^2\boldsymbol{r}'\mathrm{d}z\, \mathbf{K}^v(\boldsymbol{r}' - \boldsymbol{r}, z, \Delta) \cdot \boldsymbol{v}(\boldsymbol{r}', z) \tag{11}$$

where $\boldsymbol{r}$ and $\boldsymbol{r}'$ are two-dimensional position vectors and z is depth. Both the kernel functions and the flow are vector-valued functions of horizontal position and depth. The travel-time differences are functions of horizontal position and also the pupil size Δ (notice that for each pupil size there is a corresponding phase-speed filter, see Sect. 3).

We assume $\boldsymbol{v}$ to be given by the flow field inferred from the inversions of the ridge-filtered travel-time differences shown in Sect. 4. We neglect the effects of vertical flows as inversions for depth-dependent vertical flows have not yet been carried out.

Figure 17 compares measurements and forward models of EW travel-time differences for the cases of phase-speed filters TD1 and TD4. For the case of the filter TD1, the forward model is in qualitative agreement with the measurements in the quiet Sun. However, the forward model predicts a signature of the moat flow with opposite sign to that seen in the observations. This shows that the travel-time differences measured using phase-speed filter TD1 and those measured using ridge-filtered travel-times do not yield a consistent picture of the moat flow. Figure 17 also shows the case of measurements and forward modeling for phase-speed filter TD4. In this case, and for other filters with large phase speeds, there is qualitative agreement even in the moat.

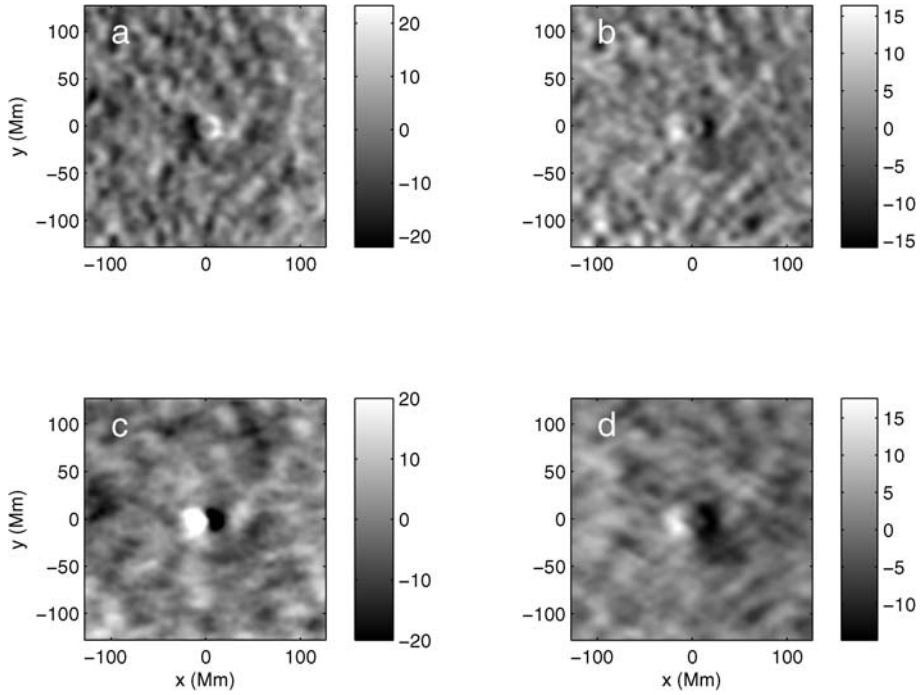


Fig. 17 Measured EW travel-time differences for phase-speed filters TD1 (panel **a**), and TD4 (panel **c**), and the corresponding modeled travel times (panels **b** and **d**) for 25 January 2002. The travel-time measurements are described in Sect. 3 and have been smoothed with a Gaussian filter with FWHM of 6.7 pixels. The *gray scales* are shown in units of seconds and have been truncated to make the details more visible. The sunspot is centered at roughly $(x, y) = (0, 0)$. Notice that the forward model is able to reproduce the moat flow for the filter TD4, but not for the case of TD1

One possible reason for the disagreements could be that the travel-time sensitivity kernels rely on an imperfect model of the power spectrum of solar oscillations. For example, the model zero-order power spectrum does not include background noise and mode linewidths may not be accurate enough.

We also emphasise that a number of assumptions have been made in carrying out the forward modeling shown in Fig. 17. It is known (e.g., Gizon and Birch 2002; Parchevsky et al. 2008; Hanasoge et al. 2007) that the reduction in the wave generation rate in sunspots can, in general, produce apparent travel-time differences. For the case of phase-speed filtered travel-time differences this effect has a magnitude of up to 10 s (Hanasoge et al. 2007). The magnitude of this effect is not known for travel-times measured using ridge filters. Similarly, wave damping in sunspots can also produce travel-time differences (e.g., Woodard 1997; Gizon and Birch 2002). The magnitude of this effect has not been carefully estimated for realistic models of wave absorption in sunspots. In addition, radiative transfer effects can cause phase shifts in sunpots (e.g., Rajaguru et al. 2007).

9.2 Conflicting Wave-Speed Profiles

Here we compare the wave-speed inversions from ring analysis (Sect. 7) and from time–distance helioseismology with phase-speed filters (Sect. 6). To make this comparison possible, we average the TD wave-speed inversion over a disk of 15° diameter centered on the

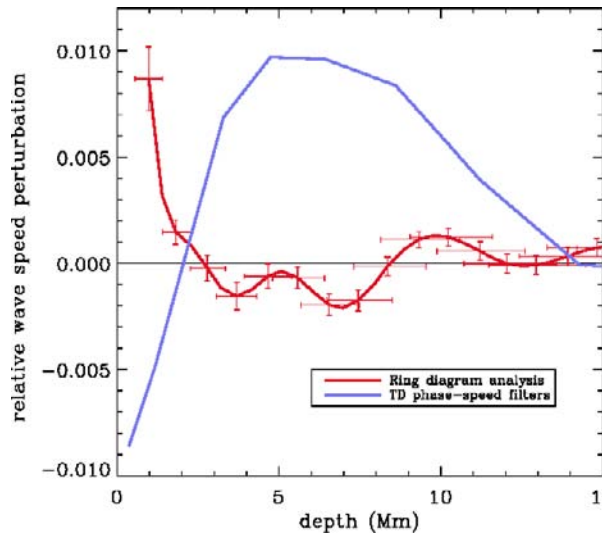


Fig. 18 Comparison of two different helioseismic methods used to infer wave speed perturbations below AR 9787 ($\delta c_{\mathrm{w}}/c$). The *red curve* shows the averaged ring-diagram results from Fig. 13. The *solid blue curve* shows the time–distance result (phase-speed filters) from Fig. 12, after averaging over the area used for ring analysis. Although they are meant to represent the same quantity, these two curves are noticeably different

sunspot, which represents the area used for ring analysis. The two wave-speed profiles are plotted as a function of depth in Fig. 18. Clearly, they do not match.

How can we explain such a strong disagreement? As already mentioned in Sects. 3 and 9.1, the details of the measurement procedures are important for the interpretation of the helioseismic observations; they may not have been fully taken into account in one or possibly both inversions. Although we have not done a TD inversion for wave speed using ridge filters, it is likely that it would give a different answer than the TD inversion using phase-speed filters, thus adding a third curve to Fig. 18.

We also note that both inversions suppose that first-order perturbation theory is valid to describe the effect of sunspots on waves. Unlike the flow perturbation, however, the perturbations in pressure and density introduced by the sunspot are not small with respect to the quiet-Sun background. Thus the concept of linear inversions is not necessarily correct for sunspots and regions of strong magnetic field. In addition, it is perhaps too naive to model the combined effects of the magnetic field in terms of an equivalent sound-speed perturbation. Ring-diagram inversions do include a contribution from changes in the first adiabatic exponent, but the direct effect of the magnetic field through the Lorentz force is not fully accounted for in either inversion. We note that the ring-diagram inversions include a treatment of near-surface effects which is different than in the TD inversions.

10 Conclusion

We have studied the sunspot in AR 9787 with several methods of local helioseismology. We have characterised the acoustic wave field near the sunspot and the surrounding plage, measured acoustic absorption by the sunspot, and showed maps of the signature of AR 9787 on the farside of the Sun. We have shown that the sunspot leaves a strong signature in the observed wave field, as evidenced by strong perturbations in travel times and frequency shifts. The interpretation of the observations, however, is difficult and we have not been able to draw an unequivocal conclusion about the subsurface structure and dynamics of

the sunspot. We have shown that one complication is the extreme sensitivity of helioseismic measurements to the choice of data analysis procedure, such as filtering in frequency-wavenumber space. In addition, our understanding of the effects of strong magnetic fields on solar oscillations is still incomplete.

On the positive side, we note that the seismically determined moat flow (TD and ridge filters) appears to be consistent with the motion of the MMFs in magnetograms. It is also clear that numerical simulations of wave propagation through model sunspots promise to provide invaluable help in interpreting the observations.

Acknowledgements This work was initiated at the second HELAS local helioseismology workshop held in Freiburg on 7–11 January 2008 and continued during the ISSI Workshop on the Origin and Dynamics of Solar Magnetism held in Bern on 21–25 January 2008. Financial support from project HELAS (European Union FP6), ISSI, project GDC-SDO (Deutsches Zentrum für Luft- und Raumfahrt), and project SISI (European Research Council) is acknowledged. SOHO is a mission of international cooperation between ESA and NASA.

References

H.M. Antia, S. Basu, Astron. Nachr. **328**, 257 (2007)
S. Basu, H.M. Antia, R.S. Bogart, Astrophys. J. **610**, 1157 (2004)
A.C. Birch, L. Gizon, Astron. Nachr. **328**, 228 (2007)
R.S. Bogart, S. Basu, M.C. Rabello-Soares, H.M. Antia, Sol. Phys. **251**, 439 (2008)
T.J. Bogdan, F. Cattaneo, A. Malagoli, Astrophys. J. **407**, 316 (1993)
D.C. Braun, Astrophys. J. **451**, 859 (1995)
D.C. Braun, Astrophys. J. **487**, 447 (1997)
D.C. Braun, A.C. Birch, Astrophys. J. **647**, 187 (2006)
D.C. Braun, A.C. Birch, Sol. Phys. **251**, 267 (2008)
D.C. Braun, C. Lindsey, Sol. Phys. **192**, 307 (2000)
D.C. Braun, T.L. Duvall Jr., B.J. Labonte, Astrophys. J. **319**, 27 (1987)
D.C. Braun, T.L. Duvall Jr., B.J. Labonte, Astrophys. J. **335**, 1015 (1988)
D.C. Braun, T.L. Duvall, B.J. Labonte, S.M. Jefferies, J.W. Harvey, M.A. Pomerantz, Astrophys. J. **391**, 113–116 (1992a)
D.C. Braun, C. Lindsey, Y. Fan, S.M. Jefferies, Astrophys. J. **392**, 739–745 (1992b)
N.S. Brickhouse, B.J. Labonte, Sol. Phys. **115**, 43–60 (1988)
P.S. Cally, Sol. Phys. **192**, 395 (2000)
P.S. Cally, A.D. Crouch, D.C. Braun, Mon. Not. R. Astron. Soc. **346**, 381 (2003)
R. Cameron, L. Gizon, K. Daiffallah, Astron. Nachr. **328**, 313 (2007)
R. Cameron, L. Gizon, T.L. Duvall Jr., Sol. Phys. **251**, 291 (2008)
J. Christensen-Dalsgaard, W. Dappen, S.V. Ajukov, E.R. Anderson, H.M. Antia, S. Basu, V.A. Baturin, G. Berthomieu, B. Chaboyer, S.M. Chitre, A.N. Cox, P. Demarque, J. Donatowicz, W.A. Dziembowski, M. Gabriel, D.O. Gough, D.B. Guenther, J.A. Guzik, J.W. Harvey, F. Hill, G. Houdek, C.A. Iglesias, A.G. Kosovichev, J.W. Leibacher, P. Morel, C.R. Proffitt, J. Provost, J. Reiter, E.J. Rhodes, F.J. Rogers, I.W. Roxburgh, M.J. Thompson, R.K. Ulrich, Science **272**, 1286 (1996)
S. Couvidat, S.P. Rajaguru, Astrophys. J. **661**, 558 (2007)
S. Couvidat, A.C. Birch, A.G. Kosovichev, Astrophys. J. **640**, 516 (2006)
A.D. Crouch, P.S. Cally, P. Charbonneau, D.C. Braun, M. Desjardins, Mon. Not. R. Astron. Soc. **363**, 1188 (2005)
A.C. Donea, C. Lindsey, D.C. Braun, Sol. Phys. **192**, 321 (2000)
T.L. Duvall Jr., L. Gizon, Sol. Phys. **192**, 177 (2000)
T.L. Duvall Jr., S.M. Jefferies, J.W. Harvey, M.A. Pomerantz, Nature **362**, 430 (1993)
T.L. Duvall Jr., S. D'Silva, S.M. Jefferies, J.W. Harvey, J. Schou, Nature **379**, 235 (1996)
Y. Fan, D.C. Braun, D.Y. Chou, Astrophys. J. **451**, 877 (1995)
L. Gizon, Mod. Phys. Lett. A **21**, 1701 (2006)
L. Gizon, A.C. Birch, Astrophys. J. **571**, 966 (2002)
L. Gizon, A.C. Birch, Astrophys. J. **614**, 472 (2004)
L. Gizon, A.C. Birch, Living Rev. Sol. Phys. **2**(6) (2005). http://www.livingreviews.org/lrsp-2005-6
L. Gizon, T.L. Duvall Jr., R.M. Larsen, J. Astrophys. Astron. **21**, 339 (2000)

S.M. Hanasoge, Astrophys. J. **680**, 1457 (2008)
S.M. Hanasoge, S. Couvidat, S.P. Rajaguru, A.C. Birch (2007). arXiv:0707.1369
F. Hill, Astrophys. J. **333**, 996 (1988)
B.W. Hindman, T.M. Brown, Astrophys. J. **504**, 1029 (1998)
J. Jackiewicz, L. Gizon, A.C. Birch, Sol. Phys. **251**, 381 (2008)
J. Jackiewicz, L. Gizon, A.C. Birch, M.J. Thompson, Astron. Nachr. **328**, 234 (2007)
J. Jackiewicz, L. Gizon, A.C. Birch, T.L. Duvall, Astrophys. J. **671**, 1051 (2007)
K. Jahn, in *NATO ASIC Proc. 375: Sunspots. Theory and Observations*, ed. by J.H. Thomas, N.O. Weiss, 1992, pp. 139–162
J.M. Jensen, F.P. Pijpers, Astron. Astrophys. **412**, 257 (2003)
J.M. Jensen, B.H. Jacobsen, J. Christensen-Dalsgaard, in *Structure and Dynamics of the Interior of the Sun and Sun-Like Stars SOHO6/GONG98*, ed. by S. Korzennik, A. Wilson (ESA Publications Division, Noordwijk, 1998), p. 635
J.M. Jensen, T.L. Duvall Jr., B.H. Jacobsen, J. Christensen-Dalsgaard, Astrophys. J. **553**, 193 (2001)
A.G. Kosovichev, T.L. Duvall Jr., in *SCORe'96: Solar Convection and Oscillations and Their Relationship*, ed. by F.P. Pijpers, J. Christensen-Dalsgaard, C.S. Rosenthal, Astrophysics and Space Science Library, vol. 225, 1997, pp. 241–260
A.G. Kosovichev, T.L. Duvall Jr., P.H. Scherrer, Sol. Phys. **192**, 159 (2000)
O.V. Ladenkov, F. Hill, S.A. Egamberdiev, D.Y. Chou, Astron. Lett. **28**, 411 (2002)
C.H. Lin, S. Basu, L. Li, Sol. Phys. (2008, in press). arXiv:0809.1427
C. Lindsey, D.C. Braun, Astrophys. J. **485**, 895 (1997)
C. Lindsey, D.C. Braun, Sol. Phys. **192**, 261 (2000)
H. Moradi, S.M. Hanasoge, P.S. Cally (2008). arXiv:0808.3628
K. Parchevsky, J. Zhao, A.G. Kosovichev, Astrophys. J. **678**, 1498 (2008)
E.N. Parker, Astrophys. J. **230**, 905 (1979)
S.P. Rajaguru, A.C. Birch, T.L. Duvall Jr., M.J. Thompson, J. Zhao, Astrophys. J. **646**, 543 (2006)
S.P. Rajaguru, K. Sankarasubramanian, R. Wachter, P.H. Scherrer, Astrophys. J. **654**, 175 (2007)
M. Rempel, M. Schüssler, M. Knölker, Astrophys. J. (2008). arXiv:0808.3294
A. Schlüter, S. Temesváry, in *Electromagnetic Phenomena in Cosmical Physics*, ed. by B. Lehnert, IAU Symposium, vol. 6, 1958, p. 263
H. Schunker, P.S. Cally, Mon. Not. R. Astron. Soc. **372**, 551 (2006)
M. Schüssler, M. Rempel, Astron. Astrophys. **441**, 337 (2005)
S.K. Solanki, Astron. Astrophys. Rev. **11**, 153 (2003)
H.C. Spruit, T.J. Bogdan, Astrophys. J. **391**, 109 (1992)
J.H. Thomas, N.O. Weiss, in *NATO ASIC Proc. 375: Sunspots. Theory and Observations*, ed. by J.H. Thomas, N.O. Weiss, 1992, pp. 3–59
M.J. Thompson, S. Zharkov, Sol. Phys. **251**, 225 (2008)
M.F. Woodard, Astrophys. J. **485**, 890 (1997)
J. Zhao, A.G. Kosovichev, T.L. Duvall, Astrophys. J. **557**, 384 (2001)

Small-Scale Solar Magnetic Fields

A.G. de Wijn · J.O. Stenflo · S.K. Solanki · S. Tsuneta

Originally published in the journal Space Science Reviews, Volume 144, Nos 1–4, 275–315.
DOI: 10.1007/s11214-008-9473-6

Abstract As we resolve ever smaller structures in the solar atmosphere, it has become clear that magnetism is an important component of those small structures. Small-scale magnetism holds the key to many poorly understood facets of solar magnetism on all scales, such as the existence of a local dynamo, chromospheric heating, and flux emergence, to name a few. Here, we review our knowledge of small-scale photospheric fields, with particular emphasis on quiet-sun field, and discuss the implications of several results obtained recently using new instruments, as well as future prospects in this field of research.

Keywords Sun · Photosphere · Magnetism · Small scale

1 Introduction

Magnetism on the Sun occurs on all scales. It manifests itself at the largest scales as a mean-field component that covers an entire hemisphere, and on progressively smaller scales as active regions, sunspots, and pores. Magnetic field in the lower solar atmosphere has structure on the smallest observable scales, up to the diffraction limit of the best telescopes.

A.G. de Wijn (✉)
University Corporation for Atmospheric Research, P.O. Box 3000, Boulder, CO 80307-3000, USA
e-mail: dwijn@ucar.edu

J.O. Stenflo
Institute of Astronomy, ETH Zurich, 8093 Zurich, Switzerland
e-mail: stenflo@astro.phys.ethz.ch

S.K. Solanki
Max-Planck-Institut für Sonnensystemforschung, Max-Planck-Strasse 2, 37191 Katlenburg-Lindau, Germany
e-mail: solanki@mps.mpg.de

S. Tsuneta
National Astronomical Observatory, Mitaka,Tokyo 181-8588, Japan
e-mail: saku.tsuneta@nao.ac.jp

M.J. Thompson et al. (eds.), *The Origin and Dynamics of Solar Magnetism*.
DOI: 10.1007/978-1-4419-0239-9_16

Theoretical arguments and simulations indicate that there is structure well beyond what can be observed today or in the foreseeable future.

A comprehensive ab-initio model of magnetic activity is currently impossible from a practical standpoint, and will remain so in the near future. The complex interaction of magnetic field, hydrodynamics, and radiative transfer requires sophisticated numerical analysis. A simulation would have to cover a substantial surface area over a good fraction of the convection zone in order to capture large-scale patterns such as supergranulation, yet also have sufficient resolution to capture interactions on scales of several kilometers or less. Such a simulation is prohibitively expensive in terms of computation time. In order to gain understanding of the physical processes involved in the creation, evolution, and eventual destruction of magnetic field, we must turn to observations to study the properties of magnetic structures.

Granular flows in the photosphere expunge field from cell interiors. Flux is swept into the intergranular lanes, where it clumps in small concentrations of mostly vertical field with strengths in excess of one kilogauss. Bright points and faculae, the most conspicuous features of magnetism in the lower solar atmosphere, correspond to these small concentrations of field. They are well known in active regions, where they group together in plages. In the quiet sun, supergranular flows concentrate them in the magnetic network that incompletely outlines supergranular cells.

Internetwork quiet-sun magnetism has been somewhat ignored historically, largely due to a lack of observations with sufficient resolution and accuracy. However, it has attracted particular interest in the past years, and this subject is currently being studied vigorously. Isolated concentrations of strong field that produce bright points and faculae also exist in supergranular interiors. Outside the concentrations, much weaker field that is not predominantly oriented perpendicular to the surface is ubiquitously present. This more horizontal field typically does not produce bright points that are easily observed using proxy-magnetometry diagnostics such as imaging in the Fraunhofer G band. Instead, sensitive magnetometers are required to observe and study weak field. The development of new instrumentation and seeing-mitigating techniques (the SpectroPolarimeter instrument on the space-borne observatory Hinode is an excellent example of both), and advanced simulations facilitated by the steadily increasing processing power of computers have made it possible to study this subject in detail.

Here, we focus our attention on magnetic fine structure of the quiet solar photosphere. In particular, we will discuss internetwork field. Quiet-sun internetwork areas cover the majority of the solar surface. Four orders of magnitude more flux emerges in the internetwork than in active regions. Consequently, field in these areas is of importance in understanding certain aspects of solar magnetic activity, such as the existence and workings of a granular dynamo and the dynamical coupling of the photosphere to higher layers.

We aim to provide a comprehensive review of quiet-sun internetwork magnetic fine structure, starting with a general overview in Sect. 2. We then address several small-scale phenomena that have recently attracted particular attention as a result of new, high-resolution observations: properties of horizontal field in the photosphere (Sect. 3), polar field (Sect. 4), and concentrations of strong vertical field (Sect. 5). Section 6 concludes the chapter with a discussion on unresolved fields.

2 Quiet Sun Magnetic Fields

The magnetic field found in the quiet sun can be categorized into network and internetwork field. The latter were discovered by Livingston and Harvey (1971, 1975) on the basis of a

weak Stokes V signal found in the interiors of supergranular cells. A separate category, the turbulent field, has also been proposed. It is not clear which of the further proposed types of quiet-sun magnetic fields, such as 'granular fields' (Lin and Rimmele 1999), horizontal quiet-sun fields (Lites et al. 1996), or 'seething fields' (Harvey et al. 2007) describe independent types of magnetic structures, and which are just different names for the same physical entity, detected in different types of observational data. The different techniques used to detect and study them pose the main difficulty with identifying one with the other. E.g., internetwork fields have traditionally been measured using the Zeeman effect, while the turbulent field has been probed mainly through the Hanle effect. Because of the cancellation of the Zeeman signal in the presence of opposite-polarity longitudinal fields in the resolution element, a tangled field may largely escape detection through the Zeeman effect, especially if the field is intrinsically weak. Only the larger scales of such a turbulent field would be seen using Zeeman-based diagnostics. These may then appear like internetwork elements, which in this scenario would represent just the tip of the iceberg of the Sun's turbulent field.

There have also been suggestions that the magnetic fluxes of all magnetic features in the photosphere form similar patterns irrespective of the scale at which they are observed. This scale invariance is consistent with the proposal that the magnetic field forms a fractal (or multifractal) pattern at the solar surface (e.g., Roudier and Muller 1987; Lawrence et al. 1995; Komm 1995; Nesme-Ribes et al. 1996; Meunier 1999, 2004; Stenflo and Holzreuter 2002, 2003; Abramenko 2005; Criscuoli et al. 2007). Since magnetic features are moved around by the evolving convection cells, possibly such an analysis provides more information on the distribution of convection at different scales, rather than on intrinsic magnetic properties. Convective eddies are expected to be self-similar for a turbulent medium, such as the solar convection zone.

2.1 Magnetic Flux in the Quiet Sun

2.1.1 Methods

In principle, it is possible to detect magnetic features and partly to estimate their magnetic flux in a variety of ways. However, the different types of measurements give different results, so that some uncertainty remains on just how much magnetic flux the quiet sun harbors.

Contrasts in more or less narrow wavelength bands are widely used as proxies of the magnetic field, since they are easy to observe at high resolution even under variable seeing conditions. They include the brightness in the G band, Ca II H or K line core, or CN bandhead. These proxies are, however, not ideal for determining the magnetic flux in the quiet sun, due to their small sensitivity. There is still some uncertainty to what extent internetwork magnetic features produce visible signatures in these proxies (however, see De Wijn et al. 2005).

The Zeeman effect not only provides quantitative measurements of the magnetic vector, but is also much more sensitive to small amounts of magnetic flux and has been shown to sense fluxes as low as 10^{16} Mx (or even less), particularly if the field is aligned along the line of sight (i.e., well visible in Stokes V, the net circular polarization). It suffers, however, from the fact that Stokes V is also sensitive to the direction in which the flux points (towards or away from the observer), so that if there is a mixture of polarities on a sufficiently small scale, the signal in Stokes V can be canceled. In Stokes Q and U cancellation, although possible, is less likely (it requires two transverse fields at right angles to each other in the resolution element).

If the aim is to measure intrinsically weak, possibly turbulent fields, then the Hanle effect is the method of choice. Basically, the Hanle effect allows the magnetic vector to be determined if the field strength lies within a fiducial range that depends on the observed spectral line. The Hanle effect is generally sensitive to low intrinsic field strengths (typical values are below a few 100 G, depending on the spectral line). Of importance for the field in the quiet sun is that the Hanle effect allows a weighted average of the field strength to be obtained even for a field that is isotropically distributed in the resolution element. Such a field would be invisible to the Zeeman effect as long as it doesn't produce any significant broadening of the line profiles (see below).

2.1.2 Measurements of Magnetic Flux in the Quiet Sun

The flux distribution in network elements has been determined by, e.g., Meunier et al. (1998) and Hagenaar (2001) (cf. Schrijver et al. 1997). They all find an exponential increase in the number density of elements with decreasing flux, down to the sensitivity limit (lying at 2×10^{18} Mx for the investigation of Hagenaar 2001). In contrast to this result, Wang et al. (1995) obtain a non-exponential, non-power law distribution for the network fluxes and a different (but also non-exponential, non-power law) distribution for the internetwork field. They use a series of criteria to differentiate between the two, including location (at the edges of supergranules or in their interior), proper motion speeds (higher speed of internetwork elements), etc. The weakest fluxes of individual internetwork features that they record are 10^{16} Mx.

Zirin (1987) found that the rate of magnetic flux emergence in internetwork fields is roughly 100 times larger than in ephemeral active regions. In the latter it is another 100 times higher than in normal active regions. Therefore, the internetwork fields completely dominate the flux emergence. However, whether the internetwork fields dominate the total flux at any given time depends on the ratio of emergence time scale to decay time scale of the fields. Intranetwork fields not just emerge at the highest rate, but also decay the most rapidly, so that their exact contribution to the instantaneous total magnetic flux is still unclear.

Prior to the Hinode mission (Kosugi et al. 2007), the typical average field strength in the quiet sun obtained from Zeeman effect measurements were a few gauss (typically 2–5 G). The estimates of Domínguez Cerdeña et al. (2003a, 2003b) and Khomenko et al. (2005a) (cf. Khomenko et al. 2005b) count as exceptions. At a spatial resolution of 0.5″, Domínguez Cerdeña et al. (2003a, 2003b) obtained an average field strength of 20 G in the internetwork. Khomenko et al. (2005a) compared the distributions of Stokes V amplitudes simultaneously observed in the infrared and the visible with the amplitudes of synthetic profiles computed in snapshots of mixed-polarity 3D MHD simulations harboring different amounts of magnetic flux. The MHD simulations of Vögler et al. (2005) can be studied at a spatial resolution nearly an order or magnitude higher than the observations, so that mixed polarity magnetic flux that is canceled out in the observations can still be counted in the simulations. The magnetic flux in the simulation snapshot that gives the best fit to the Stokes V amplitude distributions of both, the infrared and the visible lines, is taken to represent the solar flux. In this manner, to first order, the problem that a part of the flux is canceled within each spatial resolution element of the observations is circumvented. Just like Domínguez Cerdeña et al. (2003a, 2003b), Khomenko et al. (2005a) also obtained 20 G, but this value refers to a spatial resolution corresponding to the grid scale of the MHD simulations, a few 10 km. Therefore, unless there are no magnetic structures below 0.5″ in size, the value found by Khomenko et al. (2005a) is not consistent with the same value found by Domínguez Cerdeña et al. (2003a, 2003b). From high resolution observations obtained

by the Hinode Solar Optical Telescope (SOT, Tsuneta et al. 2008b; Suematsu et al. 2008; Ichimoto et al. 2008; Shimizu et al. 2008) and at the Swedish Solar Telescope we know that smaller-scaled structures are quite common.

Recently, analysis of Hinode spectropolarimeter (SP) data by Lites et al. (2008) has yielded 11 G for the longitudinal field. The fact that Hinode data (at the significantly higher and stable resolution of 0.32″) reveal only half as much flux as the investigation of Domínguez Cerdeña et al. (2003a, 2003b) suggests that fluctuations due to seeing may have affected the data of Domínguez Cerdeña et al. (2003a, 2003b) in a way that it increased the strength of the V signal, which is quite conceivable (cross-talk from brightness or velocity into Stokes V can happen quite easily).

An important question is how much flux could be hidden below the spatial scales that can be resolved? Such a hidden field (i.e., field not visible in magnetograms) is generally referred to as turbulent field, since in order to be invisible the two magnetic polarities must be mixed at scales below the spatial resolution element. Although very often no clear distinction is made to internetwork fields (which often also show a nearly random distribution of opposite polarities), we could consider internetwork fields as the large-scale and hence roughly resolved parts of the "turbulent" field, some fraction of which remains unresolved. However, given our current knowledge, we cannot rule out that the latter is physically different from the internetwork fields in some important aspect. Unno (1959) was the first to look for an unresolved "turbulent" field (that to first order was expected to be isotropic). Using differential line broadening he was able to set an upper limit of 300 G on such a field. Stenflo and Lindegren (1977) and later Stenflo (private communication) greatly improved the sensitivity of the technique by extending the investigation to hundreds of spectral lines (all the unblended Fe I lines in the visible solar spectrum), resulting in an upper limit of 100 G for the field outside the network, which includes the area-weighted contribution of the internetwork field and of any turbulent field.

Early work on the determination of unresolved magnetic flux using the Zeeman effect was also carried out by Stenflo (1987), who analyzed Stokes I, Q, and V profiles and set limits on a combination of magnetic field inclination and field strength. Tarbell et al. (1979) used high spatial resolution observations to circumvent the problem of cancellation of Stokes V by opposite polarity fields. They found that a possible turbulent field cannot exceed 100 G at spatial scales accessible to observations with a spatial resolution of 0.5″.

From the Hanle depolarization of the resonant polarization of lines formed in the quiet sun's photosphere (mainly from the Sr I line at 460.7 nm, but also from molecular lines), a turbulent magnetic field in the range of roughly 10–60 G has been inferred (Stenflo 1982; Faurobert-Scholl et al. 1995; Faurobert et al. 2001; Stenflo et al. 1998; Berdyugina and Fluri 2004; Trujillo Bueno et al. 2004; Bommier et al. 2005, 2006; Derouich et al. 2006).

With time, the investigations have increased in sophistication, now including multidimensional polarized radiative transfer and atmospheres produced by 3D radiation-hydrodynamic simulations. In general, a field of this average strength covering the whole quiet sun harbors less magnetic energy than the field in the network. Sánchez Almeida (2005) has argued, however, that (under certain assumptions) the measurements made in the Sr I line actually imply that more than half of the Sun's surface is covered by fields stronger than 60 G, even if the measurements give average field-strength values below 60 G.

Trujillo Bueno et al. (2004) also favor a higher energy density in the internetwork than in the network field (deduced from observations obtained at IRSOL by Stenflo et al. 1997). They adopt an exponential probability distribution function (PDF) for the field strength, as derived from MHD simulations. For a single PDF of the magnetic field, they find an e-folding width of 130 G (deduced from the same observations as give a 60 G average

field). Finally, they introduced different PDFs of the field in granules and intergranular lanes, with $B_0 = 15$ G in the former structures and with $B_0 = 450$ G in the latter, in order to simultaneously satisfy Sr I (atomic) and C_2 (molecular) lines. The energy density in the turbulent field in this scenario is larger than in the network. Further work on this topic, e.g., which tests the assumptions made by Trujillo Bueno et al. (2004), would be of considerable interest.

2.2 Magnetic Field Strength of Quiet Sun Fields

One question that has led to a partly heated debate over the last decade has been whether the magnetic fields in the internetwork quiet sun are intrinsically weak or strong. The magnetic field in the network has long been known to have an intrinsic strength on the order of a kilogauss (e.g., Stenflo 1973; Wiehr 1978; Solanki and Stenflo 1984; Stenflo and Harvey 1985; Stenflo et al. 1987; Solanki et al. 1987; Rabin 1992a, 1992b; Rüedi et al. 1992; Grossmann-Doerth et al. 1996, etc.) although a few advocates of weak fields, even in plage and the network remained. Thus, Zirin and Popp (1989) argued that the highly Zeeman sensitive Mg I lines at 12.3 μm only show weak fields, so that there are no strong fields in the network or in plages (except in occasional micro-pores). However, detailed radiative transfer modeling of these lines by Bruls and Solanki (1995) has shown that they are formed just below the temperature minimum in plage. At this height, due to pressure balance with the surrounding gas, the field, which in the middle and lower photosphere is well over a kilogauss, has dropped to only a few hundred gauss. The observations of Zirin and Popp (1989) actually provided confirmation of the simple model of slender flux tubes (e.g., Spruit 1976), if extended to take into account the merging of neighboring features (Pneuman et al. 1986; Steiner et al. 1986).

More recently, the debate on the intrinsic strength of quiet sun fields has been rekindled, but now concentrating on the internetwork fields. The intrinsic field strength is much more difficult to measure accurately than the magnetic flux per feature, since the Zeeman splitting often gives a non-unique result, except for kilogauss fields that fill a sufficiently large part of the aperture. Here, measurements in the infrared have an advantage, since the ratio of Zeeman splitting to Doppler width scales roughly linearly with the wavelength. It is therefore not so surprising that intrinsically weak fields in the lower photospheric layers were initially observed in the infrared at 1.56 μm (e.g., Rüedi et al. 1992). It also explains why studies of the strength of internetwork fields that employ infrared data (all have used the Zeeman sensitive line pair at 1.56 μm) give consistent results: the field strength of most internetwork features lies below roughly 600 G (Lin 1995; Solanki et al. 1996; Khomenko et al. 2003, 2005b; see also Lin and Rimmele 1999; Martínez González et al. 2007). These field strengths are partly consistent with equipartition between magnetic energy density and convective energy density, although they also provide evidence for a partial convective collapse (Solanki et al. 1996).

Observations of spectral lines in the visible have given rather varied intrinsic strengths of internetwork fields. Partly the results depend on the employed spectral lines, but they can also differ between studies using the same set of lines. An initial investigation by Keller et al. (1994) employing Stokes V measurements of Fe I 525.02 nm and Fe I 524.71 nm could not determine the true field strength, but provided evidence for a field strength below a kilogauss. Interest in these lines has been dormant until very recently when Khomenko and Collados (2007) and Socas-Navarro et al. (2008) have studied them in comparison with the more widely used 630.25 and 630.15 nm line pair as well as the 1.56 μm lines.

Most widely used have been the Fe I line pair at 630.25 and 630.15 nm which have been observed by, e.g., the Advanced Stokes Polarimeter (ASP, Elmore et al. 1992) and

now the Hinode SP. Magnetograms in these lines, recorded with the Göttingen Fabry-Perot at the VTT, have been investigated by Domínguez Cerdeña et al. (2003a, 2003b). They found magnetic flux throughout the quiet sun (covering 45% of the surface area, typically located in the intergranular lanes). In addition, this flux, which corresponds to most of the flux in the internetwork, was found to be in the form of kilogauss fields. Spectroscopic investigations employing ASP data by Lites and Socas-Navarro (2004) did not reproduce the preponderance of strong fields, while the analysis of Socas-Navarro and Lites (2004) indicated a mixture of strong and weak fields (cf. Socas-Navarro and Lites 2004). Sánchez Almeida et al. (2003), from a comparison of visible and infrared lines, also found a mixture of field strengths, although in their case the visible lines gave strong fields, while the infrared lines indicated weak ones (see further below for a more detailed discussion of this result).

Most recently, these lines, as recorded by Hinode/SP, have been analyzed by Orozco Suárez et al. (2007a, 2007b). In contrast to earlier authors, they obtained weak fields with strengths in the range of equipartition with the convection. Note that in contrast to, e.g., Domínguez Cerdeña et al. (2003a, 2003b), they inverted the full Stokes vector.

Finally, a Zeeman-effect based diagnostic has been developed by López Ariste et al. (2002, 2006). It makes use of the change in line profile shape introduced by hyperfine structure in Mn I lines in the visible part of the spectrum (553 nm) as the field strength increases. Applying this diagnostic to measurements of Stokes I and V in the internetwork they obtain mainly hectogauss fields (which cover the majority of the area and contain the majority of the flux), although they do not give precise numbers regarding the field strength. This result is confirmed by Asensio Ramos et al. (2007) employing a Mn I line in the infrared (at 1.5262 μm). Sánchez Almeida et al. (2008), however, argue that in a MISMA-like atmosphere (Micro-Structured Magnetic Atmosphere, an approximation to describe the influence on the Stokes profiles of an atmosphere with the magnetic field structured at a very small, optically thin scale; Sanchez Almeida et al. 1996) the Mn I line at 553.8 nm will indicate weak fields even if more than 50% of the magnetic flux is in the form of kilogauss fields.

The difference between the results obtained with the infrared 1.56 μm and those from the visible 630.2 nm lines has fueled the aforementioned debate on the true field strengths of internetwork fields. It has also led the groups using either one of these diagnostics to comment on the shortcomings of the other. For example, it has been argued that the visible lines miss much of the weak fields, since for incomplete Zeeman splitting (which is the case for these lines for sub-kilogauss fields) the signal in a given pixel is proportional to the magnetic flux in that pixel. Since the internetwork fields are associated with very small fluxes per pixel, these lines could miss a considerable portion of it. Also, because intrinsically weak fields change the shapes of the I and V profiles only in subtle ways, the deduced values are susceptible to noise or systematic errors. Conversely, it has been argued by Socas-Navarro and Sánchez Almeida (2003) that the infrared lines, by dint of their large Zeeman sensitivity, give too much weight to the weak fields. For these lines, the amplitude of the Stokes V signal is proportional to the fractional area covered by the field (the magnetic filling factor) rather than to the amount of magnetic flux in the pixel. Therefore, fields with a low strength give proportionately stronger signals (for, e.g., an equal amount of magnetic flux in intrinsically weak and intrinsically strong fields). Socas-Navarro and Sánchez Almeida (2003) argue that the rapid drop of the field strength with height (due to pressure balance) compounds this effect: since a spectral line is formed over a range of heights, this gradient of the field spreads the signal in the wavelength direction. Since the intrinsically strong fields are associated with the largest vertical field-strength gradients, the smearing in the wavelength direction is largest for such fields, making the Stokes amplitudes small and possibly hidden in the noise. Consequently, they argue, the infrared lines are missing much of the flux in the strong fields.

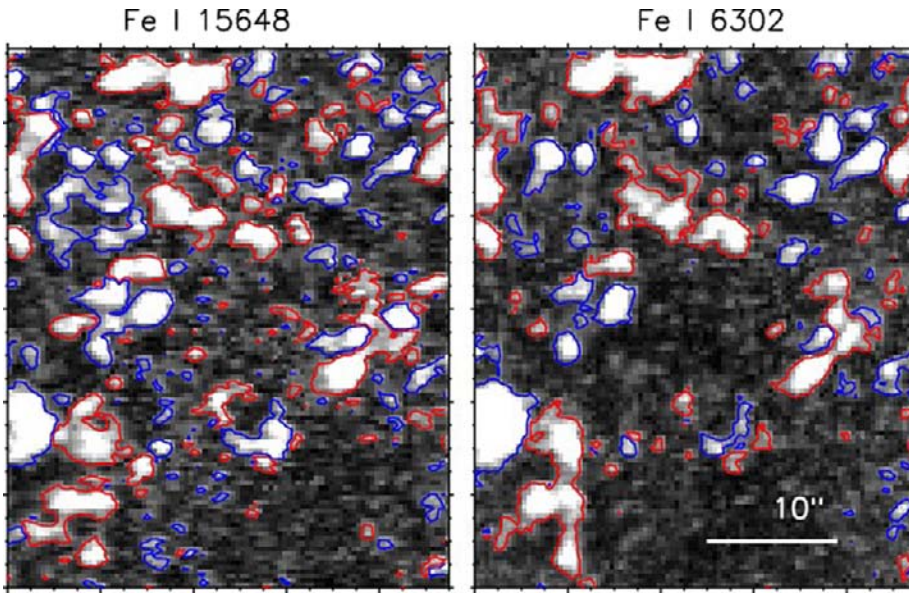


Fig. 1 Magnetograms in the infrared at 1.56 μm (*left panel*) and in the visible at 630.2 nm (*right panel*) obtained simultaneously and co-spatially with the VTT on Tenerife. Greater brightness indicates larger amounts of magnetic flux per pixel. Opposite magnetic polarities are bounded by *red* and by *blue lines*, respectively. Adapted from Khomenko et al. (2005a)

A comparison of the results obtained from the infrared and the visible lines of a simultaneously observed patch of quiet sun might be a way of deciding between the different diagnostics and associated points of view. Such a comparison was first carried out by Socas-Navarro and Sánchez Almeida (2003), who found that the Stokes maps in the two wavelength ranges looked quite different. In particular, they noted that the visible and infrared lines displayed opposite polarities in 25% of the pixels, which was a remarkably high proportion. If correct, this would indeed support the view that the infrared and visible lines were sampling rather different components of the internetwork field. The main drawback with this investigation was that data from different telescopes had been used, so that the seeing quality of the two data sets was not comparable, making their comparison less straightforward (Lites 1987). In a later analysis, Khomenko et al. (2005a) compared visible and infrared lines observed with the same telescope under identical seeing conditions and obtained a more similar distribution of polarities and fluxes from both wavelength ranges. The magnetograms obtained in both wavelength ranges are shown in Fig. 1. Remaining differences between the two images are due to the larger sensitivity of the infrared line to weak fields and to the remaining unavoidable differences in seeing (which possesses a dependence on λ).

Finally, Domínguez Cerdeña et al. (2006a) inverted a set of combined infrared and visible spectra using a 3-component model, which allowed a field-free component to co-exist with two different magnetic components. They obtained a mixture of strong and weak fields, with a clear relationship between magnetic field strength and magnetic flux in the sense that the larger the magnetic flux in a pixel, the stronger the field (left panel of Fig. 2). This result is similar to that found by Solanki et al. (1996), based purely on 1.56 μm spectropolarimetry, shown in the right panel of Fig. 2. An increase of field strength with magnetic flux of the

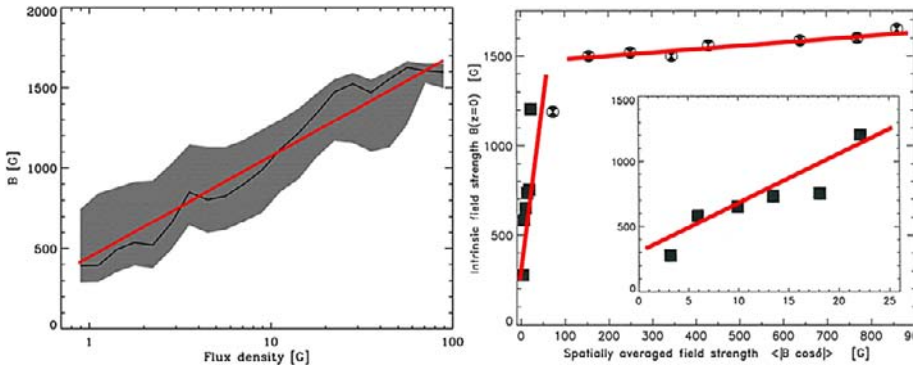


Fig. 2 *Left panel*: field strengths retrieved from a combination of 1.56 μm and 630.2 nm co-spatial observations obtained nearly simultaneously. Adapted from Domínguez Cerdeña et al. (2006a). *Right panel*: same, but from an investigation of the 1.56 μm lines alone. The results of observations covering both the quiet sun and active regions are displayed in the larger frame. In the *inset* only the results for the quiet sun are shown (each point is a value binned over numerous individual data values). The *left red line* in the main frame is drawn considering more data than in the *inset* (*circle* at $(\langle B\cos\gamma\rangle, B) = (70, 1200)$) and is not identical with the *red line* in the *inset*. Adapted from Solanki et al. (1996)

feature is in agreement with predictions of the efficiency of the convective collapse mechanism that leads to the formation of the intense flux tube (Venkatakrishnan 1986). As in their earlier papers, Domínguez Cerdeña et al. (2006a) argue that most of the flux and of the magnetic energy is in the kilogauss fields.

Many of the investigations discussed so far have been based on Stokes I and V profiles only. This can be explained partly by instrumental constraints, partly by the fact that the Stokes Q and U profiles scale as B^2, while Stokes V scales proportionally to B. For a relatively weak Zeeman splitting (typical of visible lines in the quiet sun) this implies that Q and U are much weaker than V.

The difficulty of measuring the field strength reliably from just I and V of a visible line pair, in particular from Fe I 630.2 and 630.1 nm, has been demonstrated by Martínez González et al. (2006). They fit a set of these line profiles two times, once starting from a strong-B initial guess, once from a weak-B initial guess. The final result depended strongly on the initial guess, although the fits to the profiles were equally good. The differences in the Stokes profiles produced by the different field strengths were completely compensated by slightly different temperatures and turbulence velocity values returned by the inversion code. Another test was carried out by Khomenko and Collados (2007). They used the output atmospheres from the 3D radiation-MHD simulations of Vögler et al. (2005) to test a number of diagnostics of the field strength. According to their analysis the most reliable of the tested diagnostics is the 1.56 μm line pair, the least reliable the 630.2/630.1 nm line pair. These exercises have demonstrated just how difficult it is to obtain reliable B values from this latter line pair, in particular if only Stokes I and V are available. Consequently, results obtained from such data have to be interpreted with caution.

The great advantage of also having a linear polarization profile available is that the shape of the Q and U profiles changes with the field strength, in the sense that the ratio of the strength of the π-component to the σ-components depends on B, providing further (although not in themselves unique) constraints on the field strength, as demonstrated by, e.g., Solanki et al. (1987).

More recently, the advent of Hinode has opened up new possibilities, by providing I, Q, U, and V spectra of Fe I 630.2 and 630.1 nm at a constant high spatial resolution corre-

sponding to approximately 0.3″. Recent inversions by Orozco Suárez et al. (2007a, 2007b) indicate that the Hinode data give mainly weak fields (hectogauss), possibly because of the additional constraints provided by the linear polarization signals (only pixels with profiles lying above a given threshold in Stokes Q, U, and V are inverted).

2.3 Horizontal Fields in the Internetwork

Evidence for horizontal fields in the internetwork can be noted already in data published by Martin (1988): in these magnetograms internetwork fields are visible from the center of the solar disk right to the limb, suggesting the presence of both vertical and horizontal fields. With considerable foresight, Martin interpreted these measurements as possibly due to the presence of low-lying loops in the internetwork.

Lites et al. (1996) found arcsecond scale, short-lived horizontal fields (lifetimes of minutes) in the internetwork. The size scale was determined by their spatial resolution. Meunier et al. (1998) considered the center-to-limb variation of the Stokes V amplitude of the $g = 3$ line at 1.56 μm from which they concluded that the quiet sun field is composed mainly of intrinsically weak, nearly isotropically distributed fields, in addition to strong, nearly vertical fields. Martínez González et al. (2008) also found evidence for a more or less isotropic distribution of the internetwork field (and little change in the field strength probability distribution function) from the center-to-limb variation of the polarization signal in the quiet sun, in agreement with Martin (1988) and Meunier et al. (1998). With the very sensitive SOLIS instrument on Kitt Peak, Harvey et al. (2007) deduced a "seething" horizontal field throughout the internetwork. This field of typically 1–2 G at the spatial resolution of SOLIS of 2.5–5″ changed within minutes. Further evidence for horizontal fields has been provided by Hinode: Orozco Suárez et al. (2007a, 2007b) inverted Stokes spectra to obtain a peak in the distribution of inclination angles of internetwork fields at 90°, which corresponds to horizontal fields. This interesting result may partly be an artifact of the higher sensitivity to noise of Stokes Q and U due to their weakness, unless fields are intrinsically strong. Finally, Lites et al. (2008) obtained 5 times more flux in horizontal fields than in the vertical fields in the internetwork (to be more specific: they found that the spatially averaged strength of the horizontal field is 5 times larger than of the vertical field; a precise determination of the flux for horizontal fields is rather difficult from Stokes parameters). With a strength of 50–60 G, it is comparable to the values obtained by the Hanle effect (see Sect. 6).

As the evidence for nearly horizontal internetwork fields increases, one question that comes to the fore is: what is the structure of these internetwork fields? From observations at 1.56 μm, Martínez González et al. (2007) concluded that at least some of the internetwork elements are (parts of) low-lying loop-like structures. The loops were reconstructed in a way similar to the technique applied by Solanki et al. (2003), although the 180° ambiguity inherent in the Zeeman-effect did not allow Martínez González et al. (2007) to distinguish between small Ω loops and U-loops at a granular scale. An example of a loop reconstructed by Martínez González et al. (2007) is shown in Fig. 3. These loops may correspond to the small-scale emerging loops observed by Centeno et al. (2007) in the quiet sun and by Ishikawa et al. (2008) in active region plage. These small loops carry a flux of approximately 10^{17} Mx each and are found to emerge in granules.

2.4 Source of Internetwork Fields

How could such a magnetic structure be explained? There are different possibilities.

1. Emergence of fields generated in deeper layers (e.g., by a deep convection-zone or an overshoot-layer dynamo). This could be the extension of ephemeral active-region fields

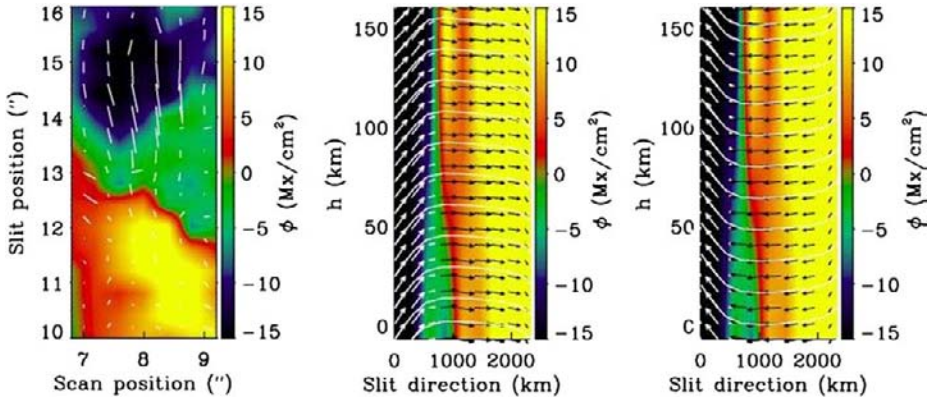


Fig. 3 Reconstructed loop in the internetwork. *Left panel*: vertical magnetic flux density in a small region of the total scan (flux density is indicated by the color) at height zero (average solar surface). The azimuthal direction of the field is overplotted. *Central panel*: the vertical dependence of the magnetic vector along a cut going from the upper part of the left frame to its lower part at the scan position marked 8″. The colors mark the magnetic flux density, while the direction of the magnetic vector is indicated by the *arrows*. The *white lines* are smoothed curves joining the arrows and outlining the loops. The *right panel* is the same, but now for the other solution allowed by the 180° ambiguity

(studied by Harvey and Martin 1973; Harvey et al. 1975; Harvey 1993; Hagenaar 2001) to still smaller scales. Note that there is a power-law distribution of flux in bipolar regions (following an inverse square law) from large active regions down to small ephemeral regions. The cutoff at the small scales is consistent with a lack of resolution and/or sensitivity. Whereas the large active regions have a strong tendency towards an E–W orientation following Hale's polarity law, increasingly smaller bipoles have increasingly weaker preferred orientations. Any lack of orientation of the smallest emerging bipoles therefore does not automatically rule out this scenario, since there is no abrupt transition, but rather a very gradual decrease of the level of orientation with decreasing area or magnetic flux.

2. Flux recycling after decay of active-regions and ephemeral active regions. The magnetic flux from a decaying region very likely partly gets dragged down by convection and can emerge again at another point on the solar surface. Such an effect has been identified in MHD simulations carried out by Ploner et al. (2001), suggesting that such recycling does take place. The work of De Wijn et al. (2005), see below, also provides evidence that either mechanism 1 or 2 (or some combination of both) is acting as the source of some of the flux in the internetwork (see Sect. 5).

3. Flux produced at or very close to the solar surface by a truly local dynamo. The first numerical experiments that sustained a local dynamo in a convective medium similar to the solar interior were carried out by Cattaneo (1999). The most realistic simulation of a local (solar surface) dynamo to date has been performed by Vögler and Schüssler (2007), who considered also a proper 3D radiative transfer etc. to simulate the conditions in the layers close to the solar surface. Starting from a very low value, the magnetic energy within the simulation box increases exponentially with time, before it saturates. The saturation value depends on the magnetic Reynolds number R_m of the simulation, being higher for larger R_m. For $R_m = 2600$ the simulations give an average, unsigned vertical field of approximately 35 G, which lies within the range of values found from the Hanle effect.

The field produced by such a simulation is structured on very small (subgranular) scales with strongly mixed opposite polarities, as can be seen from Fig. 4. It is also largely hor-

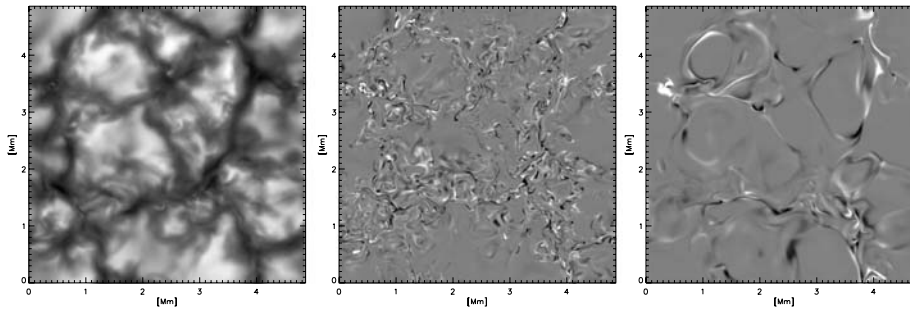


Fig. 4 Snapshot from a dynamo simulation run taken about 5 hours after introducing the seed field. The vertically emerging bolometric intensity (brightness, *left panel*) reveals a normal solar granulation pattern. The other panels show the vertical component of the magnetic field on two surfaces of constant (Rosseland) optical depth, τ_R. Near the visible surface (*middle panel*, $\tau_R = 1$, gray scale saturating at ± 250 G), the magnetic field shows an intricate small-scale pattern with rapid polarity changes and an unsigned average flux density of 25.1 G. About 300 km higher, at the surface $\tau_R = 0.01$ (*right panel*, gray scale saturating at ± 50 G), the unsigned average flux density has decreased to 3.2 G and the field distribution has become considerably smoother, roughly outlining the network of intergranular downflow lanes (*darker areas on the left panel*). Figure taken from Vögler and Schüssler (2007) by permission

izontal. It is basically composed of short, flat loops that are concentrated in intergranular lanes and generally have both their foot points within a single intergranular lane. Note that the simulations carried out so far do not allow any flux to be advected into the box (which may be the reason why relatively few larger-scale magnetic structures are visible). Note also that changes in R_m should have an influence on the magnitude of the produced magnetic field and energy, but not on its distribution, so that the shape of the PDF of the field strength and of the magnetic orientation should remain independent of R_m.

Any difference between the observed and simulated distribution of the flux may be telling us something about other effects besides a purely local dynamo acting to produce the observed field. Therefore, it is heartening that Schüssler and Vögler (2008) obtain a ratio between horizontal and vertical field that is close to the value found by Lites et al. (2008) from Hinode SP data. One difference between the two is that Lites et al. (2008) found most of their horizontal flux regions at the edges of granules, while simulations place the flux clearly in the intergranular lanes. A part of this difference may be due to the limited depth of the computation box.

Quite generally, there is an observed relationship between the weak quiet-sun fields and convective features. Best known is that the strong fields found in the network are located at the boundaries of supergranules. On a smaller scale, Lin and Rimmele (1999) find a weak field whose distribution is moulded by the granulation. The field also changes over a granular life-time (consequently they called this component of the field a granular field). Khomenko et al. (2003) find a preponderance of weak fields in the intergranular lanes, while Socas-Navarro et al. (2004) find that the field strength depends on the location of the field relative to the granule in a non-trivial manner. Arguments against the origin of at least the stronger internetwork flux from a local dynamo have been given by De Wijn et al. (2005) on the basis of the fact that this part of the flux is seen to be distributed on a mesogranular scale and displays a lifetime well in excess of that of granulation.

Such dependencies may (or may not) provide an indication of the origin of the magnetic flux. However, they do tell us that the flux must survive without complete cancellation for a sufficiently long time to be dragged to the edge of the particular convective feature it is

found to be lying at the boundary of. In the case of the network this implies a survival time of at least 10 hours, for the mesogranulation roughly an hour or two.

3 Transient Horizontal Magnetic Field

3.1 Properties of Horizontal Magnetic Field

Quiet-sun magnetism essentially consists of vertical flux tubes and horizontal magnetic fields. The field strength of vertical magnetic fields exceeds the equipartition field strength B_e of about 500 G, determined by $B_e = \sqrt{4\pi\,\rho\,v^2}$, where granules with a velocity of $v = 2 \times 10^5$ cm/s, and the plasma density $\rho = 3 \times 10^{-7}$ g/cm^3 at $\tau_{500} = 1$ are assumed. Hinode observations show that convective instability could be a mechanism used to explain the formation of such vertical flux tubes with kilogauss field strength (Nagata et al. 2008): the cooling of a flux tube at equipartition field strength precedes a transient downflow reaching 6 km/s and the intensification of the field strength to 2 kG. This is not a unique observation, but rather it is a ubiquitous phenomenon in the quiet sun.

The initial discovery of the horizontal magnetic field with ground-based telescopes was summarized in the previous section. High resolution spectroscopic observations with SOT/SP aboard Hinode have confirmed this finding and extended these studies considerably (Lites et al. 2008; Centeno et al. 2007; Orozco Suárez et al. 2007a, 2007b; Ishikawa et al. 2008; Ishikawa and Tsuneta 2009a; Tsuneta et al. 2008a). The horizontal magnetic field is highly intermittent in both the temporal and spatial domain: statistical study shows that the average life time of a horizontal field element is 4 min, and their size is smaller than the average size of the granular pattern (Ishikawa and Tsuneta 2009b). Thus, we hereafter call them elements of the "transient horizontal magnetic field" (THMF).

Figure 5 shows the histogram (i.e., PDF) of the degree of linear polarization (LP) for THMF in the quiet sun and a plage region. These two regions were located near the center of the solar disk. The degree of linear polarization is proportional to the square of the transverse magnetic field component. Vertical magnetic concentrations are masked, and thus are not included in the red histogram. This is a comparison of different areas of the Sun with different magnetic properties. The exact match of these two PDFs indicates that property of

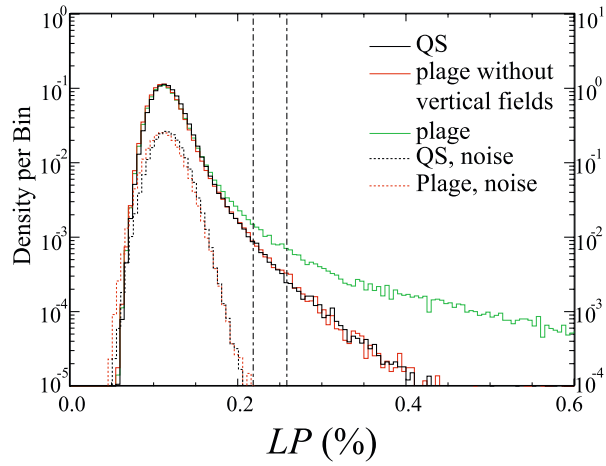


Fig. 5 Histograms of net linear polarization (LP) for plage and quiet sun. The *dotted lines* represent LP noise distributions for both datasets. The area dominated by vertical magnetic fields is masked in the plage region. The two *vertical dashed-dotted lines* indicate LP of 0.22% and 0.26%—the thresholds used in Fig. 3. From Ishikawa and Tsuneta (2009a)

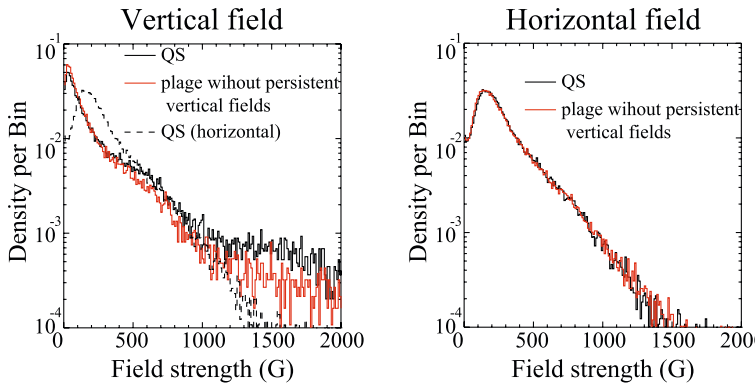


Fig. 6 PDFs of the magnetic field strength of vertical fields (*left panel*) and horizontal fields (*right panel*). *Black* and *red lines* are used for the quiet sun and the plage region, respectively. Apparent concentrations of the vertical magnetic fields are masked to obtain the PDF of the plage region. The *black dashed line* in the *left panel* shows the PDF for horizontal fields in the plage region, and is the same as the *black solid line* in *right panel*. Vertical fields refer to magnetic fields with inclination smaller than 20° or larger than 160°, and horizontal fields refer to magnetic fields with inclination larger than 70° and smaller than 110°. From Ishikawa and Tsuneta (2009a)

the THMF of the quiet sun and active regions is remarkably similar (Ishikawa and Tsuneta 2009a).

The magnetic landscape of the polar region is characterized by vertical kilogauss patches with super equipartition field strength, a coherency in polarity, and the ubiquitous weaker transient horizontal fields (Tsuneta et al. 2008a). We now know that THMFs are ubiquitous in plage regions, the quiet sun, and the extreme polar region.

The remarkably similar distributions of LP in Fig. 5 also suggest the same occurrence rates in both the quiet sun and the plage region. These occurrence rates are extremely high, as discussed by (Ishikawa et al. 2008). THMFs have lifetimes ranging from one minute to about ten minutes, comparable to the lifetime of granules. Among 52 events that they examined, 43 horizontal magnetic structures appear inside the granules, and four appear in inter-granular lanes, with the remaining five events ambiguous in position. Since 52 events are detected in the $2.5'' \times 164''$ observing area during the 40 minutes, a new event appears every 46 seconds in the same observing region. The turnover time of the granules is approximately 1000 s, with a velocity of 2 km/s and with a depth comparable to the horizontal scale of granules. There are approximately 182 granules in the observing area, assuming that the size of the granules is $1.5'' \times 1.5''$. 84% of these granules are not associated with stable strong vertical magnetic fields, and we use this smaller sample for estimating the frequency of events. If every granule were to have an embedded horizontal magnetic field structure, the horizontal field would have appeared at the surface every 6.6 s ($\sim$ 1000 s/152 granules) in the observing area. This shows that more than approximately 10% of the granules have embedded horizontal fields, suggesting a relatively common occurrence of THMFs (Ishikawa et al. 2008).

Figure 6 shows that PDFs of the intrinsic magnetic field strength for the quiet sun and the plage region are again almost identical, and the PDF of the extreme polar region (Fig. 9) is similar to those of the quiet sun and the plage region. This remarkable similarity suggests a common local dynamo process (Cattaneo 1999; Vögler and Schüssler 2007) taking place all over the Sun.

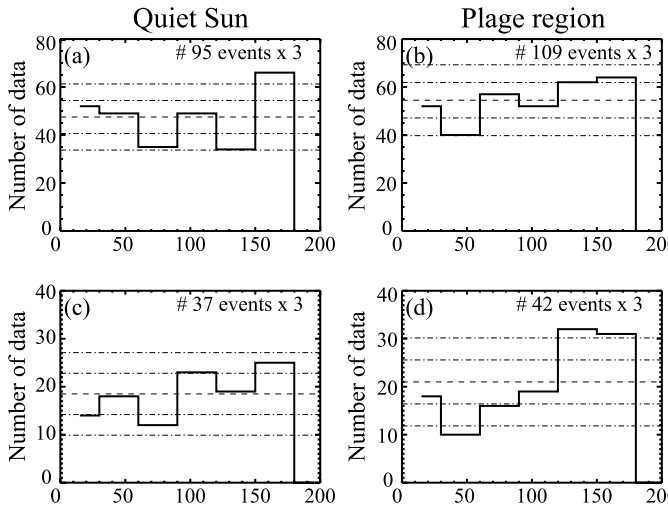


Fig. 7 The histograms of the field azimuth angles for THMFs. Three pixels with highest LP are taken from individual events, and number distributions of the azimuth angles of horizontal magnetic field for these pixels are plotted. The azimuth angle 0° is to the west, 90° to the north, and 180° to the east. Panels **a** and **b**: histograms of azimuth angle for 95 events in the quiet sun and 109 events in the plage region which have LP higher than 0.22%. Panels **c** and **d**: histograms of azimuth angle for 37 events in the quiet sun and 42 events in the plage region which have LP higher than 0.26%. The *dashed lines* indicate the case for a uniform distribution. *Dashed-dotted lines* closer to the *dashed line* show $\pm\sigma$, statistical deviation, and two other *dotted lines* show $\pm\,2\sigma$. From Ishikawa and Tsuneta (2009a)

To minimize the influence of noise in the Stokes inversion, we have analyzed only pixels whose polarization signal peaks exceed a given threshold above the noise level σ. The noise level was determined in the continuum wavelength range of the profiles. The fitting is performed for pixels whose Q, U, or V signals are larger than 4.5–5.0σ. Thus, the peaks in the PDFs at around 150 G may be an artifact: the THMFs that we observe are probably the tip of the iceberg due to our limited sensitivity, and there may be weaker but more ubiquitous magnetic fields unresolved by Hinode: the Sun's hidden magnetism inferred by, e.g., Trujillo Bueno et al. (2004) through Hanle-effect observations (see Sect. 6 for a more complete discussion).

Figure 6 indicates that 93% of horizontal magnetic fields have field strengths smaller than 700 G, and 98% smaller than 1 kG for both regions. A magnetic field strength of 700 G corresponds to the typical equipartition field strength just below the level of granules at a depth of 500 km, where the density is $\sim 10^{-6}$ g/cm^3 and the velocity is 2 km/s. Thus, the majority of horizontal fields have field strengths smaller than the equipartition field strength for average granular flows.

Figure 7 panels a and b show the magnetic field azimuth of THMFs for events with LP greater than 0.22% in the quiet sun and in the plage region discussed above. We define $\sigma = \sqrt{N}$, where N is a number of average events per 30° bin under the assumption of a uniform distribution of the azimuth of the horizontal fields. There is no statistically significant orientation in either region. If the LP threshold is $> 0.26\%$ (Fig. 7 panels c and d), we find a broad peak between 120° and 180°, and a dip between 30° and 60° that are significant at the 2σ level in the plage region. In contrast, these events in the quiet sun still show an azimuth angle distributed within 2σ of the uniform value. This peak angle corresponds to

the tilt angle of the bipolar plage region. This indicates that THMFs with higher LP in the plage region appear to be partially related to the global fields of the plage region.

3.2 THMF and Local Dynamo Process

The properties of THMFs are summarized to be: (1) an identical or similar PDF of magnetic field strength in the quiet sun, plage regions, and the extreme polar region; (2) ubiquitous occurrence all over the Sun including the extreme polar region; (3) a magnetic field strength essentially smaller than the equipartition field strength; and (4) no or weak preferred direction of the magnetic field vector.

The amount of the vertical magnetic flux in the plage in the case presented here is about 8 times larger than that of the quiet sun (Ishikawa and Tsuneta 2009a). If the THMF occurrence rate was in any way directly related to the global vertical fields forming the plage region, then we would expect the occurrence rate in the plage region to be much larger than that of the quiet sun. The similar occurrence rates we observe suggest that the emergence of the THMFs does not have a direct causal relationship with the vertical magnetic fields in the plage region. The same THMF occurrence rate, no preferred orientation, and similar field-strength distributions for both regions strongly suggest that a common local process that is not directly influenced by global magnetic fields produces THMFs (Ishikawa and Tsuneta 2009a). As (1) ubiquitous THMFs are receptive to convective motion (Centeno et al. 2007; Ishikawa et al. 2008), and (2) the field strength is essentially smaller than the equipartition field strength, a reservoir of THMFs may be located near solar surface, and these magnetic fields are carried to the surface through convective flow.

Such reservoir can be maintained by a local dynamo process due to near-surface convective motion (Cattaneo 1999; Vögler and Schüssler 2007). Indeed, numerical simulations have shown that a local dynamo can generate horizontal magnetic structures in the quiet sun (Abbett 2007; Schüssler and Vögler 2008). Such a local dynamo process could naturally explain the similarity in occurrence rates and field strength PDFs, including the fact that THMFs do not have a preferred orientation. The similarity in field-strength distribution also indicates that properties of THMFs do not depend on the seed field, e.g., global fields.

Other possibilities for the origin of THMFs include debris from decaying active region, magnetic fields that failed to emerge from the convection region to the photosphere (Magara 2001), and extended weak magnetic fields in the upper convection zone generated by "explosion" (Moreno-Insertis et al. 1995). If the reservoir is maintained by one of these processes, the THMFs would be expected to be affected by global toroidal fields in terms of properties of THMFs described above. Within the context of these simulations cited above, it may be difficult to explain the observed properties of THMFs such as the similarity in the occurrence rates and magnetic field distributions, and the lack of preferred orientation of THMFs.

A slight preferred orientation of THMFs with higher LP toward the global plage polarity suggests that these THMFs may be influenced by the global plage field. However, because any strong vertical fields associated with the emergence of these THMFs are not observed (Ishikawa et al. 2008), they are probably not directly created from the vertical magnetic fields forming the plage as suggested by Isobe et al. (2008). Thus, even if these THMFs with higher LP are related by the global toroidal system, the relationship would be indirect—the THMFs with high LP may result from fragmented elements of plage flux tossed about by the convective motions below the photosphere.

The evidence that a local dynamo is playing a significant role for the quiet sun magnetism comes from the Hanle-effect investigation by Trujillo Bueno et al. (2004), who inferred a

magnetic energy density which is the order of 20% of the kinetic energy density produced by the convective motions in the quiet solar photosphere, and showed that the observed scattering polarization signals do not seem to be modulated by the solar cycle. The papers using observations from Hinode cited here are providing us with multiple new pieces of evidence in favor of a local dynamo process taking place in the convective turbulent outer layer of the Sun.

4 Polar Field

The Sun's polar magnetic fields are thought to be the direct manifestation of the global poloidal fields in the interior, which serve as seed fields for the global dynamo that produces the toroidal fields responsible for active regions and sunspots. The polar regions are also the source of the fast solar wind. Although the polar regions are of crucial importance to the dynamo process and acceleration of the fast solar wind, its magnetic properties are poorly known. Magnetic field measurements in the solar polar regions have long been a challenge: variable seeing combined with the strong intensity gradient and the foreshortening effect at the solar limb greatly increases the systematic noise in ground-based magnetographs. Nevertheless, pioneering observations have been carried out for the polar regions (Tang and Wang 1991; Lin et al. 1994; Lites 1996; Homann et al. 1997; Okunev and Kneer 2004; Blanco Rodríguez et al. 2007). Many polar observations have also been restricted to individual polar faculae within a small field of view, and have not provided us with a global magnetic landscape of the polar region, with the exception of GONG/SOLIS (Harvey et al. 2007). Using SOT on board the Hinode spacecraft, it is possible to investigate the properties of photospheric magnetic field in polar regions with unprecedented spatial resolution, field of view, and polarimetric sensitivity and accuracy in measurements of vector magnetic fields. Such an analysis has recently been carried out by Tsuneta et al. (2008a).

4.1 The Polar Magnetic Landscape

Properties such as field strength, inclination, azimuth, filling factor, etc. may be estimated from the line profiles observed by Hinode using inversion codes. In this case, only pixels whose polarization signal exceeds 5σ above the noise level were analyzed using an inversion code that assumes a Milne-Eddington atmosphere.

Figure 8 is a map of the magnetic field strength as seen from above the south pole. Such a representation is needed to correctly see the spatial extent and size distribution of the magnetic islands in the polar region. While many of them are isolated, and some have the form of a chain of islands, complex internal structures are seen inside the individual patches. Many patchy magnetic islands have very high field strength reaching above 1 kG. They are coherently unipolar, and like plage and network fields at lower latitudes (Martinez Pillet et al. 1997), they have magnetic field vertical to the local surface.

Patches show a tendency to be larger in size with increasing latitude. The size is as large as $5'' \times 5''$ at higher latitudes and $1'' \times 1''$ at lower latitudes. Degradation in spatial resolution due to the projection effect may contribute to the larger size at high latitude. Expansion may also be caused because we observe flux tubes higher in the atmosphere close to the limb. The response function of the spectral lines observed here for a plane-parallel atmosphere viewed obliquely at an angle of 80° has a peak that is 50 to 100 km higher than if viewed straight down.

Close to the limb, it is possible to determine the inclination i of the magnetic field vector with respect to the local surface without the usual 180-degree ambiguity of the transverse

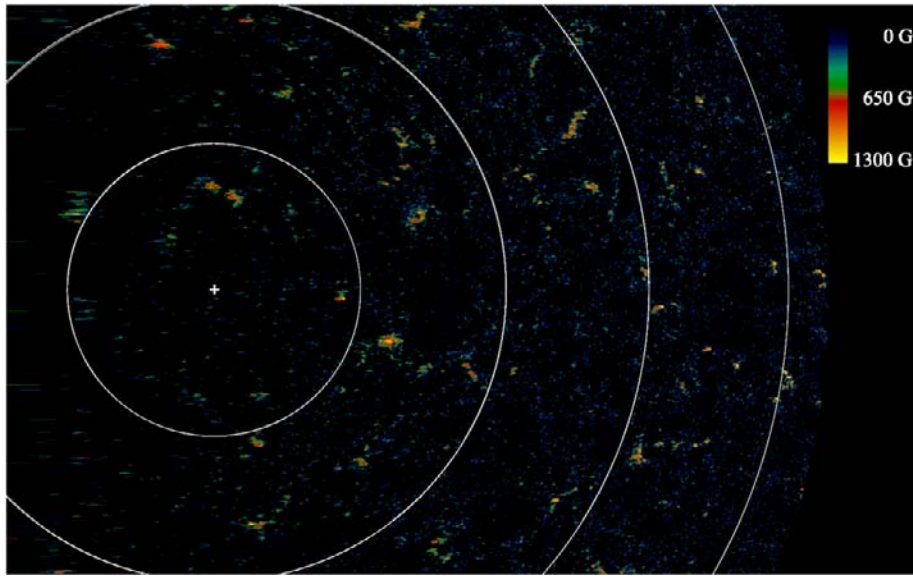


Fig. 8 View of the magnetic field strength in the south-polar region as seen from above the pole at 12:02:19–14:55:48 on March 16, 2007. East is up, and north is to the right. The original field of view of the observation is 328″ (east–west) by 164″ (north–south). Latitudinal lines for 85°, 80°, 75°, and 70° are shown as *white circles*, while the *cross mark* indicates the south pole. The spatial resolution is lost near the extreme limb (i.e., near the left of the figure). Magnetic field strength is obtained for pixels with a polarization signal exceeding 5σ above the noise level. From Tsuneta et al. (2008a)

field components (del Toro Iniesta 2003). All the large patches have fields that are vertical to the local surface, while the smaller patches tend to be horizontal. Most of the magnetic structures seen in Fig. 8 thus have either vertical or horizontal directions. These two types do not appear to be spatially correlated.

Magnetic patches of larger spatial extent coincide in position with polar faculae (Lin et al. 1994; Okunev and Kneer 2004). This is confirmed in panel c of Fig. 9. The distribution of local intensity is essentially symmetric around the average intensity for the horizontal fields, while the vertical fields tend to have higher continuum intensities.

Panel a of Fig. 9 shows the PDF of the magnetic field strength B for latitudes $> 75°$. Vertical magnetic fields with inclination $i < 25°$ dominate the stronger field regime, while horizontal fields with $i > 65°$ are much more prevalent below 250 G. A PDF of the magnetic energy is shown in Fig. 9 panel b. This shows that the vertical flux tubes with higher field strength are energetically dominant, while weaker horizontal flux tubes contrastingly carry more energy.

4.2 Total Magnetic Flux

The total vertical magnetic flux in the SOT field of view is 2.2×10^{21} Mx, while the total horizontal flux is 4.0×10^{21} Mx, assuming the filling factor given by the inversion of the data. The distribution of the filling factors has a broad peak at $f = 0.15$ with FWHM range $0.05 < f < 0.35$ (see Fig. 9 panel d). The actual filling factor may be larger than these values because of the effects of stray light (Orozco Suárez et al. 2007a). An upper bound can be computed by assuming the extreme case of $f = 1$. This yields a total vertical magnetic flux

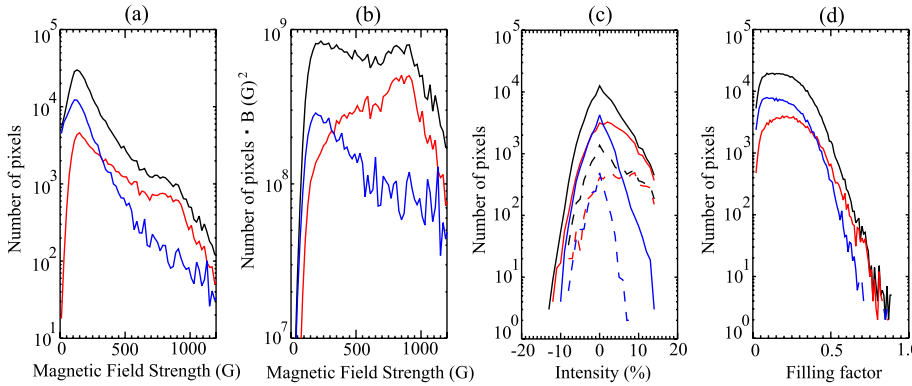


Fig. 9 Histograms of pixels at latitudes greater than 75°. *Red lines* indicate vertical field, *blue* horizontal field, and *black* both. Panel **a**: number of pixels as a function of the magnetic field strength (probability distribution function). Panel **b**: number of pixels multiplied with B^2 as a function of the magnetic field strength. The bin size of magnetic field strength in panels **a** and **b** is 20 G. Panel **c**: histogram of continuum intensity with magnetic field strength $>$ 300 G (*solid line*) and $>$ 800 G (*dashed line*). Since continuum intensity rapidly decreases toward the limb, the horizontal axis is the normalized excess continuum level with respect to the continuum level averaged over a 6.4″ box. Panel **d**: filling factor. From Tsuneta et al. (2008a)

of 9.9×10^{21} Mx, and a total horizontal magnetic flux of 2.0×10^{22} Mx. The difference is a factor of about 0.2, which roughly corresponds to the average filling factor.

The total vertical magnetic flux for the whole area with latitude above 70° is estimated to be between 5.6×10^{21} and 2.5×10^{22} Mx, assuming that the unobserved polar region has the same magnetic flux as the observed region. Since the surface area with latitude above 70° is 1.8×10^{21} cm^2, the average flux is estimated to be between 3.1 and 13.9 G. The total magnetic energy is proportional to $B^2 f S = B \Phi$. Thus, the surface poloidal magnetic energy is approximately two orders of magnitude larger than the case for the uniform magnetic field, if we take $B \sim 1$ kG, corresponding to the peak of the energy PDF in Fig. 9. Though these are the most accurate flux estimation so far made for the polar regions, these number should be regarded as minimum values due to the threshold in the selection of pixels for accurate inversion.

From the Hinode observations, the total flux of vertical magnetic field at the polar region is estimated to be at least 5.6×10^{21} Mx and at most 2.5×10^{22} Mx at the solar minimum. Various measurements indicate that the total magnetic flux of a single active region is about 10^{22} Mx (Longcope et al. 2007; Jeong and Chae 2007; Magara 2008). Thus, the measured total polar flux barely corresponds to that of single active region. The total toroidal flux would increase with time during the winding-up process by differential rotation, and the concept of the Ω-mechanism would be viable with these observational constraints.

5 Bright Points and Magnetic Elements

The magnetic field is found to be highly inhomogeneous in the lower solar atmosphere. While field is likely ubiquitously present in the photosphere (cf. Sects. 3 and 6), it is concentrated at the edges of convective cells in small-scale regions of high field strength. The convective flows expunge the field from cell interiors and concentrate the field in the intergranular downdrafts and at the borders of supergranular cells. Field is concentrated in small

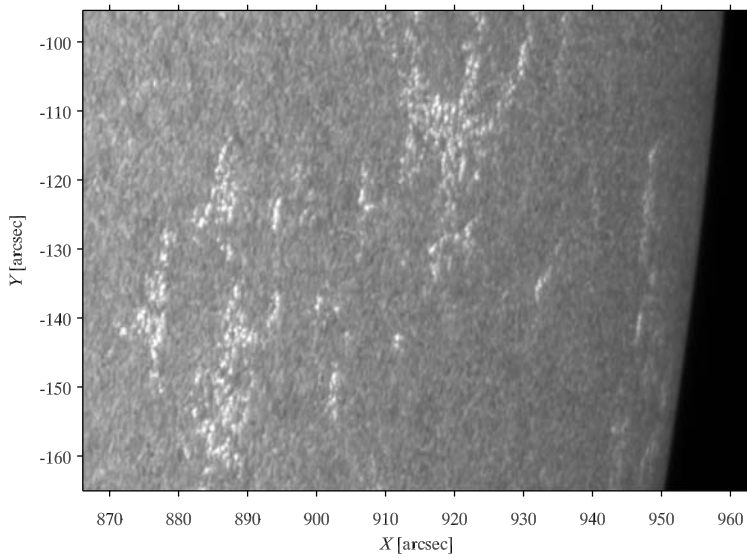


Fig. 10 A sample region in plage taken in the Fraunhofer G band near the west limb. Faculae appear as small, bright features on the disk-center-side of granules. This image was taken with Hinode/SOT on November 24, 2006, at 07:04:21 UT

"magnetic elements" that reach field strengths well beyond the equipartition field strength of about 500 G. It should be noted that these elements are not discrete structures as their name suggests. Rather, they are concentrations of strong field, with intricate structure that is expected to extend beyond what is already visible in observations at the highest spatial resolution. In addition, they frequently split into multiple apparently disjoint concentrations, or merge with other concentrations during their lifetime.

The plasma β in the photosphere outside kilogauss-strength magnetic elements, i.e., the ratio of the gas pressure to the magnetic pressure, is much larger than one. In addition, because photospheric plasma has a high conductivity, the field is "frozen in" the matter. As a result, the dynamics and evolution of magnetic fine structure in the photosphere are largely dominated by gas motions such as convection and large-scale flows associated with supergranulation. Concentrations of strong magnetic field provides an excellent conduit for conveying kinetic energy from the turbulent photosphere to higher layers of the solar atmosphere. If we are to understand the heating of the chromosphere and corona, as well as energetic events such as flares, it is important that we study the foot points of the field in the outer atmosphere.

5.1 Observations of Small-Scale Field Concentrations

There is a rich history of observations of concentrations of field in the solar photosphere. Figure 10 shows the most conspicuous small-scale magnetic features: faculae. They show up as small bright features at the limb, usually in plages or decaying active regions. For as long as the Sun has been observed through telescopes, the existence of faculae has been known. The counterparts of faculae closer to disk center are not as obvious in white light, but they do stand out in chromospheric diagnostics such as Ca II H.

Small, concentrated magnetic elements in the network were observed as "gaps" in photospheric lines around 525 nm by Sheeley (1967), and as "magnetic knots" in spec-

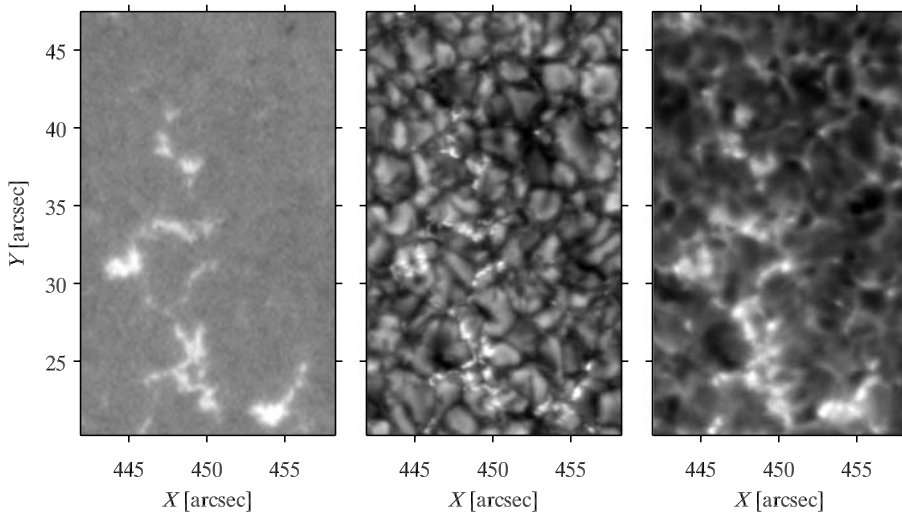


Fig. 11 Sample network region. *Left*: Fe I 630.2 nm line-of-sight magnetogram, scaled between −1 (*black*) and 1 kMx/cm^2 (*white*). *Middle*: G-band intensity. *Right*: Ca II H intensity. *Bright points* in the G band and in the Ca II H line images correlate well with positions of concentrated field. The network consists of strings of several adjacent bright points, located in the intergranular lanes. *Bright points* in the Ca II H image appear more extended and fuzzy than in the G-band image due to expansion of the flux tube with height. These images were taken with Hinode/SOT on March 30, 2007, around 00:24:30 UT. The *coordinates* indicate distance from sun center, so that $\mu \approx 0.88$

tra of plage by Beckers and Schröter (1968). It was clear from their observations that these structures were abundant in the vicinity of sunspots, but much more rare in quiet sun. In wide-band Hα images, bright features were observed to be arranged in "filigree", a long-lived, large-scale photospheric network (Dunn and Zirker 1973). Observations by Mehltretter (1974) had adequate resolution to resolve the photospheric network into strings of small "bright points" located in intergranular lanes, that change in shape and size on timescales comparable to the lifetime of granules. The photospheric network had been associated with kilogauss field in the magnetic network earlier (Stenflo 1973), suggesting that many of the observed structures were related (Muller 1977). Direct evidence that gaps, magnetic knots, faculae, filigree, and bright points were all manifestations of the same phenomenon was eventually provided by high-resolution observations, simultaneous in multiple wavelengths, of both disk and limb targets (Wilson 1981). The bright points form a dense pattern in plage and active network, while outside of active regions, they clump in patches that partially outline supergranular cells. The term "network bright point" was introduced to replace "facular points" and other terms in an effort to differentiate between bright points in regions of active and quiet sun (Stenflo and Harvey 1985; Muller 1985).

The importance of "proxy-magnetometry", the technique of determining the locations of magnetic field through a change in intensity, was recognized early on, and extensive studies of bright points in the photospheric network were quickly undertaken. Imaging in wide-band Ca II H and K or Hα were the diagnostics of choice, until Muller and Roudier (1984) switched to using the Fraunhofer G band around 430.8 nm in order to reduce the effects of chromatism in their telescope. Imaging in the G band has since caught on and is now widely used as one of the principal diagnostics for proxy-magnetometry. Figure 11 displays

an example of proxy-magnetometry using imaging in the G band and in the Ca II H line, together with a photospheric line-of-sight magnetogram. The images show a small patch of network that consists of many bright points.

Imaging at consistently high resolution at high cadence over a reasonable duration is required to study the dynamics and evolution of magnetic elements, but it not easy to achieve. Only during times of excellent seeing can an observer expect to make out these structures. Thankfully, advances in digital imaging technology, the advent of adaptive optics, and the development of sophisticated algorithms for correction of the effects of atmospheric seeing in post-processing now allow telescopes to produce diffraction-limited data with some regularity. In particular, many observers have successfully used observations with the former Swedish Vacuum Solar Telescope and the new 1-m Swedish Solar Telescope to study bright points, including size, shape, and appearance (Berger et al. 1995, 2004), dynamics (Berger and Title 1996; van Ballegooijen et al. 1998; Rouppe van der Voort et al. 2005), dispersal (Berger et al. 1998a), and contrast (Berger et al. 2007). In addition, the space-borne observatory Hinode does not suffer from seeing and has produced a vast amount of data that is highly suitable for studies of network and internetwork field.

The relation of bright points to the underlying magnetic field has also received a fair share of attention. Motivated by the relative ease with which data could be collected, observers often choose to study magnetic elements using proxy-magnetometry. Comparison of diagnostics such as imaging in the G band with magnetograms has shown clearly that strong, kilogauss field is required to form a bright point, but it is not a sufficient condition (Berger and Title 2001; Ishikawa et al. 2007). Many small concentrations of field that reach kilogauss strength do not have associated bright points, since the formation of the bright point depends strongly on the inclination of the field. The contrast of a bright point decreases as the field is angled further away from the line of sight (Beck et al. 2007). The hope is, of course, that the observed bright points are a random sample of the magnetic elements, because field orientation is independent of the line of sight. The results derived from these bright points are then expected to be valid for all magnetic elements, not just those that happen to have associated bright points, provided a statistically large number of bright points is sampled. However, there are several reasons why one should be careful with these assumptions. While proxy-magnetometry is comparatively simple, it is unable to continuously follow field concentrations if they are detected (De Wijn et al. 2005), and results based on these techniques are thus not just biased toward those concentrations that produce bright points, but also to the properties of those concentrations at the time that they are correctly angled to produce bright points. Proxy-magnetometry misses a significant portion of flux that never becomes sufficiently concentrated to produces bright points, and is insensitive to field polarity. It is important that results found through proxy-magnetometry be validated against measurements using a direct diagnostic of magnetic field.

Figure 12 shows two illustrative examples of isolated magnetic elements in line-of-sight magnetograms and proxy-magnetometry diagnostics. The elements appear similar in the magnetogram, yet only one has clear associated bright points in both G-band and Ca II H filtergrams. The other does not produce bright points, perhaps because the field is sufficiently angled away from the line of sight, or perhaps because the granulation around it is broken up and there is not enough proximity of hot granular walls to make a bright point. These magnetic elements can exhibit different dynamics, as the one in the bottom row may not be buffeted by granulation as much as the one in the top row. A study using a proxy-magnetometry diagnostic may therefore give different results than one based on true magnetometry.

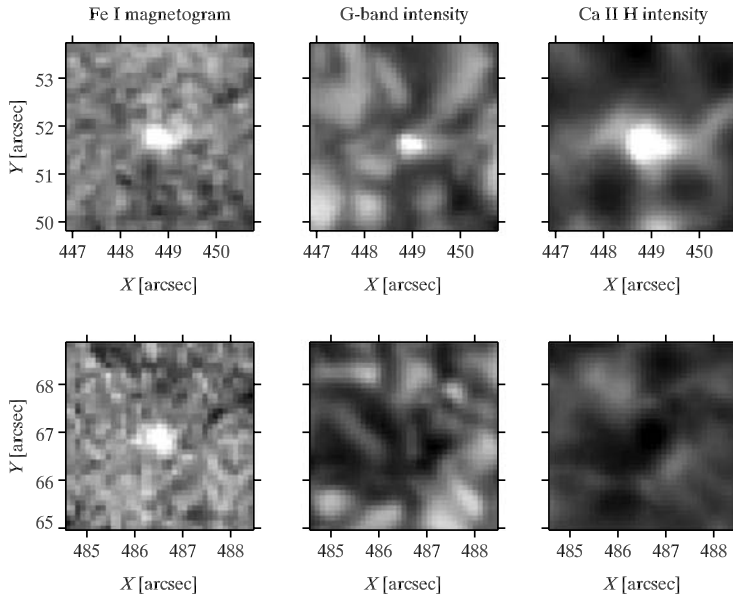


Fig. 12 Sample magnetic elements with (*top row*) and without associated bright points (*bottom panel*). These magnetic elements appear similar in the line-of-sight magnetogram, yet are very different in G-band and Ca II H intensity. From the same sequence as Fig. 11

5.2 Field Concentrations in Internetwork Areas

Kilogauss fields that produce bright points can also be found in internetwork areas. Though their existence was already noted in early studies of bright points (Muller 1983), these fields have been largely ignored historically, likely because bright points are more isolated in the internetwork and thus much harder to identify. There are fewer of them, and those that do exist are also more dynamic and have shorter lifetimes (Nisenson et al. 2003; De Wijn et al. 2005). These factors make detection and analysis difficult. Recently, however, vigorous investigation of magnetic field in internetwork areas has been undertaken. Many of the observations used in these studies are now available thanks to the development of new instruments, adaptive optics, and post-processing techniques. Analysis of the distribution of field strength and filling factor in the photosphere has attracted particular attention (Domínguez Cerdeña et al. 2003a, 2006b; Lites and Socas-Navarro 2004; Trujillo Bueno et al. 2004). While there is some disagreement between results, weak field appears to be ubiquitously present in the internetwork, structured on small spatial scales (see Sect. 2.2). In addition, horizontal field appears to pervade the photosphere (see Sects. 2.3 and 3). New observations at unprecedented resolution with the space-borne observatory Hinode have shown that it is transient in nature, structured on small scales, and occurs preferentially over the edges of granules rather than in the intergranular lanes as is the case for more vertical field (Lites et al. 2008). These observations also allow us to study emergence of field on small spatial and temporal scales (Centeno et al. 2007; Ishikawa et al. 2008). Field is brought up inside a convective cell, then quickly expelled from the interior and swept into the lanes, where it may merge with pre-existing field. The entire process takes only a few minutes.

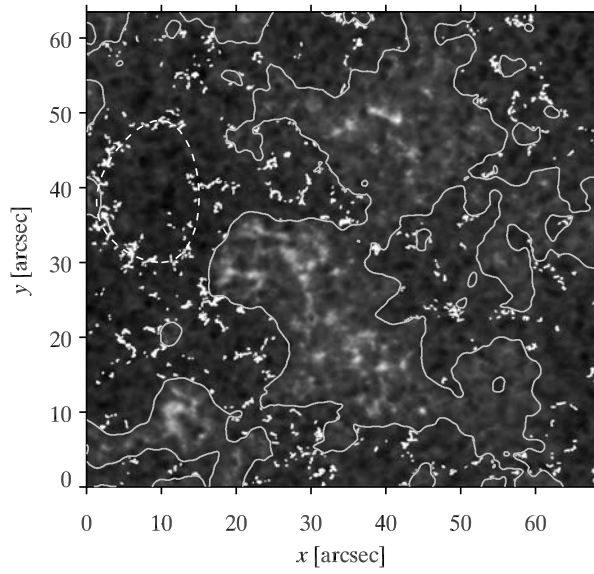


Fig. 13 Ca II H intensity averaged in time over a 1-hour sequence. The network is outlined by *solid white contours*. Locations of internetwork bright points are overlaid in white. The bright points appear to group in patches that outline edges of cell-like structures, such as around $(x, y) = (10'', 40'')$ (indicated by a *dashed line*). From De Wijn et al. (2005)

Several studies have focused on strong field that has associated bright points. The upshot is that internetwork field may become sufficiently concentrated to produce bright points, similar to network bright points, but more dynamic and with shorter lifetimes (Sánchez Almeida et al. 2004). The associated field has a longer lifetime than the bright point. The field exists before the bright point is formed and remains after it disappears, and may produce a bright point again at some later time (De Wijn et al. 2005). The lifetime of a bright point does not have a bearing on the lifetime of the underlying flux. Rather, it is a measure of the dynamics of the associated flux, i.e., how long the flux remains sufficiently aligned with the line of sight to produce a bright point, and also of the performance of the detection algorithm.

Strong field concentrations in internetwork areas also appear to outline cells on mesogranular scales. Figure 13 shows the locations of internetwork bright points detected in a 1-hour time sequence of Ca II H images. The interiors of these cells are largely devoid of field. Similar patterns have been found in active network (Berger et al. 1998b). One would expect such a pattern to be set by granular motions. Perhaps magnetic elements form these patterns as a result of flows associated with "trees of fragmenting granules" (Roudier and Muller 2004, previously called "active granules" by Müller et al. 2001) which were previously linked to mesogranules (Roudier et al. 2003). Flux is expunged by the sideways expansion of granular cells, and is collected in the downflows in intergranular lanes. In a "tree of fragmenting granules", these flows would be expected to drive flux not only to the edges of individual granules, but also to the edges of the tree, resulting in a mesogranular pattern in the positions of magnetic field in internetwork areas.

5.3 Magnetic Element Dynamics

Magnetic elements appear to obey largely Gaussian distribution of horizontal velocity, as would be expected from random buffeting by granulation. Their rms velocity is about 0.5 km/s in network and about 1.5 km/s in internetwork. Granular motions are suppressed in the network, resulting in less dynamic behavior. Magnetic elements in internetwork areas

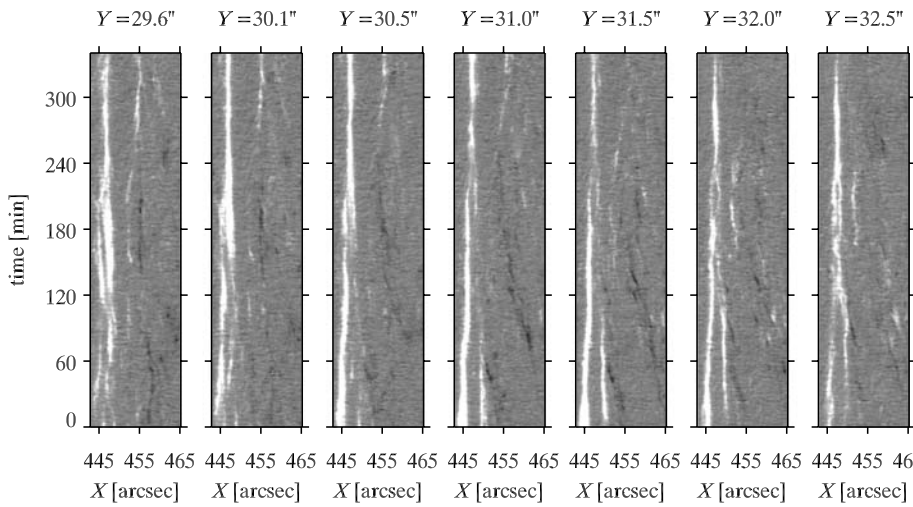


Fig. 14 Slices through a sequence of Fe I 630.2 nm line-of-sight magnetograms, scaled between −400 (*black*) and 400 Mx/cm^2 (*white*). The field of view includes some network at the left. Magnetic elements exhibit dynamic behavior, migrating distances of several arcseconds over a few hours, while experiencing many interactions with other elements during that time. They frequently seem to appear without a clearly associated opposite polarity. Examination of adjacent slices indicates that such elements are not elements that emerged as a bipole previously and are now migrating into the current 2D slice. This indicates that the flux emerged at some earlier time, and was only detected when it became concentrated enough, or that the associated flux of opposite polarity is spread out below the detection limit of the instrument. From the same sequence as Fig. 11

sometimes migrate large distances over periods of a few hours (cf. Fig. 14), while having frequent interactions with other long-lived elements and transient concentrations of field. Magnetic elements do not typically have an identity over periods longer than a few minutes because of these interactions.

Motions of bright points and magnetic elements in internetwork areas show positive autocorrelation up to at least delay times of 10 minutes, indicating that the elements retain some memory of their motions over at least that much time. Magnetic elements in internetwork areas have motions preferentially in the direction of the nearest network concentration (De Wijn et al. 2008). The likely culprit is thus supergranular flow.

5.4 Formation of Bright Points

Modeling of strong magnetic concentrations began with analytic studies of magnetostatic "flux tubes" (Spruit 1976, 1977). As computers became more powerful, numeric MHD models were created, increasing in complexity and realism over the years. Early models were used to calculate properties of "flux sheets" in two dimensions (e.g., Knölker and Schüssler 1988). Modern three-dimensional numerical models of magneto-convection now simulate mesoscale-areas (e.g., Stein and Nordlund 2006), and are successful in reproducing magnetic elements and bright points in the solar photosphere (Schüssler et al. 2003; Steiner 2005). However, two-dimensional models remain popular (Steiner et al. 1998), because adding the third dimension is computationally expensive.

These models have shown that while the cause of brightness enhancement of magnetic elements in the photosphere differs subtly between various proxy-magnetometry diagnostics, it is in all cases rooted in the partial evacuation of the flux tube as a result of magnetic

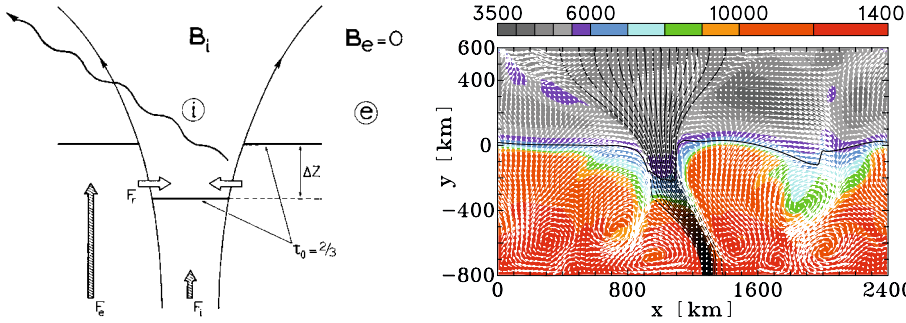


Fig. 15 Examples of modeling of magnetic elements. *Left*: analytic magnetostatic flux tube model (from Schrijver and Zwaan 2000). *Right*: sophisticated 2D numerical MHD model of a flux sheet (from http://www.kis.uni-freiburg.de/ steiner/)

pressure. The reduced density places optical depth unity inside the flux tube at a geometrically deeper layer compared to outside, thus allowing radiation to escape from deeper, hotter layers (cf. the left panel of Fig. 15). The internals of the flux tube are cooler at equal geometric height due to radiation losses, but are typically hotter at equal optical depth. This process only produces enhanced brightness in small-scale structures. Larger concentrations of strong field that inhibit convection, e.g., pores and sunspots, become dark because there is insufficient radial influx of radiation to make up for the increased losses as a result of reduced opacity. Well-known diagnostics for proxy-magnetometry are used because they show more brightness enhancement than the continuum. As an example, molecular lines such as those in the Fraunhofer G band and the CN band are weakened in small concentrations of field because of the dissociation of molecules at lower densities and higher temperatures (Kiselman et al. 2001; Steiner et al. 2001; Sánchez Almeida et al. 2001; Shelyag et al. 2004; Uitenbroek and Tritschler 2006). Opacity inside the flux tube is thus additionally reduced in these lines, and a higher emergent intensity integrated over the passband results.

One recent highlight was the confirmation of the suggestion made by Spruit and Zwaan (1981) that facular brightness enhancement is the result of radiation escape from hot granular walls. Keller et al. (2004) and Carlsson et al. (2004) used sophisticated 3D MHD simulations to model flux tubes, then "observed" them as if close to the limb using intricate codes to calculate radiative transfer. Partial evacuation of the flux tube allows the observer to look deeper into the hot granular wall than would be possible if the flux tube were absent. The dark lane often observed on the disk-ward side of faculae is formed in the cool layers above the granules and inside the flux tube.

5.5 Formation of Magnetic Elements

The prevailing theory on the formation of magnetic elements incorporates a convective instability known as "convective collapse" (Parker 1978). Field is brought up in granules and swept into the intergranular downdrafts. Flux can accumulate until the magnetic energy density is roughly equal to the kinetic energy density of granular flows. This yields an equipartition field strength of about 500 G, insufficient to produce a bright point. The gas in these regions cools because convective energy transport is suppressed by the field. The cool, dense gas enhances the intergranular downflow, so that the region is effectively evacuated by gravity and consequently compressed until the internal magnetic pressure is sufficiently

increased so that the region is again in horizontal pressure balance with the outside. Theoretical calculations indicate that magnetic elements with field strengths of 1–2 kG result from this process (Spruit 1979; Grossmann-Doerth et al. 1998).

The formation of kilogauss field concentrations from weak turbulent flux can be studied from models. Typically, a hydrodynamic model is run until it reaches a more-or-less relaxed state. A constant vertical field is then added, and the simulation is allowed to evolve further. While this obviously does not resemble what happens on the Sun, the process of convective collapse does still occur and indeed has been observed in simulations of magneotconvection (Vögler et al. 2005).

It is harder to observe convective collapse on the Sun. Accurate (spectro)polarimetric observations with high resolution and reasonable cadence are required over some period of time. Such observations are sensitive to seeing conditions, due to, e.g., long exposure times required for polarimetry. The seeing-free Hinode observatory is an obvious candidate to provide suitable observations. Indeed, formation of a kilogauss field concentration by convective collapse was recently observed using the Hinode spectropolarimeter and found to be in qualitative agreement with results from numerical simulations (Nagata et al. 2008). These observations strongly support the model of convective collapse for the formation of kilogauss field concentrations.

6 Unresolved Magnetic Fields

6.1 Range of the Unresolved Scales

Magnetoconvection in the rotating Sun is the engine of the solar dynamo. It is therefore central to our understanding of the origin of solar and stellar activity. A main problem in modeling the solar dynamo is that magnetoconvection has such a tremendous dynamic range, about 8 orders of magnitude or more, as we will see below, while numerical simulations can handle only about 3 orders of magnitude. While numerical simulations provide valuable insights, the theory needs to be guided by observations.

The magnetic-field observations refer to the surface layers (photosphere), while the properties of magnetoconvection vary with depth in the convection zone. Still, the photosphere can serve as our magnetoconvective laboratory, where we can explore the underlying physics. However, a major part of the magnetoconvective spectrum extends over scales that are too small to be resolved even with next-generation telescopes in any foreseeable future.

One may therefore question to what extent knowledge about the behavior of these small unresolved scales is really needed for understanding solar and stellar dynamos and magnetic activity. The solar dynamo gives the impression of being governed by large-scale properties like Hale's polarity law, Joy's law, the emergence and dispersion of active-region magnetic flux, shearing by differential rotation, and meridional circulation, all of which take place in the spatially resolved domain. In Sect. 6.2 we will address the connection between the scales and the role of the smallest diffusion scales for the operation of the solar dynamo.

The upper end of the scale spectrum is naturally bounded by the size of the Sun ($\approx 10^6$ km). The lower end of the magnetic spectrum is reached when the turbulent motions are unable to tangle the field lines to produce magnetic structuring. This decoupling between the plasma motions and the magnetic field happens when the frozen-in condition ceases to be valid, i.e., when the time scale of magnetic diffusion (field-line slippage through the plasma) becomes shorter than the time scale of convective transport.

The ratio between these two time scales is represented by the magnetic Reynolds number

$$R_m = \mu_0 \sigma \ell_c \, v_c \tag{1}$$

in SI units. Here σ is the electrical conductivity, ℓ_c the characteristic length scale, and v_c the characteristic velocities. $\mu_0 = 4\pi \times 10^{-7}$. When $R_m \gg 1$ the field lines are effectively frozen in and carried around by the convective motions. When $R_m \ll 1$ the field is decoupled from the turbulent motions and diffuses through the plasma.

The magnetic structuring by magnetoconvection therefore ends at scales ℓ_{diff} where $R_m \approx 1$. To calculate these scales we need to know how the characteristic turbulent velocity v_c scales with ℓ_c. Such a scaling law is given in the Kolmogorov theory of turbulence. In the relevant inertial range it is

$$v_c = k \, \ell_c^{1/3}, \tag{2}$$

where k is a constant. An estimate of $k \approx 25$ can be obtained from the observed properties of solar granulation (Åke Nordlund, private communication).

Combining these two equations and setting $R_m = 1$, we obtain the diffusion scale

$$\ell_{\text{diff}} = 1/(\mu_0 \, \sigma \, k)^{3/4}. \tag{3}$$

To evaluate this we need the expression for the Spitzer conductivity in SI units,

$$\sigma = 10^{-3} \, T^{3/2}. \tag{4}$$

This gives us

$$\ell_{\text{diff}} = 5 \times 10^5 / T^{9/8}. \tag{5}$$

For $T = 10^4$ K (a rounded value that is representative of the lowest part of the photosphere or upper boundary of the convection zone), $\ell_{\text{diff}} \approx 15$ m.

If we limit ourselves to order-of-magnitude estimates, we may say that magnetic structuring in the surface layers ends at scales of about 10 m. As the upper bound of the magnetic scale spectrum is about 10^6 km, it follows that the magnetoconvective scale spectrum spans about 8 orders of magnitude in the Sun's observable surface layers. As the diffusion limit decreases with increasing temperature, it follows that the dynamic range of magnetoconvection increases, possibly by nearly two orders of magnitude more (down to diffusion scales of cm) as we go down in the convection zone and the temperature increases towards a million degrees.

It might be objected that the rather simplistic Kolmogorov scaling law is not very applicable in the highly stratified surface layers. However, the photospheric scale height of typically 150 km is about 4 orders of magnitude larger than the diffusion scales that we have derived. Most of these scales are so small that they do not "feel" the stratification and therefore may behave in a way that is similar to isotropic Kolmogorov turbulence.

6.2 Role of the Smallest Scales for the Global Dynamo

The magnetic fields that we see at the surface of the Sun have been produced by dynamo processes in the solar interior. Lifted by buoyancy forces, the dynamo-produced fields emerge as bipolar regions into the visible photospheric layers, but with an emergence rate that is a steep function of scale size. The large-scale bipolar regions, which represent active regions (AR) with sunspots, bring up about 10^{20} Mx per day (solar-cycle average), enough

to account for the observed accumulation of flux and the large-scale background magnetic field over the course of the 11-year activity cycle. Going down in scale to the so-called ephemeral active regions (ER), the flux emergence rate goes up by two orders of magnitude, to 10^{22} Mx per day. Going down to the still smaller internetwork fields (IN), the emergence rate increases to 10^{24} Mx per day, another two orders of magnitude (Zirin 1987). While the characteristic scales of AR : ER : IN are in proportion 25 : 5 : 1 ($75'' : 15'' : 3''$), the emergence rates are in proportion 1 : 100 : 10 000.

With these high emergence rates the time scale for the turn-over or replenishment of the magnetic field pattern is not the solar cycle time scale but something much shorter. The first realization of a short turn-over time scale came two decades ago from a study of the differential rotation properties of the magnetic pattern (Stenflo 1989). When determining the proper motion of magnetic elements through cross-correlation techniques (Snodgrass 1983), a steep differential rotation law is found, which closely agrees with the law derived from Doppler measurements. When instead we form time series of the magnetic field sampled at the central meridian and perform an autocorrelation analysis to determine the period it takes for the pattern to recur after one solar rotation (or any integer number of rotation periods), then a rotation law is found that is almost rigid (Stenflo 1989). This dramatic difference between the cross-correlation and autocorrelation analyses can be naturally explained if the pattern replenishment time is much shorter than a rotation period, so that the "recurring" pattern is not actually recurring but is a new pattern that has emerged during the course of the rotation period.

The nearly rigid differential rotation law then does not represent the surface (in contrast to the steep differential rotation law), but reflects the differential rotation properties of the source region in the deep convection zone, from which the new surface fields emanate.

This behavior cannot be easily explained in terms of flux-redistribution models without high-latitude sources of new magnetic flux, like the model of Sheeley et al. (1987), which are based on meridional circulation and a smooth surface diffusion process. In such models a quasi-rigid differential rotation law for the phase velocity of the magnetic pattern results, regardless of the lag used in the correlation analysis. The observed lag-dependence of the pattern phase velocity with a steep differential rotation law for small lags would not occur without the continual supply of new magnetic flux from the Sun's interior at high latitudes. To avoid this contradiction between the flux-redistribution models and the observations, Wang and Sheeley (1994) replaced the smooth diffusion in their model with a discrete random walk process on a supergranular lattice, as a means of producing discrete flux clumps at high latitudes from the old, smooth, redistributed flux. These clumps would then drift according to a steep differential rotation law. It is, however, questionable whether supergranular random walk can continually produce flux clumps of sizes larger than one arcmin (the spatial resolution used in the correlation analysis of Snodgrass (1983) that gave the steep differential rotation law), much larger than the size of supergranules. A more natural explanation is that the magnetic pattern is really being replenished from the Sun's interior on a time scale well below the solar rotation time scale.

Support for such a short pattern replenishment time has come from SOHO MDI magnetograms, revealing a "magnetic carpet" with a pattern turn-over time of 1.5–3 days (Schrijver et al. 1997; Title and Schrijver 1998).

The problem with the high emergence rates is that they have to be matched by the flux *removal* rates for a statistically stationary situation, otherwise the photosphere would quickly get choked with magnetic flux that is all the time injected from below. It is however difficult to identify the process by which flux is removed. This problem is generally avoided in dynamo models by letting opposite polarities mathematically cancel out when they are cospatial. However, such mathematical cancellation is non-physical, magnetic flux can only

be destroyed by a reconnection process involving concentrated electric currents and Joule heating, and this can only occur fast enough if it takes place on the diffusion length scales (of order 10 m in the photosphere). This implies an extreme and highly efficient shredding of the flux elements down to these scales, something that takes place almost entirely in the spatially unresolved domain and which is therefore not directly observed.

Flux removal may occur in basically three different ways: (1) In situ cancellation of opposite magnetic polarities (reconnection). (2) Flux retraction (reprocessing in the convection zone). (3) Flux expulsion (with a possible role of CMEs). Unfortunately, the relative contributions of these three processes are completely unknown. How impervious is the solar surface to the dynamo-produced magnetic flux? How "leaky" is the solar dynamo? These are fundamental questions that still have no answers. Similar questions may be asked about the magnetic helicity.

Although reconnection is only mentioned explicitly in connection with the in situ cancellation, both flux retraction and flux expulsion could not happen without cutting off the field lines through reconnection. Therefore, also for these processes, the basic physics takes place at the diffusion length scales, down to which the flux needs to be efficiently shredded. Without this shredding, the global dynamo would not be able to operate.

6.3 Scaling Behavior of the Magnetic Field Pattern

The resolved scales now cover a dynamic range of almost four orders of magnitude (from the global scales of 10^6 km down to the neighborhood of 100 km), approximately half of the range of the magnetoconvective scale spectrum. Already back in the 1960s, in the early days of solar magnetography, when the dynamic scale range covered by the observations was only about two orders of magnitude, it was clear that the Sun's magnetic field is very fragmented or intermittent, but the degree of intermittency or the nature of the structuring was not known. To get an insight into the hidden nature of the field it was necessary to develop indirect diagnostic techniques to overcome the resolution limit and derive intrinsic field properties that were not dependent on the quality of the telescopes used.

A similar situation is encountered in stellar physics, where we derive the physical properties of the stellar atmospheres although the stars remain unresolved point objects. While crucial information on key field parameters like magnetic field strengths and filling factors can be obtained this way, we have no information on the unresolved field *morphology*, and the results depend on the interpretative models used. An exception is Zeeman-Doppler imaging of rapid rotators. By necessity these models have to be idealized to limit the number of free parameters, and they need to be tailored to the type of diagnostics that we use. Thus the Zeeman and Hanle effects are sensitive to very different parameter domains of the field, as we will see in Sect. 6.4.

The situation has improved dramatically during the last decades. Advances in spatial resolution have significantly extended the dynamic range of the resolved scales, allowing us to get a glimpse of how the magnetic pattern scales as we zoom in on ever smaller scales. Numerical simulations have given us insights into the nature and scaling behavior of magnetoconvection when we go beyond the resolution limit into the unresolved domain. This allows us to get a better understanding of the nature of the field pattern and gives us guidance in the choice of the most realistic interpretative models to use to diagnose the spatially unresolved domain.

Until a few years ago the "standard model" of photospheric magnetic fields was that the basic building blocks are strong-field (mostly kilogauss) highly intermittent flux tubes occupying a small fraction of the photospheric volume, and that the space between these

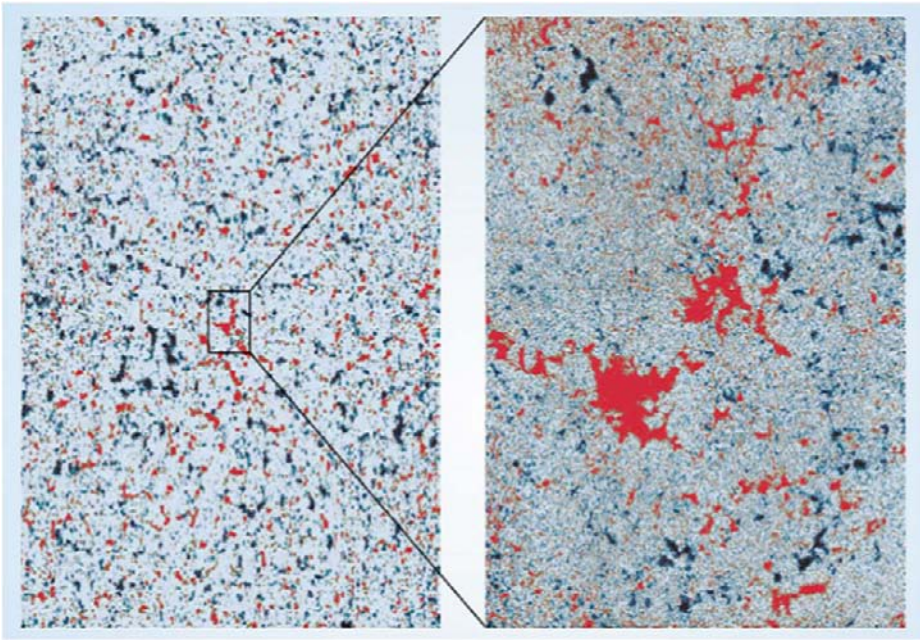


Fig. 16 Illustration of the fractal-like nature of the magnetic-field pattern on the quiet sun. The *left map*, extracted from the central part of a Kitt Peak magnetogram of 9 February 1996, covers 15% of the solar disk, while the *right map*, obtained on the same day at disk center with the Swedish La Palma telescope (courtesy Göran Scharmer) covers an area that is 100 times smaller (Stenflo and Holzreuter 2002; Stenflo 2004)

flux tubes is filled with much weaker and highly tangled (or "turbulent") fields. We now realize that this "two-component picture" is mainly a product of the idealizations used when interpreting Zeeman and Hanle signatures of the spatially unresolved domain. Instead the field appears to behave like a fractal.

Figure 16 illustrates this fractal appearance of quiet-sun magnetic fields. If a magnetogram is presented without tick marks that indicate the spatial scale, it is very hard to guess what the scale is. The pattern seems to have a high degree of scale invariance, it looks statistically the same as we zoom in on ever smaller scales. Further we have a coexistence of strong and weak fields over a large dynamic field strength range. The probability distribution function for the field strengths appears to be nearly scale invariant and can be described in terms of a Voigt function with a narrow Gaussian core and "damping wings" that extend out to the kilogauss values (Stenflo and Holzreuter 2002; Stenflo and Holzreuter 2003, but see also Domínguez Cerdeña et al. 2006b). Such scale invariant properties are typical of a fractal. A fractal dimension of 1.4 has been found for both the observed magnetic field pattern and the pattern that results from numerical simulations of magnetoconvection at scales that are smaller than the resolved ones (Janßen et al. 2003).

6.4 Field Diagnostics Beyond the Spatial Resolution Limit

Like in any other area of astrophysics where we are dealing with spatially unresolved objects, we have to extract information about the physical conditions that is encoded as various types of signatures in the spectrum. To enable this extraction we make use of models that

must have a smaller number of free parameters than the number of independent observables that we can use to constrain the model. A fundamental issue is the uniqueness and numerical stability of any such inversions.

6.4.1 Zeeman Diagnostics and the Line-Ratio Technique

Back in the 1960s it became clear that the measured field strengths on the quiet sun increased with the spatial resolution of the instrument, which led to the question what the strength would be if we had infinite resolution (Stenflo 1966). To answer this question the line-ratio technique was devised, which led to the conclusion that more than 90% of the net magnetic flux in the photosphere, as seen with modest spatial resolution (larger than a few arcseconds), comes from highly bundled fields with a strength of 1–2 kG and a small volume filling factor (typically 1%) (Howard and Stenflo 1972; Frazier and Stenflo 1972; Stenflo 1973). Due to the tiny filling factor the average net field strength is only of order 10 G or less, although most of the field lines come from kilogauss flux patches in the photosphere.

This result led to the concept of discrete magnetic flux tubes as the theoretical counterpart of the unresolved kilogauss flux fragments. The mechanism of convective collapse (Parker 1978; Spruit 1979; Spruit and Zweibel 1979) gained wide acceptance as the process leading to the spontaneous formation of kilogauss flux tubes. Empirical flux tube models at increasing levels of sophistication were built (Solanki 1993). Observational support for the convective collapse mechanism could be found (Solanki et al. 1996), while also showing the existence of a family of weaker flux tubes that had been theoretically predicted (Venkatakrishnan 1986).

The classical line-ratio technique that allows a robust determination of the intrinsic field strength is based on the simultaneous observation of the circular polarization in the Fe I 524.7 and 525.0 nm line pair (Stenflo 1973). The observed circular polarization due to the longitudinal Zeeman effect, illustrated in the FTS spectrum of Fig. 17, depends on many combined factors, like the line depth and detailed line shape (which in turn depend on the temperature-density stratification of the atmosphere and the details of line formation), the Landé factor, and the line-of-sight component of the magnetic field. In traditional magnetography one calibrates away the line-profile factors by recording the magnetograph response to artificial line shifts of the spatially averaged spectral line. This calibration procedure then gives us the average line-of-sight field strengths (averaged over the spatial resolution element of the instrument), under the assumption that the relation between circular polarization and field strength is a linear one (being the relation that is valid in the weak-field limit), and assuming that the line depth and line shape are the same in the magnetic elements as for the spatially averaged sun. However, both these assumptions are generally wrong. As the field strength increases, the relation between circular polarization and field strength becomes increasingly non-linear. Further, the temperature-density stratification of the atmosphere (and thus the line profile) is significantly different in the unresolved magnetic elements than outside them. The line-ratio technique allows us to isolate the magnetic-field effects from the line formation and temperature-density effects, to obtain a signature that can only occur if the field is intrinsically strong (meaning that the polarization dependence on field strength lies in the non-linear regime). From the measured degree of non-linearity (Zeeman saturation) the value of the field strength can be extracted.

The robustness of the method depends on the choice of line pair. No line pair has been found that better optimizes this robustness than the Fe I 524.7 and 525.0 nm one. These two lines both belong to multiplet no. 1 of iron, have almost identical excitation potential, oscillator strength, and line depth, and therefore respond in the same way to the temperature-density stratification of the atmosphere, with the same line formation properties. The only

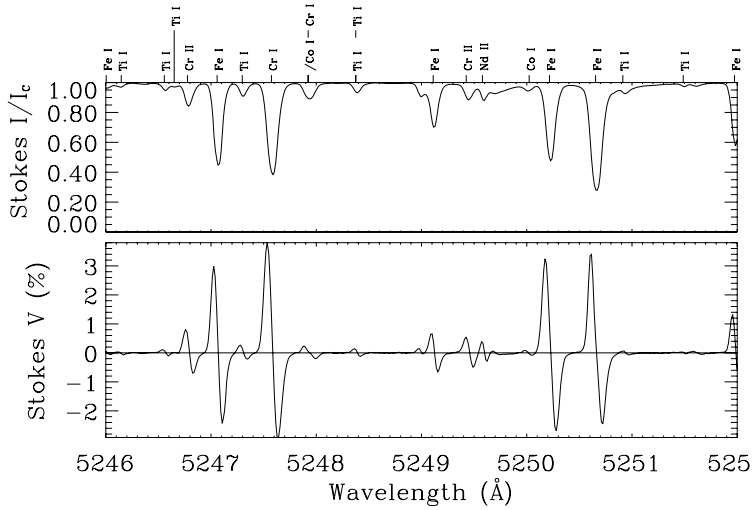


Fig. 17 Portion of a recording in a facula at disk center with the Fourier Transform Spectrometer at the McMath-Pierce facility (Kitt Peak) (Stenflo et al. 1984). The classical line-ratio technique is based on the comparison between the Stokes V amplitudes in the two Fe I 524.7 and 525.0 nm lines (Stenflo 1973). Zeeman saturation due to strong, unresolved fields leaves a characteristic signature in the profile ratio, which allows the intrinsic field strength of the unresolved fields to be determined

significant difference between them is their effective Landé factors: 2.0 for the 524.7 nm line, 3.0 for the 525.0 nm line. If all fields were intrinsically weak, the circular-polarization Stokes V profiles of the two lines would have the same shapes and only differ in terms of a global amplitude scaling factor in proportion to their Landé factors ($g_{524.7} : g_{525.0} = 2 : 3$). If we form the ratio $g_{524.7}\, V_{525.0}/(g_{525.0}\, V_{524.7})$, it would be unity if all fields were weak, regardless of the temperature-density stratification or line-formation properties of the solar atmosphere. It differs from unity only because of the *differential non-linearity*: the 525.0 nm line with its larger Landé factor deviates more from linearity than the 524.7 nm line.

This line-ratio technique was applied before the advent of Stokesmeters, using magnetograph exit slits in fixed positions of the line profiles (Stenflo 1973). This was sufficient for obtaining robust field-strength determinations. With fully resolved Stokes V line profiles with high S/N ratio (cf. Fig. 17) it became possible to test and verify the interpretation in great detail, since the Zeeman saturation does not only suppress the Stokes V amplitudes but also broadens the Stokes V profile in a way that gives the $g_{524.7}\, V_{525.0}/(g_{525.0}\, V_{524.7})$ ratio a very characteristic profile shape when plotted as a function of wavelength $\Delta\lambda$. Thus the self-consistency and validity of the interpretational model could be verified (for details, see Stenflo 1994).

This interpretational model contained two components: one magnetic component with field strength and filling factor as the free parameters, and one non-magnetic component. The measured line ratio does not depend on filling factor, only on field strength. The filling factor enters when explaining the V amplitudes of each line, since the amplitudes scale with both filling factor and field strength. Since the line ratio was found to be practically identical in quiet network regions with little magnetic flux and in strong faculae with much flux, the conclusion was that the magnetic building blocks (flux tubes) have rather unique properties (Frazier and Stenflo 1972), almost always with field strengths of 1–2 kG. Different regions on the Sun (outside sunspots) then differ not so much in field strength, but rather in the

number density or filling factor of the flux elements. This implies that the magnetograms, which show a continuous range of apparent field strengths, basically are maps of the filling factor, not of field strength.

This view of solar magnetism has been confirmed with other combinations of spectral diagnostics, in particular with infrared lines (e.g., Rüedi et al. 1992), which however have also revealed the existence of intrinsically weaker flux elements that are mixed in with the kilogauss ones. Due to the larger Zeeman splitting in the infrared it was possible to extend the 2-component approach to a 3-component one (with two magnetic components), which revealed the existence of intrinsically weaker fields. With advances in spatial resolution it became possible to actually resolve and see the flux tubes that had been predicted by the line-ratio method, as first done with speckle polarimetry (Keller 1992).

6.4.2 Hanle Diagnostics

While all these results were self-consistent, the Zeeman-effect observations left us with a picture where about 99% of the photospheric volume (outside the kilogauss flux elements) was field free, which is non-physical, since nothing in the highly electrically conducting and turbulent photospheric plasma can be field free. The introduction of a "non-magnetic" component is exclusively for mathematical convenience. The question is what the magnetic nature of this component is. Since its contribution to the Zeeman-effect polarization signals is very small, it must either mean that the field is indeed extremely weak, or that the field is highly tangled with mixed polarities within the spatial resolution element, such that one has nearly perfect cancellation of the opposite signs of the spatially unresolved Stokes V signals (in which case the field does not have to be weak). We now know through applications of the *Hanle effect* that the second case is much closer to the truth.

In contrast to the Zeeman-effect polarization, the Hanle effect is a coherency phenomenon that only occurs when coherent scattering contributes to the line formation. Such scattering can produce linear polarization also in the absence of magnetic fields. The term Hanle effect covers all the magnetic-field induced modifications of the scattering polarization. Since it has different sensitivity and symmetry properties than the Zeeman-effect polarization, it both responds to much weaker fields and does not suffer from the cancellation effects that make the Zeeman effect "blind" to a tangled field. This property was first exploited by Stenflo (1982) to derive a *lower* limit of 10 G for the strength of the tangled field in the 99% of the volume between the kilogauss flux tubes.

Examples of Hanle-effect signatures and how they differ from the Zeeman effect are shown in Fig. 18. The photospheric Sr I line in the left panels has been extensively used by various authors (Faurobert-Scholl 1993; Faurobert-Scholl et al. 1995; Stenflo et al. 1998; Trujillo Bueno et al. 2004) to improve the constraints on the properties of the turbulent field for which the Zeeman effect is blind. The most sophisticated constraints based on the use of probability distribution functions (PDF) have been derived by (Trujillo Bueno et al. 2004), indicating turbulent field strengths of order 100 G. Such volume-filling fields contain so much magnetic energy that they may play a major role in the energy balance of the solar atmosphere. The Hanle signatures of the strong Ca I 422.7 nm line (right panels in the figure) can be used to diagnose the horizontal magnetic fields in the solar chromosphere.

Assume that we have chosen our Stokes coordinate system such that the non-magnetic scattering polarization is along the Stokes Q direction. This direction is parallel to the nearest solar limb when observing on the solar disk in a zone near the limb (which we most often do for such observations, since the scattering polarization amplitude increases as we

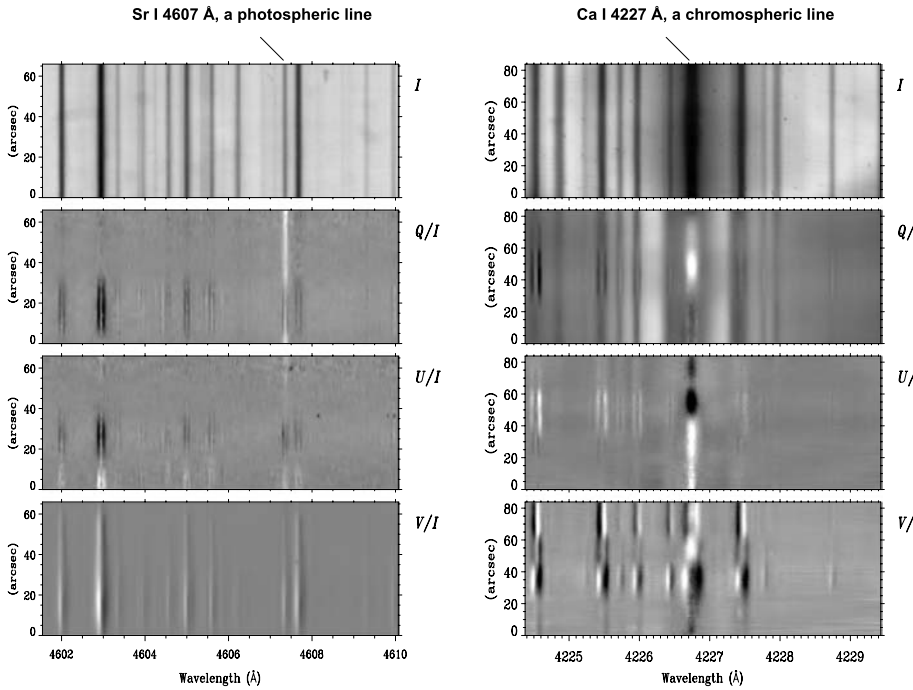


Fig. 18 Illustration of the different signatures of the Zeeman and Hanle effects in images of the Stokes vector (represented by the images of the intensity I and the three fractional polarizations Q/I, U/I, and V/I). The Hanle effect appears in the linear polarization (Stokes Q/I and U/I) in the line cores of certain lines, like Sr I 460.7 nm (*left panels*) and Ca I 422.7 nm (*right panels*), while the Zeeman effect exhibits its usual polarization signatures in the surrounding lines. The recordings were made with the Zurich Imaging Polarimeter (ZIMPOL, cf. Povel 1995; Gandorfer et al. 2004) at the McMath Pierce facility (Kitt Peak)

get closer to the limb). The main polarization signatures of the Hanle effect are depolarization (reduction of the Stokes Q amplitude) and rotation of the plane of polarization (appearance of signals in Stokes U). Since however the Hanle rotation angle can have both signs, a highly tangled field with equal contributions of plus and minus will lead to cancellations like for the Zeeman effect, so there will be no Stokes U signatures from such fields. In contrast, the depolarization effect in Stokes Q has only one "sign" (reduction of the polarization amplitude), regardless of the field polarity, and is therefore immune to the above-mentioned cancellation effects. This is the signature of the turbulent fields that we have to work with.

For Hanle diagnostics of the turbulent fields Hanle depolarization gives us one observable per spectral line. For single-line observations, the interpretative model therefore cannot contain more than one free parameter. Since the introduction of this diagnostic technique, the traditional model has been to assume a single-valued field with an isotropic angular distribution (Stenflo 1982). The free parameter is the single-valued field strength. However, other more realistic model choices are beginning to be used, which are guided by the insights gained from numerical simulations of magnetoconvection and from analysis of magnetic-field distribution functions that have been determined from observations in the spatially resolved domain. From the observed scaling behavior of the resolved fields and the behavior of the smaller-scale fields in numerical simulations one can make educated guesses for the analytical shapes of the field strength probability distribution functions (PDFs) that should be used to model the behavior in the spatially unresolved domain. With a single

Hanle observable (the depolarization for a single spectral line) we must then limit ourselves to characterize the model PDF with a single free parameter, for instance by keeping the relative shape invariant and using a stretching factor as the free parameter. Such an approach has been applied with success for data with the Sr I 460.7 nm line (Trujillo Bueno et al. 2004).

The Hanle effect however does not at all limit us to use such simplistic, one-parameter models. They are only used because the application of the Hanle effect to the diagnostics of spatially unresolved magnetoconvection is still in its infancy, and one needs to start with and fully understand the simplest approaches before proceeding to higher levels of sophistication. Like with the line-ratio technique in the case of the Zeeman effect, with the Hanle effect one can also use a multi-line approach with simultaneous Hanle observations in several spectral lines with different sensitivities to the Hanle effect. Such an application of the *differential Hanle effect* (Stenflo et al. 1998) increases the number of independent observables, which allows us to increase the number of free parameters of the interpretative models and thus enhance the degree of realism. While most lines differ not only in their Hanle sensitivities but also in their line formation properties, which adds considerable complication to the inversion problem, there exist certain pairings of molecular lines for which the line formation properties are identical, the only difference being the Hanle sensitivities (Berdyugina and Fluri 2004). This allows the magnetic-field effects to be isolated from the other non-magnetic effects, similar to what is done with the 525.0/524.7 Zeeman-effect line ratio. This has the great advantage of making the inversion much more robust and the derived field strengths less model dependent.

6.4.3 Unified Zeeman-Hanle Diagnostics with Distribution Functions

The Zeeman and Hanle effects are highly complementary. The longitudinal Zeeman-effect signals represent the net magnetic flux that often (but not always) has its main sources in the highly bundled strong fields, but they carry nearly zero information on the spatially unresolved volume-filling weaker, tangled fields between the intermittent stronger fields. Let us here recall that nearly four orders of magnitude in spatial scales lie unresolved below the current spatial resolution limit of magnetograms (as represented by Hinode, cf. Sect. 6.1). In contrast, the Hanle effect is almost blind to the flux-tube like fields, for three reasons: (1) The effect scales with the filling factor, which is very tiny for the flux-tube fields (of order 1%). (2) The Hanle effect is insensitive to vertical fields, and the strong fields tend to be nearly vertical due to the strong buoyancy forces acting on them. (3) The Hanle effect completely saturates for fields stronger than a few hundred gauss. The complementary nature of the two effects has in the past led to the choice of two apparently contradictory interpretative models used for each effect: for the Zeeman effect the two-component model (or extended variations thereof, with additional components) with the concept of a magnetic filling factor, for the Hanle effect a volume-filling field (filling factor of unity) with an isotropic distribution of field vectors.

This apparent dichotomy in the diagnostic methods arises because each of the Zeeman and Hanle effects provides an incomplete, filtered view of the underlying reality, which in a unified picture is fractal-like, and which may best be characterized in terms of probability distribution functions (PDFs). When we “put on our Zeeman goggles”, we project out properties of the strong-field tail of the PDF, which appears flux-tube like. When, on the other hand, we “put on our Hanle goggles”, we project out properties of the weak-field portion of the PDF. However, the application of unified PDF models for both Zeeman and Hanle diagnostics is still in its infancy, and the initial results are only tentative, because the information we have that could guide our choice of distribution functions is still very incomplete.

The incompleteness mainly lies in the lack of information on the angular distribution function of the field, not so much in the PDF for the field strengths, for which we have reasonably good analytical functions to work with. The angular distribution is expected to be closely coupled to the field-strength distribution. From theoretical considerations we expect the stronger fields to have an angular distribution that is fairly peaked around the vertical direction, since they are more affected by the vertical buoyancy forces while resisting bending and tangling by the turbulent motions. The weakest fields on the other hand are expected to have a much wider angular distribution, since the dominating effect is the turbulent tangling of the passive fields. For intermediate field strengths we should have a gradual transition between the wide and the peaked angular distributions. The Stokes profile signatures from such combinations of distribution functions for the spatially unresolved magnetic fields have recently been explored by radiative-transfer modeling (Sampoorna et al. 2008), but such calculations have not yet been applied to model fitting of observational data.

A unique opportunity to obtain lacking observational information on the angular distribution functions would be with the SOT data from the Hinode spacecraft. A detailed exploration of the distribution functions of the quiet-sun vertical and horizontal magnetic fields with Hinode data (Lites et al. 2008) has given the surprising result that there seems to be five times more horizontal magnetic flux than vertical flux. Furthermore, the patches of flux concentrations of vertical and horizontal fields are observed to be well separated, rather than co-spatial. There is convincing indirect evidence that most of the horizontal flux patches are not spatially resolved even with Hinode but have a small filling factor, indicating intrinsic sizes of the underlying flux elements of at most 50 km (Ishikawa et al. 2008). These intriguing results are not yet properly understood, so the angular distribution functions needed for our diagnostic models still remain elusive.

7 Conclusion

While we have attempted to give a comprehensive overview of small-scale magnetic field in the solar atmosphere, this review is by no means complete. In particular, we have not touched upon chromospheric fields, which besides being structured on small scales, also display dynamic behaviour on short timescales.

New instruments that are able to measure photospheric magnetic field in high-resolution and at sufficient cadence to study dynamics, either directly through (spectro)polarimetry, or indirectly through proxy-magnetometry, are now available to observers. In particular, we have discussed several results from the Hinode mission. These results, while dealing with small-scale field, have great repercussions for important questions surrounding magnetism in the Sun, and in particular for the existence and workings of both the local and global solar dynamos.

With new instruments and sophisticated modeling enabled by advances in computing, we have greatly improved our understanding of magnetic activity on all scales in the Sun. Yet, we have also seen that the end is not yet in sight: field is likely structured on scales well beyond what can be observed or simulated today or in the foreseeable future. Our understanding of the processes that give rise to small-scale magnetic field will continue to improve as more observations are analyzed, models become more sophisticated and lifelike, and new instruments are developed, such as the sunrise balloon-borne observatory (Gandorfer et al. 2006) or the Advanced Technology Solar Telescope (Keil et al. 2000).

References

W.P. Abbett, Astrophys. J. **665**, 1469 (2007)
V.I. Abramenko, Sol. Phys. **228**, 29 (2005)
A. Asensio Ramos, M.J. Martínez González, A. López Ariste, J. Trujillo Bueno, M. Collados, Astrophys. J. **659**, 829 (2007)
C. Beck, L.R. Bellot Rubio, R. Schlichenmaier, P. Sütterlin, Astron. Astrophys. **472**, 607 (2007)
J.M. Beckers, E.H. Schröter, Sol. Phys. **4**, 142 (1968)
S.V. Berdyugina, D.M. Fluri, Astron. Astrophys. **417**, 775 (2004)
T.E. Berger, A.M. Title, Astrophys. J. **463**, 365 (1996)
T.E. Berger, A.M. Title, Astrophys. J. **553**, 449 (2001)
T.E. Berger, C.J. Schrijver, R.A. Shine et al., Astrophys. J. **454**, 531 (1995)
T.E. Berger, M.G. Löfdahl, R.A. Shine, A.M. Title, Astrophys. J. **506**, 439 (1998a)
T.E. Berger, M.G. Löfdahl, R.S. Shine, A.M. Title, Astrophys. J. **495**, 973 (1998b)
T.E. Berger, L.H.M. Rouppe van der Voort, M.G. Löfdahl et al., Astron. Astrophys. **428**, 613 (2004)
T.E. Berger, L. Rouppe van der Voort, M. Löfdahl, Astrophys. J. **661**, 1272 (2007)
J. Blanco Rodríguez, O.V. Okunev, K.G. Puschmann, F. Kneer, B. Sánchez-Andrade Nuño, Astron. Astrophys. **474**, 251 (2007)
V. Bommier, M. Derouich, E. Landi degl'Innocenti, G. Molodij, S. Sahal-Bréchot, Astron. Astrophys. **432**, 295 (2005)
V. Bommier, E. Landi Degl'Innocenti, N. Feautrier, G. Molodij, Astron. Astrophys. **458**, 625 (2006)
J.H.M.J. Bruls, S.K. Solanki, Astron. Astrophys. **293**, 240 (1995)
M. Carlsson, R.F. Stein, Å. Nordlund, G.B. Scharmer, Astrophys. J. **610**, L137 (2004)
F. Cattaneo, Astrophys. J. **515**, L39 (1999)
R. Centeno, H. Socas-Navarro, B. Lites et al., Astrophys. J. **666**, L137 (2007)
S. Criscuoli, M.P. Rast, I. Ermolli, M. Centrone, Astron. Astrophys. **461**, 331 (2007)
A.G. De Wijn, R.J. Rutten, E.M.W.P. Haverkamp, P. Sütterlin, Astron. Astrophys. **441**, 1183 (2005)
A.G. De Wijn, B.W. Lites, T.E. Berger et al., Astrophys. J. **684**, 1469–1476 (2008)
J.C. del Toro Iniesta, *Introduction to Spectropolarimetry* (Cambridge University Press, Cambridge, 2003)
M. Derouich, V. Bommier, J.M. Malherbe, E. Landi Degl'Innocenti, Astron. Astrophys. **457**, 1047 (2006)
I. Domínguez Cerdeña, F. Kneer, J. Sánchez Almeida, Astrophys. J. **582**, L55 (2003a)
I. Domínguez Cerdeña, J. Sánchez Almeida, F. Kneer, Astron. Astrophys. **407**, 741 (2003b)
I. Domínguez Cerdeña, J.S. Almeida, F. Kneer, Astrophys. J. **646**, 1421 (2006a)
I. Domínguez Cerdeña, J. Sánchez Almeida, F. Kneer, Astrophys. J. **636**, 496 (2006b)
R.B. Dunn, J.B. Zirker, Sol. Phys. **33**, 281 (1973)
D.F. Elmore, B.W. Lites, S. Tomczyk et al., in Proc. SPIE, vol. 1746, Polarization analysis and measurement, ed. D.H. Goldstein, R.A. Chipman (1992), pp. 22–33
M. Faurobert, J. Arnaud, J. Vigneau, H. Frisch, Astron. Astrophys. **378**, 627 (2001)
M. Faurobert-Scholl, Astron. Astrophys. **268**, 765 (1993)
M. Faurobert-Scholl, N. Feautrier, F. Machefert, K. Petrovay, A. Spielfiedel, Astron. Astrophys. **298**, 289 (1995)
E.N. Frazier, J.O. Stenflo, Sol. Phys. **27**, 330 (1972)
A.M. Gandorfer, H.P.P.P. Steiner, F. Aebersold et al., Astron. Astrophys. **422**, 703 (2004)
A.M. Gandorfer, S.K. Solanki, P. Barthol et al., in Proc. SPIE, vol. 6267, Ground-based and Airborne Telescopes, ed. L.M. Stepp (2006)
U. Grossmann-Doerth, C.U. Keller, M. Schüssler, Astron. Astrophys. **315**, 610 (1996)
U. Grossmann-Doerth, M. Schüssler, O. Steiner, Astron. Astrophys. **337**, 928 (1998)
H.J. Hagenaar, Astrophys. J. **555**, 448 (2001)
K.L. Harvey, PhD thesis, Univ. Utrecht, 1993
K.L. Harvey, S.F. Martin, Sol. Phys. **32**, 389 (1973)
K.L. Harvey, J.W. Harvey, S.F. Martin, Sol. Phys. **40**, 87 (1975)
J.W. Harvey, D. Branston, C.J. Henney, C.U. Keller, Astrophys. J. **659**, L177 (2007)
T. Homann, F. Kneer, V.I. Makarov, Sol. Phys. **175**, 81 (1997)
R. Howard, J.O. Stenflo, Sol. Phys. **22**, 402 (1972)
K. Ichimoto, B. Lites, D. Elmore et al., Sol. Phys. **249**, 233 (2008)
R. Ishikawa, S. Tsuneta, Astron. Astrophys. (2009a, in press)
R. Ishikawa, S. Tsuneta, Astron. Astrophys. (2009b, in preparation)
R. Ishikawa, S. Tsuneta, Y. Kitakoshi et al., Astron. Astrophys. **472**, 911 (2007)
R. Ishikawa, S. Tsuneta, K. Ichimoto et al., Astron. Astrophys. **481**, L25 (2008)
H. Isobe, M.R.E. Proctor, N.O. Weiss, Astrophys. J. **679**, L57 (2008)
K. Janßen, A. Vögler, F. Kneer, Astron. Astrophys. **409**, 1127 (2003)

H. Jeong, J. Chae, Astrophys. J. **671**, 1022 (2007)
S.L. Keil, T.R. Rimmele, C. Keller, F. Hill, Bull. Am. Astron. Soc. **32**, 1433 (2000)
C.U. Keller, Nature **359**, 307 (1992)
C.U. Keller, F.-L. Deubner, U. Egger, B. Fleck, H.P. Povel, Astron. Astrophys. **286**, 626 (1994)
C.U. Keller, M. Schüssler, A. Vögler, V. Zakharov, Astrophys. J. **607**, L59 (2004)
E. Khomenko, M. Collados, Astrophys. J. **659**, 1726 (2007)
E.V. Khomenko, M. Collados, S.K. Solanki, A. Lagg, J. Trujillo Bueno, Astron. Astrophys. **408**, 1115 (2003)
E.V. Khomenko, M.J. Martínez González, M. Collados et al., Astron. Astrophys. **436**, L27 (2005a)
E.V. Khomenko, S. Shelyag, S.K. Solanki, A. Vögler, Astron. Astrophys. **442**, 1059 (2005b)
D. Kiselman, R.J. Rutten, B. Plez, in IAU Symposium, vol. 203, Recent Insights into the Physics of the Sun and Heliosphere: Highlights from SOHO and Other Space Missions, ed. P. Brekke, B. Fleck, J.B. Gurman (2001), p. 287
M. Knölker, M. Schüssler, Astron. Astrophys. **202**, 275 (1988)
R.W. Komm, Sol. Phys. **157**, 45 (1995)
T. Kosugi, K. Matsuzaki, T. Sakao et al., Sol. Phys. **243**, 3 (2007)
J.K. Lawrence, A.C. Cadavid, A.A. Ruzmaikin, Phys. Rev. E **51**, 316 (1995)
H. Lin, Astrophys. J. **446**, 421 (1995)
H. Lin, T. Rimmele, Astrophys. J. **514**, 448 (1999)
H. Lin, J. Varsik, H. Zirin, Sol. Phys. **155**, 243 (1994)
B.W. Lites, Appl. Opt. **26**, 3838 (1987)
B.W. Lites, Sol. Phys. **163**, 223 (1996)
B.W. Lites, H. Socas-Navarro, Astrophys. J. **613**, 600 (2004)
B.W. Lites, K.D. Leka, A. Skumanich, V. Martinez Pillet, T. Shimizu, Astrophys. J. **460**, 1019 (1996)
B.W. Lites, M. Kubo, H. Socas-Navarro et al., Astrophys. J. **672**, 1237 (2008)
W. Livingston, J. Harvey, in IAU Symposium, vol. 43, Solar Magnetic Fields, ed. R. Howard (1971), p. 51
W.C. Livingston, J. Harvey, Bulletin of the American Astronomical Society **7**, 346 (1975)
D. Longcope, C. Beveridge, J. Qiu et al., Sol. Phys. **244**, 45 (2007)
A. López Ariste, S. Tomczyk, R. Casini, Astrophys. J. **580**, 519 (2002)
A. López Ariste, S. Tomczyk, R. Casini, Astron. Astrophys. **454**, 663 (2006)
T. Magara, Astrophys. J. **549**, 608 (2001)
T. Magara, S. Tsuneta, Publ. Astron. Soc. Jpn. **60**, 1181–1189 (2008)
S.F. Martin, Sol. Phys. **117**, 243 (1988)
M.J. Martínez González, M. Collados, B. Ruiz Cobo, Astron. Astrophys. **456**, 1159 (2006)
M.J. Martínez González, M. Collados, B. Ruiz Cobo, S.K. Solanki, Astron. Astrophys. **469**, L39 (2007)
M.J. Martínez González, A. Asensio Ramos, A. López Ariste, R. Manso Sainz, Astron. Astrophys. **479**, 229 (2008)
V. Martinez Pillet, B.W. Lites, A. Skumanich, Astrophys. J. **474**, 810 (1997)
J.P. Mehltretter, Sol. Phys. **38**, 43 (1974)
N. Meunier, Astrophys. J. **515**, 801 (1999)
N. Meunier, Astron. Astrophys. **420**, 333 (2004)
N. Meunier, S.K. Solanki, W.C. Livingston, Astron. Astrophys. **331**, 771 (1998)
F. Moreno-Insertis, P. Caligari, M. Schüssler, Astrophys. J. **452**, 894 (1995)
R. Muller, Sol. Phys. **52**, 249 (1977)
R. Muller, Sol. Phys. **85**, 113 (1983)
R. Muller, Sol. Phys. **100**, 237 (1985)
R. Muller, T. Roudier, Sol. Phys. **94**, 33 (1984)
D.A.N. Müller, O. Steiner, R. Schlichenmaier, P.N. Brandt, Sol. Phys. **203**, 211 (2001)
S. Nagata, S. Tsuneta, Y. Suematsu et al., Astrophys. J. **677**, L145 (2008)
E. Nesme-Ribes, N. Meunier, B. Collin, Astron. Astrophys. **308**, 213 (1996)
P. Nisenson, A.A. van Ballegooijen, A.G. De Wijn, P. Sütterlin, Astrophys. J. **587**, 458 (2003)
O.V. Okunev, F. Kneer, Astron. Astrophys. **425**, 321 (2004)
D. Orozco Suárez, L.R. Bellot Rubio, J.C. Del Toro Iniesta et al., Publ. Astron. Soc. Jpn. **59**, 837 (2007a)
D. Orozco Suárez, L.R. Bellot Rubio, J.C. del Toro Iniesta et al., Astrophys. J. **670**, L61 (2007b)
E.N. Parker, Astrophys. J. **221**, 368 (1978)
S.R.O. Ploner, M. Schüssler, S.K. Solanki, A.S. Gadun, in Astronomical Society of the Pacific Conference Series, vol. 236, Advanced Solar Polarimetry—Theory, Observation, and Instrumentation, ed. M. Sigwarth (2001) p. 363
G.W. Pneuman, S.K. Solanki, J.O. Stenflo, Astron. Astrophys. **154**, 231 (1986)
H. Povel, Opt. Eng. **34**, 1870 (1995)
D. Rabin, Astrophys. J. **390**, L103 (1992a)
D. Rabin, Astrophys. J. **391**, 832 (1992b)

T. Roudier, R. Muller, Sol. Phys. **107**, 11 (1987)
T. Roudier, R. Muller, Astron. Astrophys. **419**, 757 (2004)
T. Roudier, F. Lignières, M. Rieutord, P.N. Brandt, J.M. Malherbe, Astron. Astrophys. **409**, 299 (2003)
L.H.M. Rouppe van der Voort, V.H. Hansteen, M. Carlsson et al., Astron. Astrophys. **435**, 327 (2005)
I. Rüedi, S.K. Solanki, W. Livingston, J.O. Stenflo, Astron. Astrophys. **263**, 323 (1992)
M. Sampoorna, K.N. Nagendra, H. Frisch, J.O. Stenflo, Astron. Astrophys. **485**, 275–287 (2008)
J. Sánchez Almeida, Astron. Astrophys. **438**, 727 (2005)
J. Sanchez Almeida, E. Landi degl'Innocenti, V. Martinez Pillet, B.W. Lites, Astrophys. J. **466**, 537 (1996)
J. Sánchez Almeida, A. Asensio Ramos, J. Trujillo Bueno, J. Cernicharo, Astrophys. J. **555**, 978 (2001)
J. Sánchez Almeida, I. Domínguez Cerdeña, F. Kneer, Astrophys. J. **597**, L177 (2003)
J. Sánchez Almeida, I. Márquez, J.A. Bonet, I. Domínguez Cerdeña, R. Muller, Astrophys. J. **609**, L91 (2004)
J. Sánchez Almeida, B. Viticchié, E. Landi Degl'Innocenti, F. Berrilli, Astrophys. J. **675**, 906 (2008)
C.J. Schrijver, C. Zwaan, Cambridge Astrophysics Series, vol. 34, Solar and Stellar Magnetic Activity (Cambridge University Press, Cambridge, 2000)
C.J. Schrijver, A.M. Title, A.A. van Ballegooijen, H.J. Hagenaar, R.A. Shine, Astrophys. J. **487**, 424 (1997)
M. Schüssler, A. Vögler, Astron. Astrophys. **481**, L5 (2008)
M. Schüssler, S. Shelyag, S. Berdyugina, A. Vögler, S.K. Solanki, Astrophys. J. **597**, L173 (2003)
N.R. Sheeley Jr., Sol. Phys. **1**, 171 (1967)
N.R. Sheeley Jr., A.G. Nash, Y.-M. Wang, Astrophys. J. **319**, 481 (1987)
S. Shelyag, M. Schüssler, S.K. Solanki, S.V. Berdyugina, A. Vögler, Astron. Astrophys. **427**, 335 (2004)
T. Shimizu, S. Nagata, S. Tsuneta et al., Sol. Phys. **249**, 221 (2008)
H.B. Snodgrass, Astrophys. J. **270**, 288 (1983)
H. Socas-Navarro, B.W. Lites, Astrophys. J. **616**, 587 (2004)
H. Socas-Navarro, J. Sánchez Almeida, Astrophys. J. **593**, 581 (2003)
H. Socas-Navarro, V. Martínez Pillet, B.W. Lites, Astrophys. J. **611**, 1139 (2004)
H. Socas-Navarro, J.M. Borrero, A. Asensio Ramos et al., Astrophys. J. **674**, 596 (2008)
S.K. Solanki, Space Sci. Rev. **63**, 1 (1993)
S.K. Solanki, J.O. Stenflo, Astron. Astrophys. **140**, 185 (1984)
S.K. Solanki, C. Keller, J.O. Stenflo, Astron. Astrophys. **188**, 183 (1987)
S.K. Solanki, D. Zufferey, H. Lin, I. Rüedi, J.R. Kuhn, Astron. Astrophys. **310**, L33 (1996)
S.K. Solanki, A. Lagg, J. Woch, N. Krupp, M. Collados, Nature **425**, 692 (2003)
H.C. Spruit, Sol. Phys. **50**, 269 (1976)
H.C. Spruit, PhD thesis. University of Utrecht, The Netherlands, 1977
H.C. Spruit, Sol. Phys. **61**, 363 (1979)
H.C. Spruit, C. Zwaan, Sol. Phys. **70**, 207 (1981)
H.C. Spruit, E.G. Zweibel, Sol. Phys. **62**, 15 (1979)
R.F. Stein, Å. Nordlund, Astrophys. J. **642**, 1246 (2006)
O. Steiner, Astron. Astrophys. **430**, 691 (2005)
O. Steiner, G.W. Pneuman, J.O. Stenflo, Astron. Astrophys. **170**, 126 (1986)
O. Steiner, U. Grossmann-Doerth, M. Knölker, M. Schüssler, Astrophys. J. **495**, 468 (1998)
O. Steiner, P.H. Hauschildt, J. Bruls, Astron. Astrophys. **372**, L13 (2001)
J.O. Stenflo, Arkiv for Astronomi **4**, 173 (1966)
J.O. Stenflo, Sol. Phys. **32**, 41 (1973)
J.O. Stenflo, Sol. Phys. **80**, 209 (1982)
J.O. Stenflo, Sol. Phys. **114**, 1 (1987)
J.O. Stenflo, Astron. Astrophys. **210**, 403 (1989)
J.O. Stenflo, *Solar Magnetic Fields: Polarized Radiation Diagnostics* (Kluwer, Dordrecht, 1994)
J.O. Stenflo, Nature **430**, 304 (2004)
J.O. Stenflo, J.W. Harvey, Sol. Phys. **95**, 99 (1985)
J.O. Stenflo, R. Holzreuter, in ESA Special Publication, vol. 505, SOLMAG, Proceedings of the Magnetic Coupling of the Solar Atmosphere Euroconference, ed. H. Sawaya-Lacoste (2002), pp. 101–104
J.O. Stenflo, R. Holzreuter, in Astronomical Society of the Pacific Conference Series, vol. 286, Current Theoretical Models and Future High Resolution Solar Observations: Preparing for ATST, ed. A.A. Pevtsov, H. Uitenbroek (2003), p. 169
J.O. Stenflo, L. Lindegren, Astron. Astrophys. **59**, 367 (1977)
J.O. Stenflo, S. Solanki, J.W. Harvey, J.W. Brault, Astron. Astrophys. **131**, 333 (1984)
J.O. Stenflo, S.K. Solanki, J.W. Harvey, Astron. Astrophys. **173**, 167 (1987)
J.O. Stenflo, M. Bianda, C.U. Keller, S.K. Solanki, Astron. Astrophys. **322**, 985 (1997)
J.O. Stenflo, C.U. Keller, A. Gandorfer, Astron. Astrophys. **329**, 319 (1998)
Y. Suematsu, S. Tsuneta, K. Ichimoto et al., Sol. Phys. **249**, 197 (2008)
F. Tang, H. Wang, Sol. Phys. **132**, 247 (1991)

T.D. Tarbell, A.M. Title, S.A. Schoolman, Astrophys. J. **229**, 387 (1979)
A.M. Title, C.J. Schrijver, in Astronomical Society of the Pacific Conference Series, vol. 154, Cool Stars, Stellar Systems, and the Sun, ed. R.A. Donahue, J.A. Bookbinder (1998), p. 345
J. Trujillo Bueno, N. Shchukina, A. Asensio Ramos, Nature **430**, 326 (2004)
S. Tsuneta, K. Ichimoto, Y. Katsukawa et al., Astrophys. J. **688**, 1374 (2008a)
S. Tsuneta, K. Ichimoto, Y. Katsukawa et al., Sol. Phys. **249**, 167 (2008b)
H. Uitenbroek, A. Tritschler, Astrophys. J. **639**, 525 (2006)
W. Unno, Astrophys. J. **129**, 375 (1959)
A.A. van Ballegooijen, P. Nisenson, R.W. Noyes et al., Astrophys. J. **509**, 435 (1998)
P. Venkatakrishnan, Nature **322**, 156 (1986)
A. Vögler, M. Schüssler, Astron. Astrophys. **465**, L43 (2007)
A. Vögler, S. Shelyag, M. Schüssler et al., Astron. Astrophys. **429**, 335 (2005)
Y.-M. Wang, N.R. Sheeley Jr., Astrophys. J. **430**, 399 (1994)
J. Wang, H. Wang, F. Tang, J.W. Lee, H. Zirin, Sol. Phys. **160**, 277 (1995)
E. Wiehr, Astron. Astrophys. **69**, 279 (1978)
P.R. Wilson, Sol. Phys. **69**, 9 (1981)
H. Zirin, Sol. Phys. **110**, 101 (1987)
H. Zirin, B. Popp, Astrophys. J. **340**, 571 (1989)

Coupling from the Photosphere to the Chromosphere and the Corona

S. Wedemeyer-Böhm · A. Lagg · Å. Nordlund

Originally published in the journal Space Science Reviews, Volume 144, Nos 1–4, 317–350.
DOI: 10.1007/s11214-008-9447-8

Abstract The atmosphere of the Sun is characterized by a complex interplay of competing physical processes: convection, radiation, conduction, and magnetic fields. The most obvious imprint of the solar convection and its overshooting in the low atmosphere is the granulation pattern. Beside this dominating scale there is a more or less smooth distribution of spatial scales, both towards smaller and larger scales, making the Sun essentially a multi-scale object. Convection and overshooting give the photosphere its face but also act as drivers for the layers above, namely the chromosphere and corona. The magnetic field configuration effectively couples the atmospheric layers on a multitude of spatial scales, for instance in the form of loops that are anchored in the convection zone and continue through the atmosphere up into the chromosphere and corona. The magnetic field is also an important structuring agent for the small, granulation-size scales, although (hydrodynamic) shock waves also play an important role—especially in the internetwork atmosphere where mostly weak fields prevail. Based on recent results from observations and numerical simulations, we attempt to present a comprehensive picture of the atmosphere of the quiet Sun as a highly intermittent and dynamic system.

Keywords Sun · Photosphere · Chromosphere · Corona · Convection · Magnetohydrodynamics · Radiative transfer

S. Wedemeyer-Böhm (✉)
Institute of Theoretical Astrophysics, University of Oslo, Oslo, Norway
e-mail: svenwe@astro.uio.no

A. Lagg
MPI für Sonnensystemforschung, Katlenburg-Lindau, Germany
e-mail: lagg@mps.mpg.de

Å. Nordlund
Niels Bohr Institute, University of Copenhagen, Copenhagen, Denmark
e-mail: aake@nbi.dk

M.J. Thompson et al. (eds.), *The Origin and Dynamics of Solar Magnetism*.
DOI: 10.1007/978-1-4419-0239-9_17

1 Introduction

Observations of the solar atmosphere reveal a wealth of different phenomena, which occur over an extended range of different temporal and spatial scales. This is not surprising, considering the fact that already basic parameters such as gas density and temperature span many orders of magnitude, from the convection zone below the photosphere to the corona. At a first look, it may thus appear rather hopeless to construct an overall picture that can account for all the phenomena. At a closer look, however, many connections between apparently independent phenomena can be found, ultimately implying a multitude of couplings through the atmosphere. In addition, there seems to be a hierarchical arrangement of approximately selfsimilar convective motions, with the granulation pattern embedded in increasingly larger meso- and supergranulation patterns.

The key to a comprehensive picture of the solar atmosphere thus lies in relaxing too strict and oversimplified concepts, even when they are didactically nicer than the reality. The solar atmosphere should not be seen as a static stack of layers but rather as intermittent domains that are dynamically coupled together. One example is magnetic flux structures (or "flux tubes") fanning out with a wine-glass geometry. Such regular building blocks put certain constraints on the implied atmospheric structure, which can make it difficult to fit in other observational findings. Accepting that magnetic field structures are far less regular offers room for a more generally valid comprehensive picture. This trend became more and more obvious during the recent years, both from the observational and theoretical side (see, e.g., Carlsson 2007; Gudiksen 2006; Hansteen 2007; Judge 2006; Rutten 2007; Steiner 2007 and many more).

The advantages of a relaxed picture can be seen from the example of the quiet Sun chromosphere above internetwork regions, which in itself is a complex and intriguing phenomenon (see, e.g., Judge 2006; Rutten 2006; De Pontieu et al. 2004; Lites et al. 1999 and many more). Despite tremendous progress, there are still many open questions concerning its structure, dynamics and energy balance. Recent observations now prove—beyond any doubt—the chromosphere to be a highly dynamic and intermittent layer. The internetwork chromosphere is the product of a dynamic interplay of shock waves and magnetic fields. This picture, which was already suggested by many earlier investigations, offers a key to resolve some apparent contradictions that lead to much confusion in the past. A prominent example concerns the observation of carbon monoxide (see, e.g., Ayres 2002), which now can be explained as an integral part of a dynamic and intermittent atmosphere (Wedemeyer-Böhm and Steffen 2007; Wedemeyer-Böhm et al. 2005a, 2006). And still the chromosphere cannot be investigated without also taking into account the layers above and below. The shock waves, which are so essential at least for the lowest, weak-field parts of the chromosphere, are generated in the layers below, while significant amounts of mass and energy are exchanged between the chromosphere and the corona above. Obviously, the whole atmosphere must be seen as an integral phenomenon.

In the following sections, we report on a selection of results from observations and numerical simulations, which will help us put together an updated, revised view of the structure of the quiet Sun atmosphere.

2 The Sun—A Multi-Scale Object

An overarching point in this discussion is the fact that the Sun is fundamentally a multi-scale object. This is a major difficulty for modeling and understanding, since it requires (computationally expensive) modeling over a large range of scales.

But the Sun also displays aspects of self-similarity and scale invariance in several respects, which on the other hand helps a lot. To illustrate the self-similarity, Fig. 1a shows temperature patterns in horizontal planes in a large scale simulation of solar convection (Zhao et al. 2007), and Fig. 1b shows patterns of vertical velocities from the same simulation. The temperature patterns show very intermittent cold structures, embedded in a background of horizontally nearly constant temperature (images of entropy would look essentially identical, with near-constancy also in the vertical direction). The set of panels also shows that the pattern scales increase systematically with depth.

Figure 1b, on the other hand, which displays vertical velocity on a color scale that changes from yellow to blue with sign (with a narrow band of grey for velocities near zero) gives a completely different impression. With this rendering choice one can see that, at least from a morphological point of view, the patterns at different depths are quite similar. Displaying in this way, signed velocity reveals that the sharply defined dark (cold) patterns in Fig. 1a indeed correspond to the strongest downward velocities, but that there are also relatively broad areas of much milder downflows. This shows that, as the ascending gas is forced (by mass conservation) to overturn, it does so at first gently, then to finally be accelerated more strongly by the positive feedback that comes from merging with colder gas from above.

At the visible surface the horizontal velocity patterns from various depths are superposed. This happens because the depth dependencies of the large scale horizontal velocity patterns are rather weak; at least over depth intervals small compared to their horizontal extents. On the other hand, as illustrated by Fig. 1, the dominant scale becomes smaller for layers increasingly close to the surface. As the amplitudes of these smaller scales are larger, they mask the presence of the larger scale patterns, whose presence, however, can still be revealed, e.g., with Fourier analysis or with low-pass filtering. The hierarchy is illustrated in a side view in Fig. 2. The combination of streamlines and colors illustrate how near-surface, small scales fluctuations are carried along in larger scale flows.

The hierarchy of scales displayed reveals no particular preferred scale above the granular one; the transition to larger and larger scales with depth is smooth. A direct way to illustrate this from observations is to use power spectra of solar velocities, as observed with SOHO/MID (Georgobiani et al. 2007). Figure 3 shows velocity (mainly horizontal) as a function of size, produced by filtering the velocity power observed by MDI into sonic and sub-sonic parts. The velocity spectrum displayed is produced by then taking the square root of the velocity power times wave number; this is a quantity—a velocity spectrum—that nicely illustrates the dependence of velocity amplitude on size. Note that there is very little (less than a factor two) extra power at scales traditionally associated with supergranulation, and that there is a smooth and increasing distribution of velocity amplitude across the "meso-granulation" range to granulation scales. The same behavior is found in large scale numerical simulations (Georgobiani et al. 2007).

Supergranulation patterns can be brought forward by averaging over either time or space; the dashed line in Fig. 3 shows the effect of a 24-hour time average. A low-pass wavenumber spatial filter has a similar effect; it cuts away the larger amplitudes at smaller scales and exposes aspects of the underlying larger scale pattern. The relatively distinct appearance of a supergranulation scale network in magnetically related diagnostics indicates that the transport and diffusion of magnetic field structures at the solar surface results in what is effectively a low-pass wavenumber filter.

As shown by Fig. 4 it is practically impossible to tell the difference between velocity patterns on different scales, once they are filtered to effectively the same resolution. As illustrated by Stenflo and Holzreuter (2002, 2003) magnetic field patterns and distributions also show a degree of self-similarity.

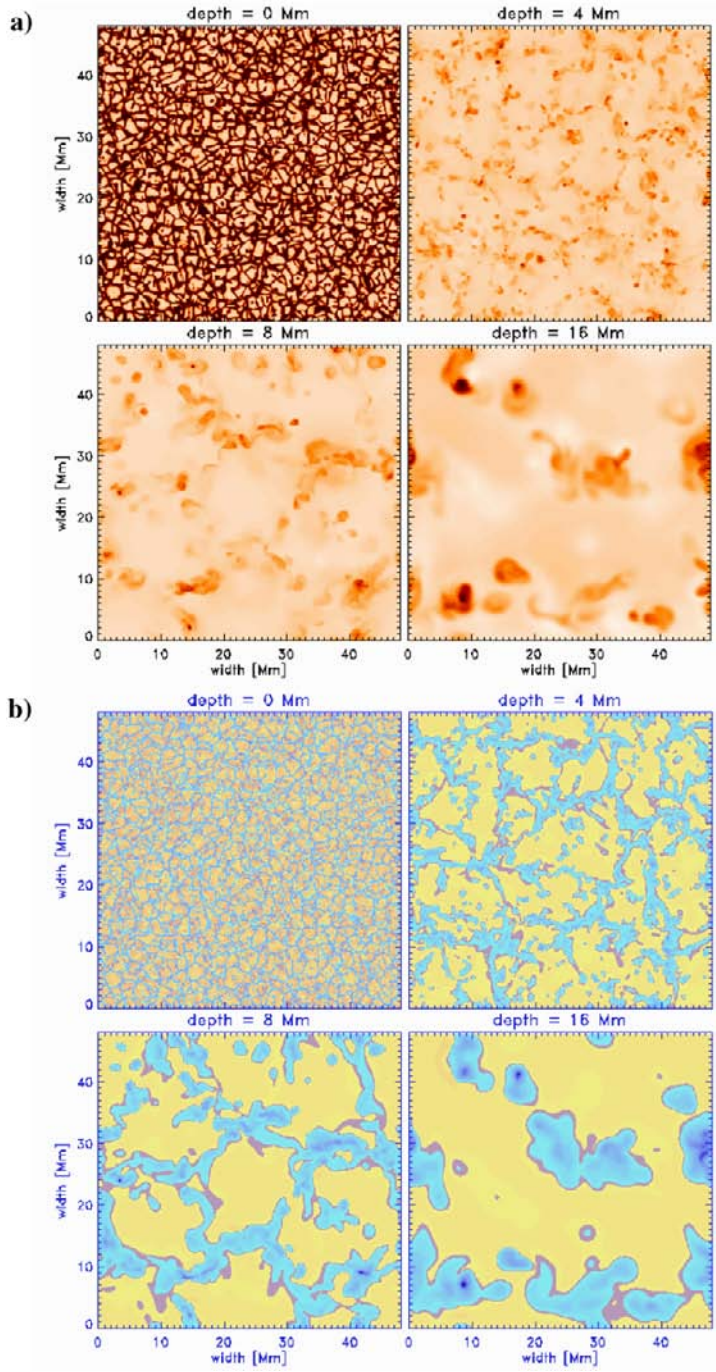


Fig. 1 Large scale solar convection (48 Mm × 48 Mm × 20 Mm). **a** Temperature (*top four panels*) and **b** velocity (*bottom four panels*) patterns at four different depths. Temperature is shown on a linear scale. Velocities are rendered with positive (downward) values *blue* and negative (upward) velocities *yellow*. A narrow band near zero velocity is rendered in *grey*

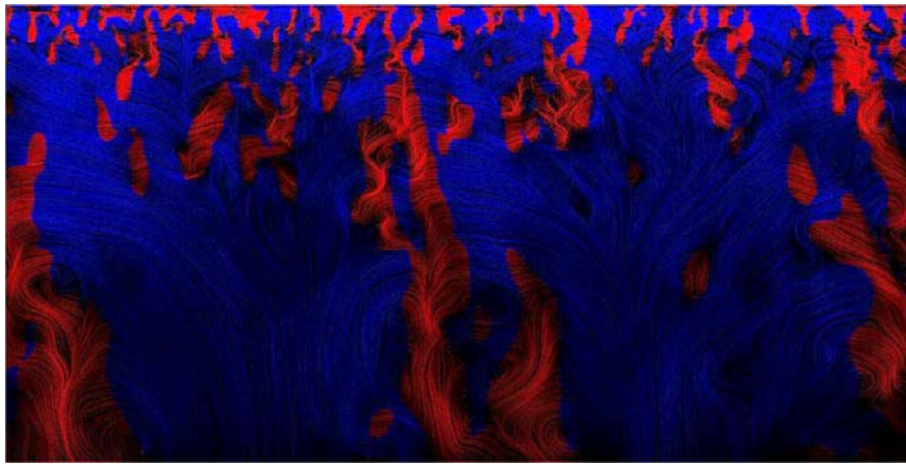


Fig. 2 A side view of a 48 Mm × 48 Mm × 20 Mm simulation, showing velocity streamlines, with brightness increasing with increasing magnitude. Up- and down-flows are rendered in *blue* and *red*, respectively

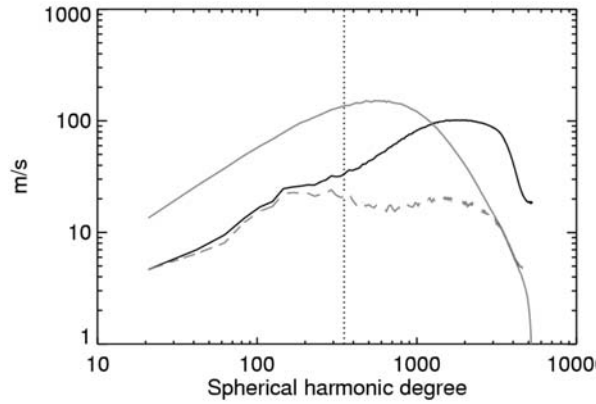


Fig. 3 Solar velocity spectrum: A subsonic (7 km s^{-1}) filter has been used to separate the velocity into oscillatory (*grey*) and convective (*black*) components. The *dashed line* shows the convective component resulting from first taking a 24 h average. Adapted from Georgobiani et al. (2007)

The magnetic energy equation

$$\frac{\partial}{\partial t}\left(\frac{\mu B^2}{2}\right) = -\nabla \cdot (\mathbf{E} \times \mu\mathbf{B}) - \mathbf{v} \cdot (\mathbf{j} \times \mathbf{B}) - Q_{\text{Joule}} \tag{1}$$

illustrates that balance of the magnetic energy at each depth is achieved by Lorentz force work (by the flow on the field) being used to balance magnetic dissipation, with net magnetic energy transported up or down by the Poynting flux, $\mathbf{E} \times \mu\mathbf{B}$. As shown by the work of Vögler and Schüssler (2007) and Steiner et al. (2008) the actual direction of net transport is systematically downwards, at least below the solar surface. It appears likely that there is net dynamo action at each depth in the convection zone, with net magnetic energy delivered to the next layer down. This naturally leads to a pile up near the bottom of the convection zone. The downward transport, which is known from direct studies (Tobias et al. 1998, 2001; Dorch and Nordlund 2001), is often referred to as "turbulent pumping", and is associated with the asymmetry between up and downflows (illustrated in Fig. 1b).

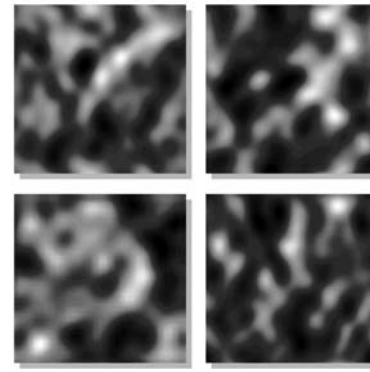


Fig. 4 Solar horizontal velocities observed with SOHO/MDI. A patch some distance away from solar center has been compensated for projection effects and filtered to effective resolutions that differ by factors of 2. Which is which?

On the largest scales (largest depths), and only there, differential rotation enables a large scale global dynamo action, with patterns clearly controlled by being stretched out by differential rotation. Buoyancy eventually pushes the fields back up.

Another evidence for self-similarity comes from the power law behavior of flare energy distribution as a function of time. This behavior is also recovered in numerical simulations of 3D magnetic reconnection (Galsgaard and Nordlund 1996). These are signs that magnetic reconnection occurs in a multi-scale hierarchy, where magnetic dissipation at large magnetic Reynolds number (low resistivity) creates a hierarchy. Large scale structures generate subsidiary small scale structures, which do it again (on shorter time scales) and again, until the spatial scales are small enough to support the dissipation.

Note the remarkable and wonderful argument, made already by Parker a long time ago, which shows that driven magnetic dissipation must, if anything, *increase* with decreasing resistivity—quite contrary to naive expectations. This has been verified in numerical experiments by at least three different groups (Galsgaard and Nordlund 1996; Hendrix et al. 1996; Dmitruk and Gómez 1999).

The chromosphere and corona are likely to be heated in much the same way, as is illustrated by the well known flux-flux relations between coronal and chromospheric diagnostic. It is hard to even avoid, as in models of coronal heating there is a tendency of dumping much more energy in the chromosphere as a side effect (Gudiksen and Nordlund 2002, 2005b).

As pointed out by Phil Judge: The chromosphere is not a mess; the upper chromosphere looks nearly force-free like the corona, whereas the lower chromosphere is less force-free. Complexity comes from both the temperature and density. A central question is: What drives the flows (particularly the cool upflows)?

Semi-realistic models of coupling of the horizontal photospheric velocity field to the corona were first computed by Gudiksen and Nordlund (2005b), who showed that a correctly normalized photospheric (model) velocity field injects sufficient power into the corona to create and maintain coronal temperatures (cf. Fig. 5).

The mechanism is, in the absence of explicit flux emergence, essentially the ‘braiding mechanism’ introduced by Parker (1972, 1981, 1983). The heating is quite intermittent and drives up- and down-flows along magnetic field lines into and out of the corona. Emulated TRACE images and animations show qualitative agreement with observations (Peter et al. 2004). Silicon IV, for example, picks up cooling condensations (cf. Fig. 6). In addition, spectral lines formed at different temperatures show semi-quantitative agreement of the dependency with depth of their Doppler shifts and mean emission measures (cf. Figs. 7 and 9, Peter et al. 2006). The differential emission measure is a ‘fingerprint’ type diagnos-

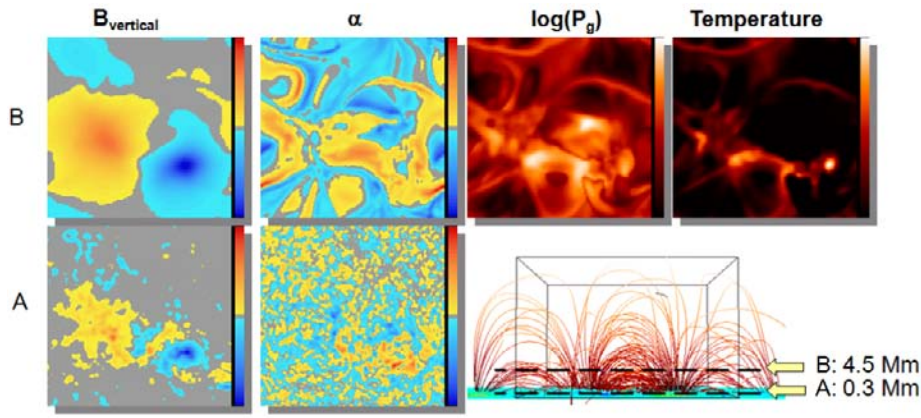


Fig. 5 The panels show the vertical magnetic field (*far left*), the ratio (often referred to with the symbol α) of the magnitude of the electric current along the magnetic field to the magnitude of the magnetic field itself (*left*), the gas pressure (*right*) and the log. temperature (*far right*). The positions of the cutting planes are indicated in the *inset* at *bottom right*. Adapted from Gudiksen and Nordlund (2005b)

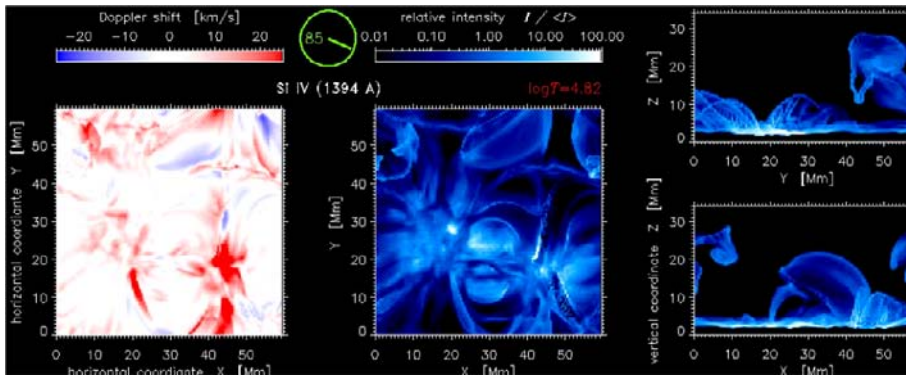


Fig. 6 The Si IV 1394 Å Doppler shift (*left*) and emission measure (*mid*) as they would be observed from above, and the emission measure projected along the Y-axis (*right*). Positive (downward) Doppler shifts are in *red* and negative Doppler shifts are in *blue*. Adapted from Peter et al. (2004)

tic, in much the same way as spectral line asymmetries are for photospheric spectral lines (Dravins et al. 1981; Dravins and Nordlund 1990; Asplund et al. 2000a, 2000b). Subsequently there has been much progress due to the work of the Oslo group (Hansteen and Gudiksen 2005; Hansteen et al. 2006, 2007; Martínez-Sykora et al. 2008)—cf. also the discussion in Sect. 4.3.

3 Observations—Measuring the Magnetic Field in the Solar Atmosphere

The Hα line core images in Figs. 7h and 8 show a well-known but still barely understood and intricate picture: fibrils that spread from regions of enhanced magnetic field strength, occasionally connecting to neighboring regions or apparently fading in between (see e.g. Rutten 2007). The structure gradually changes as one goes from line center into the wings,

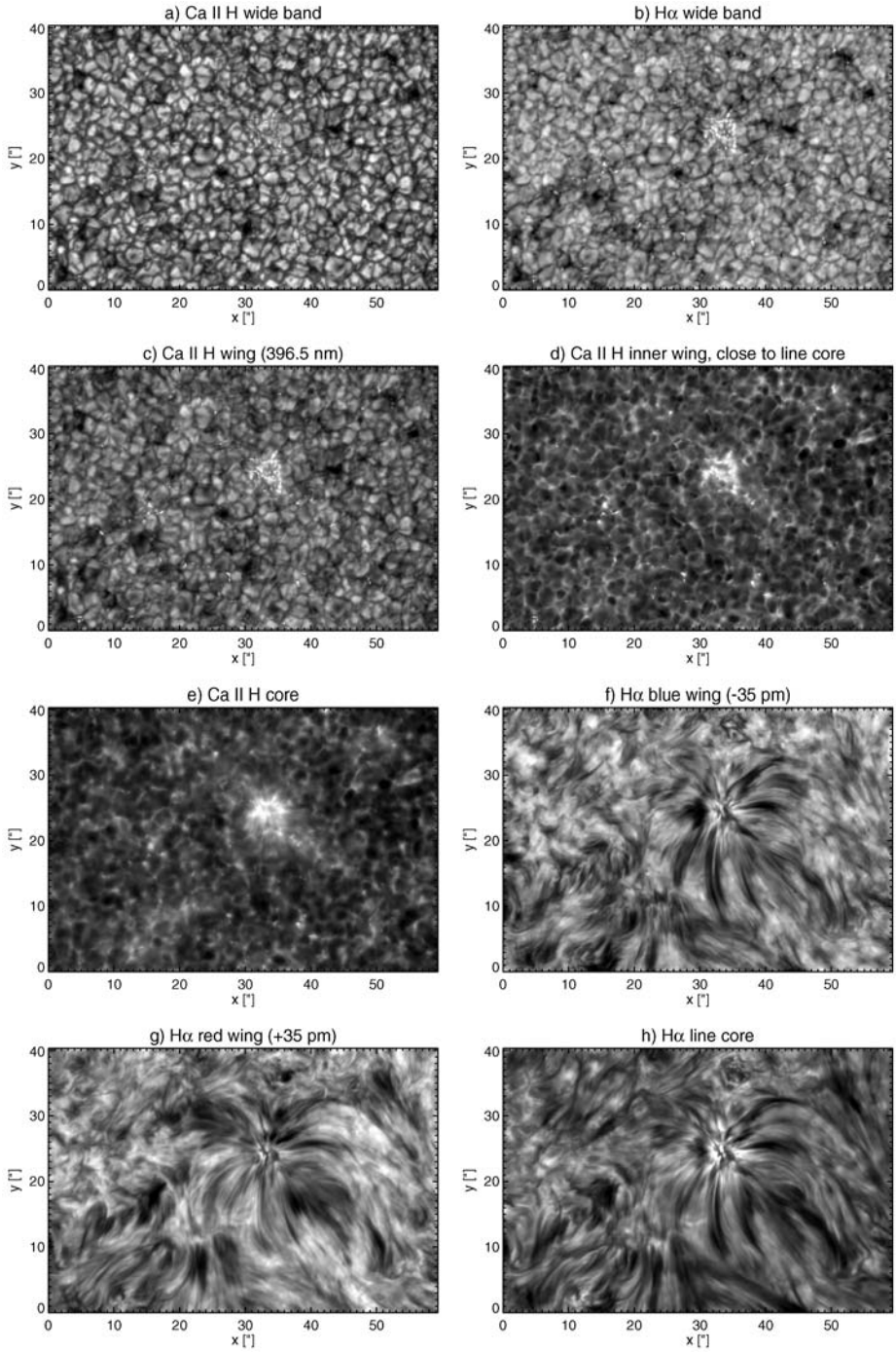


Fig. 7 Observations of a quiet Sun region close to disc-centre: **a** Ca II H wide band, **b** Hα wide band image (FWHM 0.8 nm), **c** Ca II H wing (396.5 nm), **d** Ca II H inner wing (close to line core), **e** Ca II H core, **f** Hα blue wing at -35 pm, **g** Hα red wing $+35$ pm, **h** Hα line core. The observations were carried out with the Swedish 1-m Solar Telescope (SST). Data courtesy: L. Rouppe van der Voort (University of Oslo)

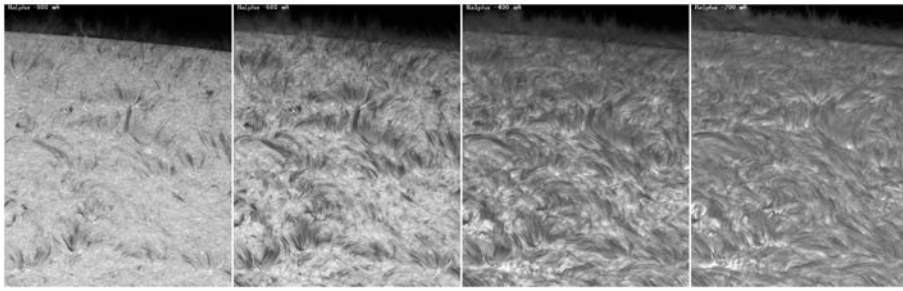


Fig. 8 Hα fine structure at -800, -600, -400, -200 mÅ from line center recorded with the Dutch Open Telescope (DOT) on October 4, 2005 (taken from Rutten 2007). Cell-spanning fibrils are visible around line center (*right*). The decreasing line opacity in the blue wing of the line opens the view to the solar photosphere, intercepted by dark fibrils resulting from the Doppler shift of the line core

as the corresponding intensity is due to lower layers. Finally, a mesh-like background pattern shines through in the internetwork regions. It is most likely due to reversed granulation in the middle photosphere with some possible contributions from the low chromosphere. The gradual change of the pattern in Hα with wavelength gives some clues about the atmospheric structure, in particular the magnetic field in the chromosphere (the "canopy" field), and definitely shows us that the photosphere and chromosphere are coupled via magnetic fields on medium to large spatial scales and via fields and shock waves on the small scales.

Therefore, the understanding of the coupling between photosphere and corona is intimately connected to the measurement of the chromospheric magnetic field. The following subsections exemplify the difficulties of chromospheric magnetic field measurements and present promising approaches to determine the vector magnetic field of the chromosphere.

3.1 Improving Magnetic Field Extrapolations

Measuring the magnetic field in the photosphere has a long tradition (Hale 1908). After 100 years of solar magnetic field measurements, the level of sophistication, both in terms of instrumentation and in analysis technique, has reached a very high level of maturity. Numerous telescopes, ground based and space born (e.g. GONG, MDI), investigate the global structure of the solar magnetic field on a routine basis. High resolution measurements allow the characterization of magnetic elements with a size in the 100 km range (e.g., SST-CRISP, Hinode SP). With HMI on SDO and the balloon-borne 1 m telescope Sunrise (launch: summer 2009) a major improvement in the determination of photospheric magnetic fields will be achieved in both directions—the global magnetic field configuration as well as the smallest scale structures down to a size of 25 km.

The availability of such high quality photospheric vector magnetograms and the low plasma-β in the chromosphere are the basic ingredients needed for reliable, force-free magnetic field extrapolations. Starting with Sakurai (1981) these extrapolations nowadays have reached a high level of sophistication (see reviews by Sakurai 1989; Amari et al. 1997; Wiegelmann 2008). To further improve the accuracy of the chromospheric magnetic field extrapolations additional information on the complex structure of the chromosphere must be taken into account. One of the most promising advances in this direction was proposed by Wiegelmann et al. (2008): the basic assumption for applying non-linear, force-free magnetic field extrapolations is the force-freeness of the photospheric vector magnetograms. Measured magnetograms do not fulfill this requirement, therefore a preprocessing of the

measured data is required. Wiegelmann et al. (2008) developed a minimization procedure that yields a more chromosphere-like field by including the field direction information contained in, e.g., chromospheric Hα images. Including this information into the extrapolation algorithm significantly enhances the reliability of the extrapolations.

3.2 Direct Measurements of the Chromospheric Magnetic Fields

Measurement techniques for chromospheric magnetic fields have to overcome a variety of hurdles: (i) the plasma density is several orders of magnitude lower than in the photosphere, (ii) the energy transport is dominated by radiation, (iii) the magnetic field strength is on average lower than in the photosphere, and (iv) anisotropic illumination induces population imbalances between atomic sublevels that are modified by weak magnetic fields. The low plasma density leads to weak signals in the absorption (on-disk observations) or emission (off-limb observations) of spectral lines. The absorption signatures of chromospheric lines often show a strong photospheric contribution. Only highly spectrally resolved observations of the line core carry the chromospheric information. As a consequence of the low density, the simplifying assumption of local thermodynamic equilibrium breaks down. The interpretation of the observations is thus by far more involved than in the case of photospheric observations. Additionally, the low chromospheric magnetic field strengths weakens the Zeeman signals in spectral lines. Scattering polarization and its modification by the Hanle effect introduce an additional complication in the analysis of the polarization signal of spectral lines.

During the last decade major progress has been achieved in circumventing these hurdles. Radio observations are able to determine the magnetic field strength in and around active regions (see review by Lee 2007). Acoustic mapping techniques (Finsterle et al. 2004) use the reflection of high-frequency acoustic waves (mHz-range) from the region in the atmosphere where the gas pressure and the magnetic pressure are equal to reveal the structure of the magnetic canopy. The biggest leap in the direct determination of chromospheric magnetic fields was achieved by combining state of the art instrumentation for full Stokes polarimetry with recent progress in atomic physics. Bommier (1980), Landi Degl'Innocenti (1982), Stenflo and Keller (1997) and Trujillo Bueno et al. (2002) opened a new diagnostic window in solar physics: magnetic fields influence the strength and the direction of the linear polarization resulting from atomic or scattering polarization. This effect, discovered by Hanle (1991), allows the determination of the magnetic vector from Milligauss to several tens of Gauss, a range not accessible by Zeeman diagnostics.

The following sections describe examples of measurements in this new diagnostic window, focused around two of the most popular spectral lines for combined Hanle and Zeeman measurements: the triplet of He I 10 830 Å and the He I D_3 5876 Å multiplet. The formation of these lines requires ionization of para-He by ultraviolet radiation or collisions, followed by recombination to populate the lower sates of ortho-He. Since the main source for the ultraviolet radiation is the corona, these He I lines lacks almost any photospheric contribution. Additionally, they are (generally) optically thin and narrow, allowing the use of rather simple analysis techniques, like Milne-Eddington inversions of the radiative transfer equations (Solanki et al. 2003; Lagg et al. 2004, 2007). With the inversion code HAZEL (HAnle and ZEeman Light, see Asensio Ramos et al. 2008), involving the joint action of atomic level polarization and the Hanle and Zeeman effect in these lines, a standard tool for the analysis of Stokes spectra is now available.

3.2.1 Spicules

Spicules are an ubiquitous phenomenon on the Sun. At any time, the number of these needle-like structures on the Sun is on the order of 4×10^5. These dynamic and short lived features (lifetimes typically 5–10 minutes) can be considered as magnetic tunnels through which the refueling of the coronal plasma takes place (Athay 2000). High cadence Hinode SOT observations in Ca II H (Okamoto et al. 2007; Suematsu et al. 2007) revealed details in terms of size and dynamics and led to the discovery of a new type of spicules (type II spicules, De Pontieu et al. 2007a) with shorter lifetimes (10–150 s), smaller diameters (< 200 km compared to < 500 km for type I spicules), and shorter rise times. According to De Pontieu et al. (2007b), they (i) act as tracers for Alfvén waves with amplitudes of the order of 10 to 25 $\mathrm{km\,s^{-1}}$ and (ii) carry, in principle, enough energy to play an important role for the heating of the quiet Sun corona and for acceleration of the solar wind. See also Sect. 5.2.

Measurements of the magnetic field of spicules, both type I and type II, are essential for the understanding of this phenomenon. (Trujillo Bueno et al. 2005) were the first to directly demonstrate the existence of magnetized, spicular material. Full Stokes polarimetric data in the He I 10 830 Å line, obtained with the Tenerife Infrared Polarimeter (Collados et al. 1999), were analyzed by solving the radiative transfer equation assuming an optically thick atmosphere. The application of a combined Hanle and Zeeman diagnostic revealed a magnetic field strength for the observed type I spicule of 10 G and an inclination angle of 37° at a height of 2000 km above the photosphere. The authors state that 10 G is the typical field strengths for spicules at this height, but significantly stronger fields may also be present. This result agrees with the measurements from López Ariste and Casini (2005) using full Stokes polarimetry in the He I D_3 line. They find field strengths not higher than 40 G and a good correlation between the magnetic field orientation and the visible structure in Hα (see Fig. 9). An independent confirmation of these measurements was presented by Socas-Navarro and Elmore (2005) by using full Stokes observations from SPINOR (Spectro-Polarimeter for INfrared and Optical Regions, at the Dunn Solar Telescope). Their multi-line approach removes the dependence of the strength of the Hanle signals on the zero-field polarization produced by the scattering of anisotropic radiation in the higher atmosphere.

3.2.2 Prominences and Filaments

The spectacular eruptions of prominences and filaments and the resulting coronal mass ejections (CME) can cause sudden changes in the terrestrial magnetosphere. A typical CME releases an energy of 10^{25} J and 10^{12} kg of solar material into the interplanetary space (Harrison 1994). Before eruption, solar magnetic field holds this dense and relatively cool material in the hot coronal environment and supports it against the solar gravity for time periods as long as weeks. The knowledge of the magnetic field within these structures therefore is of great interest to understand the mechanisms leading to a possible eruption.

Casini et al. (2003) were the first to present magnetic maps of prominences using full Stokes polarimetry in the He I D_3 line. Their results confirm previous measurements of the average field in prominences, ranging between 10 and 20 G and oriented horizontally with respect to the solar surface. However, they also find the presence of organized structures in the prominence plasma embedded in magnetic fields that are significantly larger than average (50 G and higher). Merenda et al. (2007) extended this work to include the forward scattering case, applied it to a filament located at disk center and obtained the first magnetic maps of a filament. In this preliminary work they restricted their analysis to the saturated

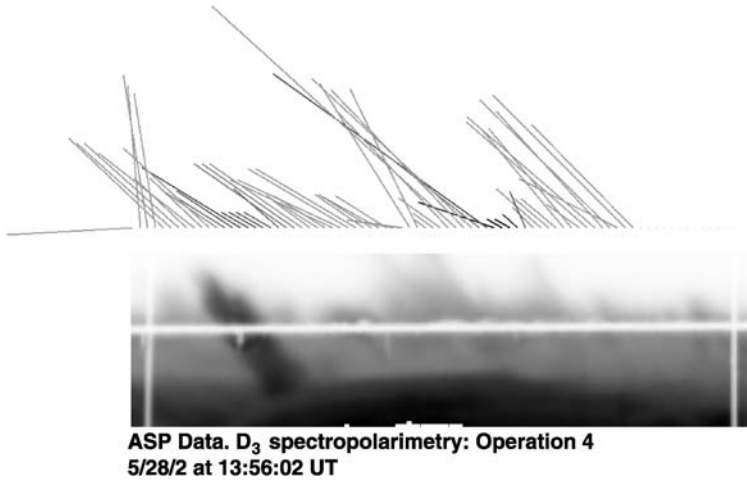


Fig. 9 Magnetic field measurement in the He I D_3 line: the magnetic field vector determined by a combined Hanle and Zeeman diagnostic traces the visible structures in the Hα slit-jaw image (*bottom*, adapted from López Ariste and Casini 2005)

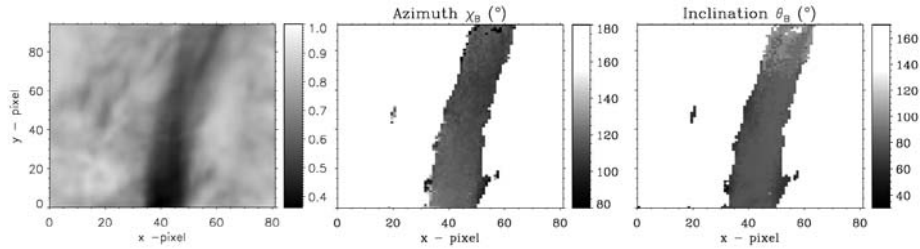


Fig. 10 Intensity image in the center of the He I 10 830 Å line and derived azimuthal and inclination angle of the magnetic field (adapted from Merenda et al. 2007)

Hanle effect regime between 10 and 100 G. Here the linear polarization is only sensitive to the direction of the magnetic field and does not change with intensity variations. The results are reliable maps for the azimuth and inclination angle for the magnetic field (see Fig. 10). In agreement with Casini et al. (2003) they find horizontal fields in the central part of the filament and a change of the azimuth according to the orientation of the main axis of the filament. In order to detect, for example, the small-scale and rapidly moving filaments mentioned in Sects. 4.2 and 5.3, significant improvements in signal to noise ratio and temporal resolution of polarimetric observations in this spectral line are required. The complex magnetic field and velocity structure of an erupting filament in the He I 10 830 Å line was analyzed by Sasso et al. (2007): besides the magnetic field topology they identify the presence of up to 5 different atmospheric components, distinguished by their velocities ranging from -50 to $+100\,\mathrm{km\,s^{-1}}$, within the resolution element of approximately 1.5″. This measurement clearly demonstrates the fibrilar structure of the chromosphere (see also Lagg et al. 2007) and the need for higher spatial resolution measurements in this line.

3.2.3 Canopy

Following previous work by W. Livingston, Gabriel (1976) introduced the term canopy to explain the emission measures of chromospheric and transition region UV lines. In the "classical" picture, the magnetic pressure wins over the gas pressure with increasing height, so that the magnetic flux concentrations rooted in the network expand and cover the internetwork cells with horizontal fields (see Sect. 5 for an updated view). Giovanelli and Jones (1982) and Jones and Giovanelli (1982) performed detailed studies of the magnetic canopy close to the limb by determining magnetograms using chromospheric spectral lines like the Ca II triplet at 8542 Å or the Mg I b_2 line at 5173 Å. These magnetograms are characterized by a polarity inversion line parallel to the limb, on either side surrounded by diffuse fields above the internetwork region (see Steiner and Murdin 2000 for a sketch of the magnetic configuration).

Especially during the last decade diagnostic tools involving the Hanle effect significantly improved the possibilities to characterize the canopy fields. Using spectropolarimetric data in the Sr II 4078 Å line "Hanle histograms", showing the statistical distributions of the Hanle rotation and depolarization effects, Bianda et al. (1998) determined the magnetic field strength of horizontal, canopy-like fields to be in the range of 5 to 10 G. The first spatial mapping of Hanle and Zeeman (Stenflo et al. 2002) effect revealed details of canopy fields in a semi-quiet region measured close to the limb in the Na I D_1–D_2 system. The authors found direct evidence for horizontal magnetic fields, slightly stronger than the field strengths determined by Bianda et al. (1998) (25–35 G), that remain coherent over a spatial scale of at least three supergranules.

The concept of a magnetic canopy around sunspots and in active regions is well established. Over quiet regions, the formation of this layer of horizontal fields is matter of debate: Schrijver and Title (2003) showed that concentrations of magnetic flux in the network in the order of a few tens of Mx cm^{-2} will destroy the classical, wineglass-shaped magnetic field topology. Such flux concentrations, suggested by simulations, were identified by Trujillo Bueno et al. (2004) in terms of ubiquitous tangled magnetic field with an average strength of $\approx$130 G, much stronger in the intergranular regions of solar surface convection than in the granular regions. A significant fraction of this hidden magnetic flux has now been clearly identified with the spectropolarimeter of the Hinode spacecraft (Lites et al. 2008). However, narrow-band (0.1 Å) observations in the Ca K line with a spatial resolution of 0.1″ obtained with the Swedish Solar Telescope (SST) provide evidence that magnetic fibrils, originating from network flux concentrations, do span over a large distance above the quiet Sun network (see Figs. 7 and 11). Magnetic field measurements using the He I 10 830 Å line also indicate the presence of a uniform, horizontal magnetic field topology over the internetwork at mesogranular scales (Lagg and Merenda 2008). These measurements, presented in Fig. 12, were obtained with the Tenerife Infrared Polarimeter II (TIP-2) mounted behind the Vacuum Tower Telescope (VTT) on Tenerife (Collados et al. 2007) at a heliocentric angle of 49° ($\mu = \cos\Theta = 0.65$). Inversions involving the Hanle and Zeeman effects prove the presence of a horizontal "canopy" magnetic field on mesogranular scales with strengths of the order of 50 to 100 G, similar to the value of the averaged magnetic field of the underlying photosphere. Both, the recent narrow-band Ca K observations of (e.g., Pietarila et al. 2008; Rouppe van der Voort et al. 2005) and the magnetic field measurements (e.g., Lagg and Merenda 2008) seem to be in *apparent* contradiction to Schrijver and Title (2003), pointing out the necessity of a more detailed analysis on the validity of the concept of the magnetic canopy over quiet Sun regions. Nevertheless, the different finding can be fit into a common picture, when taking into account the sampled height ranges and a field topology, which is

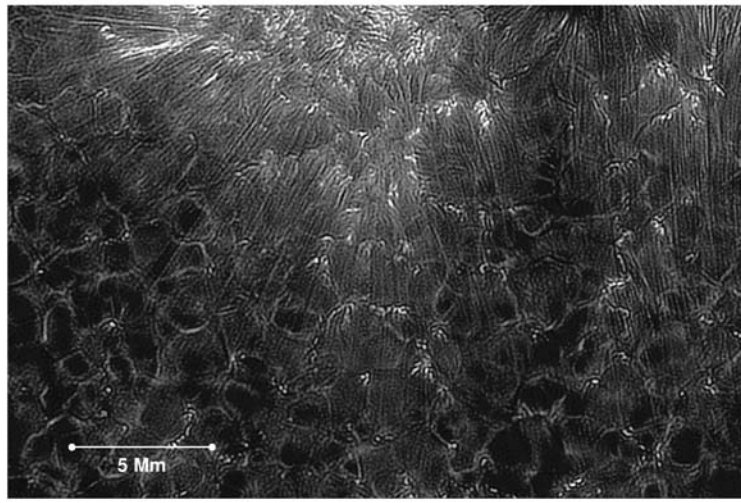


Fig. 11 Speckle-reconstructed, narrow band image (contrast-enhanced) of a plage region observed in the line core of Ca K using the SST (Pietarila et al. 2008, cf. Rouppe van der Voort et al. 2005). The Ca K fibrils extend over quiet Sun regions. The mesh-like background pattern is nevertheless dominated by the reversed granulation pattern in the middle photosphere (cf. Fig. 7)

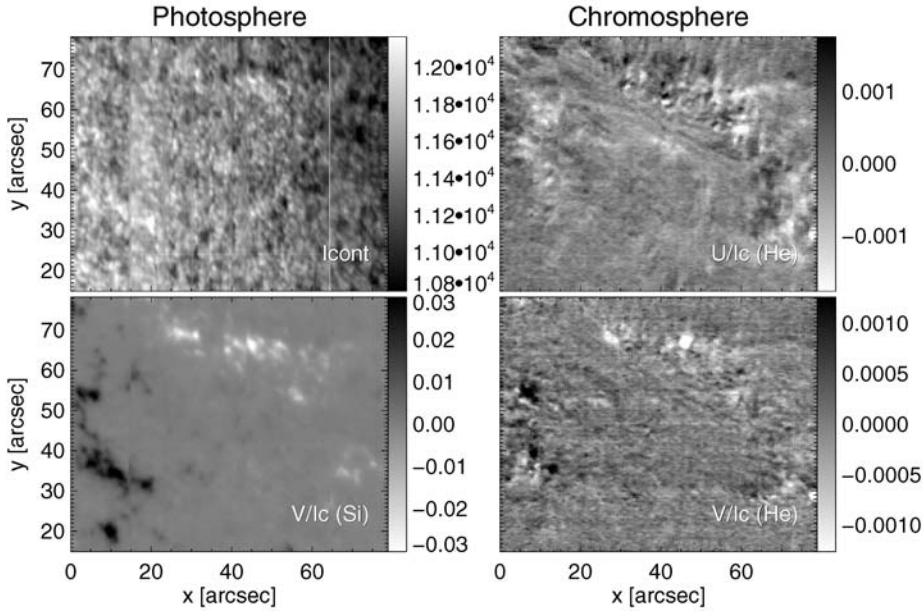


Fig. 12 Measurement of the magnetic field over a supergranular cell in the photosphere and the chromosphere (German Vacuum Tower Telescope, Tenerife Infrared Polarimeter 2, May 10, 2008): continuum close to the 10 830 Å line (*top left*), Stokes V signal integrated over the red wing of the photospheric Si I 10 827 Å line (*bottom left*), Stokes U and V signal integrated over red wing of the chromospheric He I 10 830 Å line (*top* and *bottom right*, respectively). The chromospheric maps suggest the presence of magnetic structures organized on mesogranular scales within the supergranular cell outlined by the photospheric Stokes V map

more complex and entangled on small scales than usually assumed (see Sect. 5). The "classical" canopy might be in some ways a too simplified and thus potentially misleading concept.

4 Numerical Simulations of the Quiet Sun

4.1 Internetwork Photosphere

The solar granulation is now well reproduced by modern radiation (magneto-)hydrodynamical simulations. The contrast of continuum intensity or "granulation contrast" is often used for comparisons between observations and simulations. For many years, the contrast derived from observations were much lower than those found in numerical simulations. One reason is the often unknown but crucial effect of an optical instrument and the Earth atmosphere, resulting in a significant decrease of the granulation contrast. This problem can partially be overcome by using observations with space-borne instruments. Recent observations with the Broadband Filter Imager (BFI) of the Solar Optical Telescope (SOT) onboard the Hinode spacecraft now show higher contrast values. After application of a realistic point spread function (Wedemeyer-Böhm 2008; Danilovic et al. 2008), state-of-the-art numerical simulations indeed reproduce important characteristics of "regular" granulation.

The convective flows in and just above granule interiors advect magnetic field laterally towards the intergranular lanes, where the field is concentrated in knots and sheets with up to kilo-Gauss field strengths. In the granule interiors, usually only weak field remains, although in some situations flux concentrations of up to a few hundred Gauss can occur within the granules (Steiner et al. 2008). The latter finding is in agreement with the observations by Centeno et al. (2007) and Ishikawa et al. (2008). This process of "flux expulsion" has been known since early simulations (Galloway and Weiss 1981; Nordlund 1986). It is now an integral part of magnetoconvection simulations (see, e.g.,Weiss et al. 1996; Stein and Nordlund 1998; Steiner et al. 1998; Schaffenberger et al. 2005; Vögler et al. 2005). The close-up from a simulation by Schaffenberger et al. (2005) in Fig. 13 illustrates the process. The magnetic field in the low photosphere is not only advected laterally. It is also lifted upwards and is concentrated above the reversed granulation layer at a height, which roughly corresponds to the classical temperature minimum in semi-empirical models (Fontenla et al. 1993). There, the convective overshooting effectively dies out and most of the upward directed flows above the granule interiors turn into lateral flows (Fig. 13). In the models by Wedemeyer et al. (2004), Wedemeyer (2003), the rms velocity amplitudes are smallest at these heights. The result is that the magnetic field is "parked" there and forms a mostly horizontally aligned field. It connects to the photospheric flux funnels, which spread out from the intergranular lanes below. The enclosed regions below, on the other hand, are virtually field-free with field strengths of possibly down to a few Gauss only (Steiner 2003). The field configuration around these granular voids was referred to as a dynamic "small-scale canopy" by Schaffenberger et al. (2005, 2006). Virtually field-free granule interiors are very common in their simulations. In the more recent simulations by Steiner et al. (2008), this phenomenon is also existent but less pronounced, although the field in the granular interiors is still much weaker than in the surrounding lanes. The main differences between these simulations is the average field strength (10 G and 20 G, resp.) and the injection of horizontal field at the lower boundary in the latter simulation. Obviously, the exact occurrence of small-scale canopies still depends on details of the simulations and thus needs to be checked by comparison with observations. The recent detection of so-called "horizontal inter-network fields" (HIFs) can be regarded as observational support for the small-scale field structure

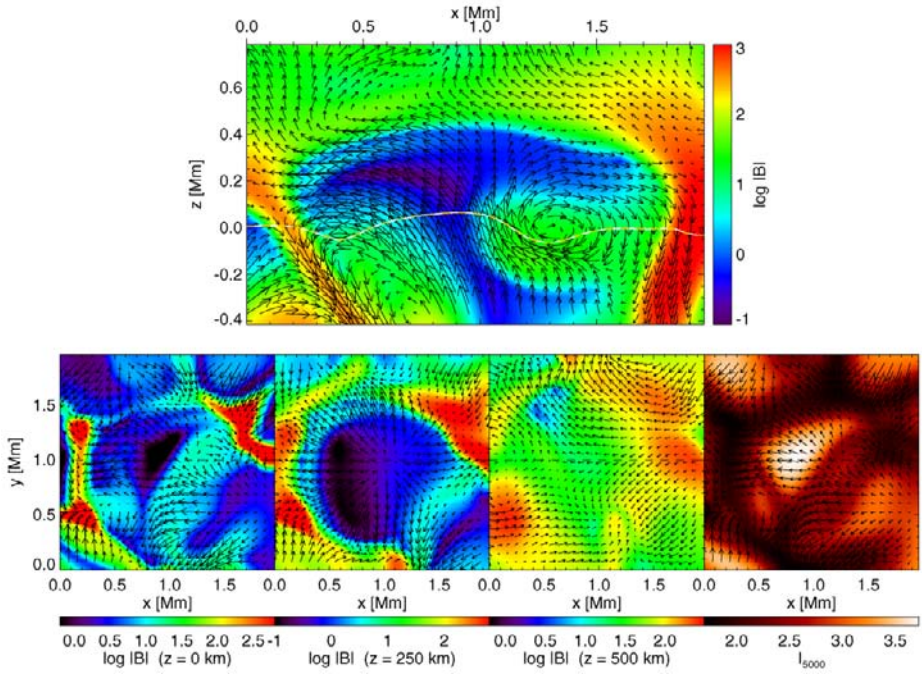


Fig. 13 Flux expulsion in a close-up from a MHD simulations by Schaffenberger et al. (2005): Logarithmic magnetic field strength in a vertical cross-section (*top*) and in three horizontal cross-sections (*bottom*) at heights of 0 km, 250 km, and 500 km. The emergent intensity is displayed in the *rightmost panel*. The *arrows* represent the velocity field in the shown projection planes. The *white line* in the *upper panel* marks the height of optical depth unity

seen in the simulations. It is observed that the horizontal field component in the granular interiors is stronger than the vertical component (Lites et al. 1996, 2007, 2008; Orozco Suárez et al. 2007). HIFs are also clearly present in simulations (Schaffenberger et al. 2006; Schüssler and Vögler 2008; Steiner et al. 2008) and are in good agreement with the observations.

The direction of the horizontal magnetic field, which is continuously lifted to the upper photosphere and lower chromosphere, varies. Consequently, current sheets form where different field directions come close to each other. In the simulations by Schaffenberger et al. (2006), a complex stacked meshwork of current sheets is generated at heights from ~ 400 km to ~ 900 km. The lower limit of this range, which is the typical height of the small-scale canopies can be considered as the upper boundary of the photosphere.

4.2 Internetwork Chromosphere

In recent years models have been extended in height to include the chromosphere. Modeling this layer is an intricate problem as many simplifying assumptions, which work fine for the lower layers, are not valid for the thinner chromosphere. Rather, time-dependent three-dimensional non-equilibrium modeling is mandatory. This is in particular true for the radiative transfer, for which deviations from the (local) thermodynamic equilibrium should be taken into account. Numerically, this is a demanding task. It is unavoidable to

make simplifications and compromises when implementing at least the most important non-equilibrium effects in a time-dependent multi-dimensional simulation code. A practicable way is to start with simplified models and increase the amount and the accuracy of physical ingredients step by step. In their pioneering work, Carlsson and Stein (1994, 1995) implemented a detailed radiative transfer, which was affordable by restricting the simulation to one spatial dimension. This simplification made it necessary to implement an artificial piston below the photosphere to excite waves as the convection cannot be realistically simulated in one spatial dimension. The high computational costs for such detailed radiative transfer calculations forced Skartlien et al. (2000) to use a simplified description for their three-dimensional model. Nevertheless, their treatment included scattering. Simplifications of the radiative transfer are necessary for three-dimensional simulations in order to make them computationally feasible. This class of 2D/3D numerical simulations cover a small part of the near-surface layers and extent vertically from the upper convection zone to the middle chromosphere. This way the shock-waves are excited by the simulated convection without any need for an artificial driver. The chromospheric layer of these models is usually characterized by intense shock wave action, putting high demands on the stability of numerical codes. Wedemeyer et al. (2004) made experiments with simplified 3D models without magnetic fields, using CO^{5}BOLD (Freytag et al. 2002). As in the aforementioned simulations, they found that overshooting convection in the photosphere triggers acoustic waves that propagate upwards and steepen into shock fronts. The result is a dynamic layer above a height of ~ 700 km, which is composed of hot shock fronts and cool post-shock regions. The gas temperature in horizontal cross-sections through the model exhibits highly dynamic mesh-like pattern with spatial scales comparable to the granulation. The same can be seen in the follow-up simulations by Schaffenberger et al. (2005), which include weak magnetic fields (see Fig. 14). The gas temperature in the CO^{5}BOLD model chromospheres range from about 7000 K down to 2000 K, owing to the adiabatic expansion of the post-shock regions. A similar pattern is also present in the simulations by Martínez-Sykora et al. (2008). The temperature range is very similar in both models, but the temperature amplitudes differ. Some snapshots of the simulation by Martínez-Sykora et al. (2008) also show a double-peaked temperature distribution at chromospheric heights, but the cool background component is usually much weaker than in the CO^{5}BOLD model. Possible reasons for the differences are related to the numerical treatment of the radiative transfer in the upper layers. A shock-induced pattern can already be perceived in the temperature maps by Skartlien et al. (2000), although it less pronounced due to the relatively coarse grid spacing in this earlier simulation.

Not only the modeling but also the observation of the shock-dominated layer (hereafter referred to as "fluctosphere", see Sect. 5) is non-trivial. A clear detection in Ca II H, K or the infrared lines requires a high spatial, temporal, and spectral resolution, all at the same time. A too broad filter wavelength range leads to significant contributions from the photosphere below. The fluctospheric pattern is then easily masked by a reversed granulation signal. The situation is complicated by the fact that both patterns have very similar spatial scales, i.e. roughly granulation scales. This is due to the fact that the generation of both patterns is due to processes in the low photosphere. This is illustrated in Fig. 15, which shows preliminary synthetic intensity maps in the Ca II infrared line at $\lambda = 854$ nm. The maps were calculated with the non-LTE radiative transfer code MULTI (Carlsson 1986) column by column from the model by Leenaarts and Wedemeyer-Böhm (2006). We use the non-equilibrium electron densities, which are output from the time-dependent simulation. The top rightmost panel of Fig. 15 shows the mesh-like pattern in the line core, whereas the reversed granulation is visible in the line wing (middle column). Even further out in the wing, the granulation

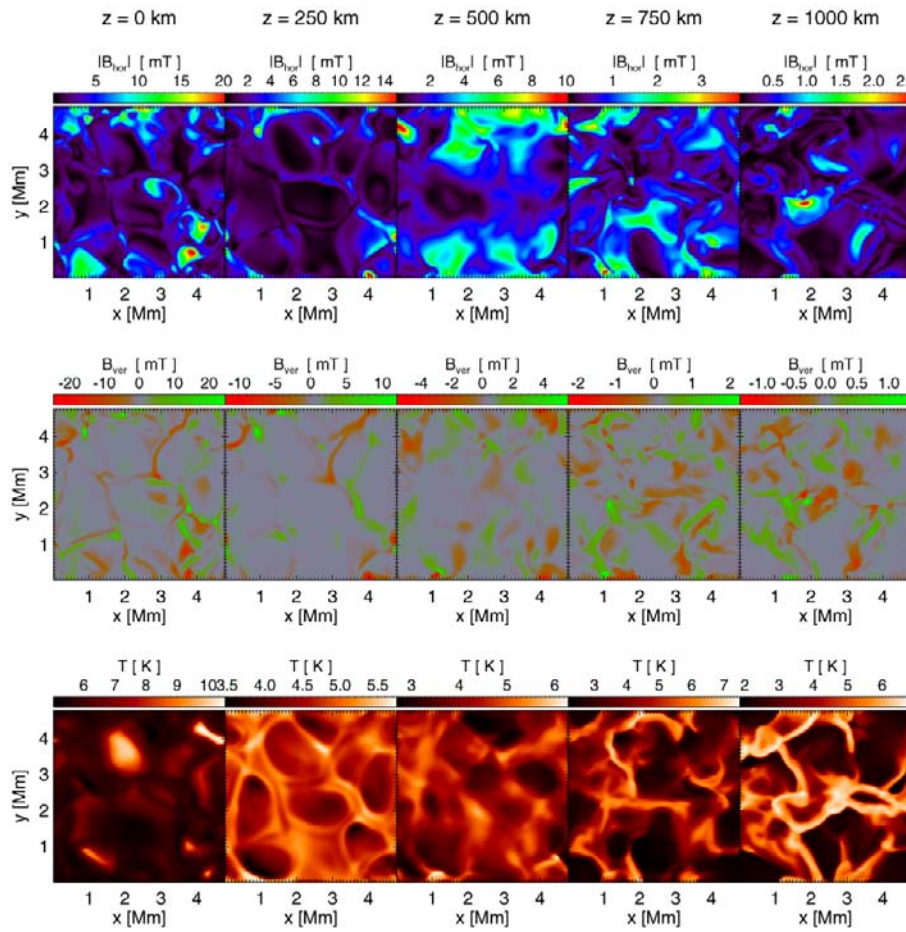


Fig. 14 Horizontal cross-sections through the model by Schaffenberger et al. (2005, 2006) showing the horizontal magnetic field component (*top*), the vertical component (*middle row*), and the gas temperature (*bottom*) at different heights: $z = 0$ km (granulation), 250 km (reversed granulation), 500 km, 750 km (fluctosphere), and 1000 km (from left to right)

pattern appears (left column). The mesh-like fluctosphere pattern can be seen Ca H, K, and the IR triplet, too.

A comparison of the line core map with the temperature maps in Fig. 14 shows that primarily the hottest regions of the pattern are seen in the Ca intensity, whereas a lot of atmospheric fine-structure remains invisible. The hot regions are caused by "collision" of neighboring shock fronts, ultimately compressing the gas in the region in-between and rising its temperature. This effect enhances in particular the Ca brightness at the vertices of the mesh. These small bright areas most likely are observed as Ca grains, while the emission along the mesh is so faint that it is hard to detect. The formation of Ca II grains by propagating shock waves was already explained by Carlsson and Stein (1997) over a decade ago. The fact that their detailed 1D simulations closely match observations of grains, clearly shows that Ca II grains are indeed a phenomenon related to shock waves. In 1D but also in 3D, the formation takes place at heights of ~ 1 Mm above optical depth unity. In both

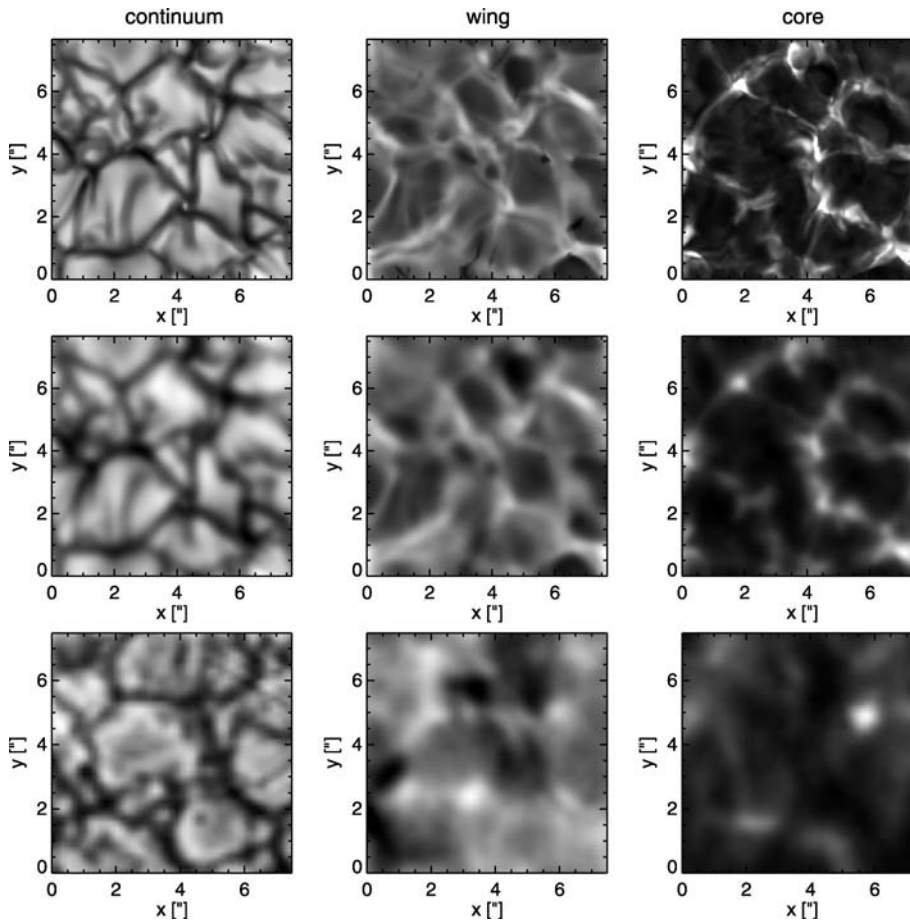


Fig. 15 Small-scale structure of the solar atmosphere seen in the Ca II infrared line at 854 nm continuum (*left column*), line wing (*middle*), and line core (*right column*). *Top row*: Synthetic maps based on a simulation with non-equilibrium hydrogen ionization; *middle row*: after application of a PSF and filter transmission; bottom: observations with IBIS at the DST (Courtesy of F. Wöger). See text for details

cases, the shocks propagate upwards into down-flowing material. The difference, however, is that in 1D shocks are plane-parallel so that interaction between individual waves is essentially reduced to shock-merging and shock-overtaking. In 3D, shock wave interaction is more complex. And still, the compression zones between shocks—the most likely candidate for grain formation in 3D—moves upwards with the waves and thus certainly show very similar observational signatures. While it seems to be well established that Ca grains are produced by shock waves, some details of the formation process have to be revisited in a 3D context.

However, the grains might just be the "tip of the iceberg". Progress in observational techniques and instrumentation now finally allow us to detect the dark details of the fluctosphere. The middle row of Fig. 15 illustrates this observational effect. A point spread function (PSF) has been applied to the synthetic maps. The PSF accounts for a circular, unobstructed aperture of 70 cm diameter and a non-ideal Voigt-like contribution due to instrumental stray-light and atmospheric seeing. Finally, the degraded maps are integrated

over wavelength with a synthetic transmission filter with a FWHM of 5 pm. The assumptions are rather optimistic and represent excellent observational conditions. And yet the resulting image degradation has a significant effect on the visible patterns. Obviously, a lower spatial or spectral resolution would further suppress the faint mesh-like pattern in the line core. Please note that the calculations are still preliminary. A full 3D treatment of the radiative transfer and the included scattering, which will soon be possible, might increase the area of enhanced brightness. Also it is not clear yet how the possible interaction of the shock waves with the overlying "canopy" field would alter the properties of the pattern and its observational mesh/grain signature. The resulting pattern nevertheless in many aspects resembles the recent observations by F. Wöger et al. with (i) the Interferometric BIdimensional Spectrometer (Cauzzi et al. 2008, IBIS) at the Dunn Solar Telescope (DST) of the National Solar Observatory at Sacramento Peak (Wedemeyer-Böhm and Wöger 2008) and (ii) with the German Vacuum Tower Telescope (VTT) at the Observatorio del Teide (Wöger et al. 2006). See the lower row of Fig. 15 for examples of IBIS data.

Based on the models by Wedemeyer et al. (2004), weak magnetic fields were taken into account in the simulations by Schaffenberger et al. (2005, 2006) and Steiner et al. (2008) (see Fig. 14). Different initial magnetic field configurations and strengths from $B_0 = 10$ G to 20 G were tried, all resembling quiet Sun internetwork conditions (see Wedemeyer-Böhm et al. 2005b for an experiment with $B_0 = 100$ G). The computational domains again comprise several granules and extend into the chromosphere, typically to heights of ~ 1400 km. The MHD models are very similar to their hydrodynamic precursors with respect to structure and dynamics. The ubiquitous shock waves produce a very similar pattern in the gas temperature but also shape the small-scale structure of the magnetic field in the upper model atmosphere. Consequently, the magnetic field in the fluctosphere is highly dynamic and has a complex topology. A look at horizontal cross-sections at different heights in Fig. 14 implies that the field in the upper layers is much weaker ($|B| < 50$ G) and more homogenous than in the photosphere below. On the other hand, the fluctospheric field evolves much faster. The horizontal field component $B_{\rm hor}$ in the range 500 km to 750 km is (i) stronger than the vertical one, $B_{\rm z}$ and (ii) has a rather large filling factor there.

In the small-scale internetwork simulations carried out with CO^{5}BOLD, the strongly varying surface of plasma $\beta = 1$ is found on average at heights of the order of 1000 km to 1400 km or even higher, depending on model details. Heights of the same order are also found by, e.g., Hansteen (2007). The exact location certainly depends on the field strengths in the internetwork, which are still under debate. Instead of plasma $\beta = 1$, one can also talk about an equivalent surface, where sound speed and Alfvén speed are equal. It makes clear that these regions are important for the propagation and eventual dissipation. Simulations show that this surface indeed separates two domains that differ in their dynamical behavior: A slow evolving lower part and a highly dynamic upper part. This is certainly related to the finding that wave mode conversion and refraction occurs under the condition of plasma $\beta \approx 1$ (Rosenthal et al. 2002; Bogdan et al. 2003; Cally 2007; Steiner et al. 2007). The current sheets, which are present below and above the plasma $\beta = 1$ surface, differ in their orientation. While they are mostly stacked with horizontal orientation in the lower part down to the top of the small-scale canopies at the boundary to the photosphere, the thin current sheets above plasma $\beta = 1$ are formed along shock fronts and can thus show oblique or even vertical orientation.

4.3 Large-Scale Simulations

The models described in Sects. 4.1 and 4.2 do not take into account the large-scale canopy fields but rather concentrate on the small spatial scales of quiet Sun internetwork regions. In contrast, the simulations discussed in this section comprise larger computational domains. To make this possible, one usually has to make compromises such as, e.g., reduce the spatial resolution or develop efficient numerical methods. Stein et al. (2006) made impressive progress by extending the computational box towards supergranulation scales. Their models do not include the upper atmosphere but extend deep into the convection zone. Gudiksen and Nordlund (2002, 2005a), on the other hand, succeeded in creating time-dependent numerical models, which extend from the photosphere all the way into the corona. An important aspect, which can be investigated with this kind of models, it the (magnetic) connection between the atmospheric layers all the way from the top of the convection to the corona (see also Abbett 2007). Furthermore, extended simulations allow for investigating phenomena that are connected to spatial scales between granulation and supergranulation. For instance, the simulations by Hansteen and Gudiksen (2005) and Hansteen et al. (2006) revealed the formation of dynamic chromospheric features similar to dynamic fibrils. Being driven by upward propagating waves in the chromosphere, they are an example of the coupling between different atmospheric layers. Another type of coupling is provided in the form of horizontal magnetic flux structures with extensions of a few Mm, which emerge from the upper convection and rise upwards through the atmosphere. See Cheung et al. (2007) and Martínez-Sykora et al. (2008) for recent examples of flux emergence simulations. The simulations by Leenaarts et al. (2007) confirm once more (cf. Carlsson and Stein 2002) that the ionization degree of hydrogen has to be treated in non-equilibrium in the upper atmosphere. Although the simulation is two-dimensional, it features weak-field sub-canopy domains with upward propagating shock waves and a magnetic-field dominated "canopy" domain above (see their Fig. 1). A strong coupling of the individual layers is very obvious.

5 An Updated Picture of the Quiet Sun Atmosphere

The results of the previous sections are summarized in a schematic sketch of the quiet Sun atmosphere (see Fig. 16) with particular emphasis on the low atmosphere in internetwork regions. It is based on (and should be interpreted in comparison with) recent sketches by, e.g., Judge (2006), and Rutten (2006, 2007) but contains many modifications to incorporate new results derived from observations and numerical simulations.

5.1 The Large-Scale Magnetic Field

The large-scale building blocks of the quiet Sun atmosphere are the magnetic network patches, which outline supergranulation cells. The large-scale convective flows (see long arrows) advect magnetic field to the lanes of the supergranulation. Consequently, the magnetic field is highly structured and concentrated close to the "surface" ($\tau_{500} = 1$) with kG field strengths. The visible result is the so-called magnetic network (see Fig. 7. More recent observations with high spatial resolution (e.g., Orozco Suárez et al. 2007) reveal that the magnetic network patches consist of a conglomerate of smaller magnetic elements or "flux bundles" of different field strength with a wealth of substructure. This finding is incorporated in Fig. 16, in contrast to earlier sketches that feature the magnetic network as rather massive flux tubes. The heights where sound speed and Alfvén are equal ($c_s = c_A$), or equivalently where the

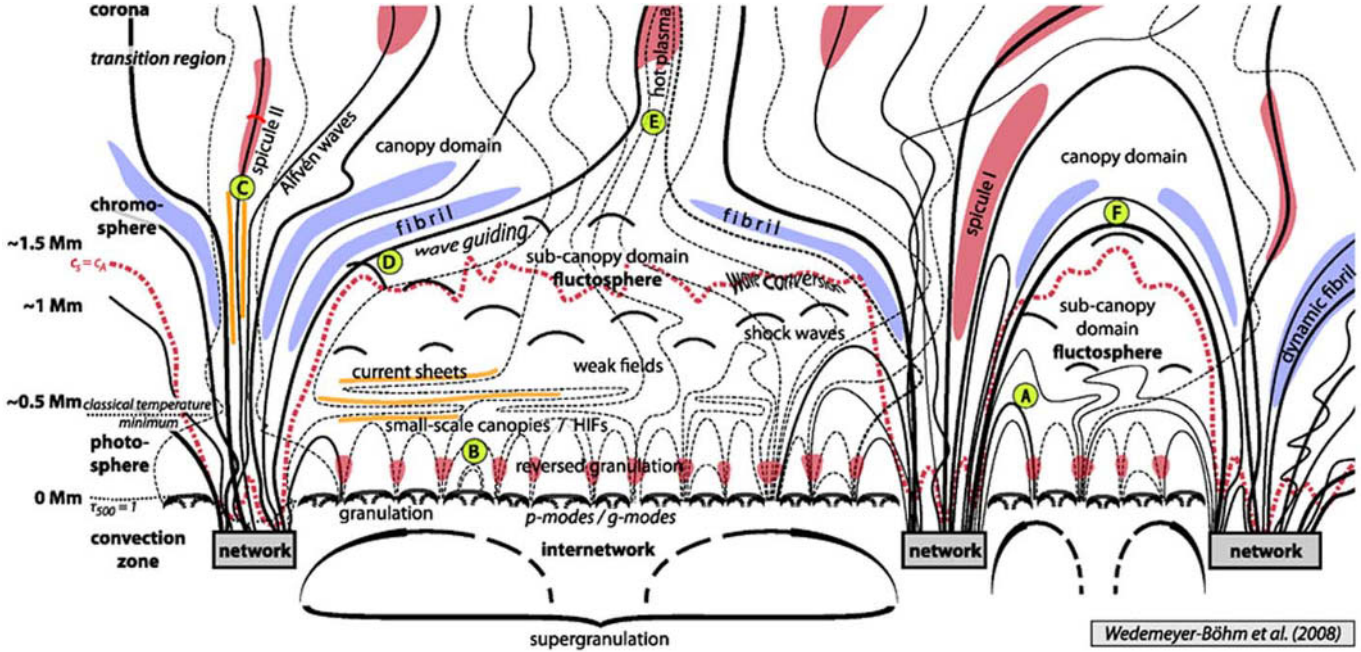


Fig. 16 Schematic, simplified structure of the lower quiet Sun atmosphere (dimensions not to scale): The *solid lines* represent magnetic field lines that form the magnetic network in the lower layers and a large-scale ("canopy") field above the internetwork regions, which "separates" the atmosphere in a canopy domain and a sub-canopy domain. The network is found in the lanes of the supergranulation, which is due to large-scale convective flows (*large arrows* at the *bottom*). Field lines with footpoints in the internetwork are plotted as *thin dashed lines*. The flows on smaller spatial scales (*small arrows*) produce the granulation at the bottom of the photosphere ($z = 0$ km) and, in connection with convective overshooting, the weak-field "small-scale canopies". Another result is the formation of the reversed granulation pattern in the middle photosphere (*red areas*). The mostly weak field in the internetwork can emerge as small magnetic loops, even within a granule (point B). It furthermore partially connects to the magnetic field of the upper layers in a complex manner. Upward propagating and interacting shock waves (*arches*), which are excited in the layers below the classical temperature minimum, build up the "fluctosphere" in the internetwork sub-canopy domain. The *red dot-dashed* line marks a hypothetical surface, where sound and Alfvén are equal. The labels D-F indicate special situations of wave-canopy interaction, while location D is relevant for the generation of type-II spicules (see text for details). Please note that, in reality, the 3D magnetic field structure in the canopy and also in the subcanopy is certainly more complex and entangled than shown in this schematic sketch

plasma $\beta = 1$, will certainly show large variations, depending on the (local) field strength. It may even reach below the surface of optical depth unity (at a reference wavelength of 500 nm) within strong field concentrations but may stay up at a few hundreds kilometers in weaker network patches. And still the variation in field strength and topology, incl. the width of the network patches, is even larger than can be presented in the simplified sketch here.

The magnetic field spreads out in the layers above the patches. Depending on the polarity of neighboring flux concentrations, they can form funnels or connect via loops that span the internetwork regions in-between. These two cases are illustrated in Fig. 16 in a simplified way. In the classical picture, the large-scale field enclosing the weak-field internetwork regions is referred to as "magnetic canopy". The corresponding flux funnels are often depicted with a wineglass-like geometry and have their footpoints in the photospheric network only. In reality, where the third spatial dimension offers an important additional degree of freedom, the field topology is more complex (see, e.g., Gudiksen 2006; Peter et al. 2006; Jendersie and Peter 2006; Schrijver and van Ballegooijen 2005). Schrijver and Title (2003) state that as much as half of the field could actually be "rooted" in the internetwork regions. From there, it can connect directly to the coronal field or via small loops to the photospheric network. The network patches could thus be surrounded by "collars" of loops with spatial scales comparable to one or a few granules. Consequently, the concept of a regular canopy structure seems questionable. Instead, the field topology should rather be understood as a set of individual field lines. Nevertheless, we stick here to the term "canopy" but use it in a wider sense. The height of the canopy and the field structure as a whole varies significantly from region to region and with time. The height indicators to the left in Fig. 16 should therefore only be used for rough orientation. In principle, the lower boundary of the "canopy" field separates two distinct domains: a *canopy domain* and a *subcanopy domain*. In reality, however, the boundary is certainly less strict than the sketch may imply. Rather, the magnetic field of both domains may be interconnected, e.g., by small loops, which extend on granular scales (point A). This way, the dynamics of the internetwork photosphere could have a direct influence on the properties of the upper layers, e.g., with respect to wave propagation and heating.

5.2 The Canopy Domain

The canopy domain is dominated by (large-scale) magnetic fields. It is this layer, which, due to the emission in Hα, appears as a purple-red rim at the beginning and end of a total solar eclipse. Therefore, only the canopy domain represents the chromosphere in a strict and original sense. At a closer look, a rich fibrilar structure can be seen in chromospheric Hα observations. They are found in rosette-like formations that funnel out from the magnetic network below and in many cases connect to neighboring network fields. A few fibrils are shown in Fig. 16 in connection with plasma that is trapped in the chromospheric field. Such fibrils and also the larger dynamic fibrils (Hansteen et al. 2006; Langangen et al. 2008a, 2008c shown at the right in the figure here) are an integral part of the quiet Sun chromosphere and even more frequent than can be shown in the 2D sketch here. According to De Pontieu et al. (2007c), fibrils could be the result of chromospheric shock waves that occur when convective flows and global oscillations leak into the chromosphere along the field lines of magnetic flux concentrations. In general, magnetohydrodynamic waves are an integral and ubiquitous part of the canopy domain. (Alfvén waves are indicated in Fig. 16 but represent just one of several possible wave modes.) Such perturbations can be excited by a number of processes, e.g., by the shuffling and braiding of the magnetic footpoints in the photosphere by convective flows. As the large-scale magnetic field continues from the

lower layers into the *transition region* into the *corona* above, the whole canopy domain is dynamically coupled. Again, it must be emphasized that the field topology is certainly more complex than can be expressed in the sketch here (see, e.g., Fig. 7). Indeed, Schrijver and van Ballegooijen (2005) state that instead of the plasma-β surface being closely connected to the (classical) canopy, regions with low and high β can well be mixed up into the corona.

As already mentioned in Sect. 3, a most obvious constituent of the chromosphere, at least when observed at the solar limb, are spicules (see, e.g., De Pontieu et al. 2004). Now, two types of spicules are distinguished based on differences in their dynamic behavior (De Pontieu et al. 2007a). Spicules of type I are the result of shock waves that are excited by disturbances in the photosphere (e.g., in connection with p-modes) and propagate from there along the magnetic field lines photosphere into the upper layers (Hansteen et al. 2006; Rouppe van der Voort et al. 2007). Spicules of type II, on the other hand, are more dynamic but thinner, exhibit higher velocities and have shorter lifetimes (see, e.g., Langangen et al. 2008b). They are most likely generated by magnetic reconnection events. Alfvén waves, which by many are considered as an ubiquitous phenomenon in the upper atmosphere, can be detected in connection with spicules (De Pontieu et al. 2007b). An example is drawn in the upper chromosphere above some vertically orientated current sheets (point C).

Another ingredient of the sketch are blobs of hot plasma in the corona, although their exact position and shape needs further investigation. De Pontieu et al. (2003) showed that the emission is not correlated with the centers of flux concentrations. Rather, the emission seems to appear at random locations. Although De Pontieu et al. (2003) refer to “moss” (Berger et al. 1999; Fletcher and De Pontieu 1999), which is related to active regions, there is no obvious reason why the situation should be different for the quiet Sun corona. Also, hot plasma regions like the one marked with “E” in the sketch are certainly not preferentially located directly above the middle of an internetwork region. In reality, the entangled and skewed field topology will make such blobs—if existent in the way depicted here—appear rather uncorrelated with the field topology of the underlying magnetic network.

5.3 The Sub-Canopy Domain

The magnetic field in the sub-canopy domain is mostly weak (see, e.g., Trujillo Bueno et al. 2004; Orozco Suárez et al. 2007), so that the plasma is larger than one in the lower layers. There, the field is essentially passively advected by the hydrodynamic flow fields. Convective motions and overshooting at the “surface” are the fundamental structuring agents, making the granulation the dominant spatial scale. Nevertheless, the weak fields in the sub-canopy domain most likely connect at least partially with the stronger canopy field. This feature is taken into account as integral part of the atmosphere sketch. Unfortunately, the presentation remains rather speculative at this point as many details of how and where the connections exactly take place are still unknown.

Beside the magnetic field, the consequences of convective overshooting allow to divide the subcanopy domain into layers with distinct dynamics (from bottom to top): low photosphere, middle photosphere, high photosphere, fluctosphere.

The Lower and Middle Photosphere exhibit the visible imprints of the solar surface convection. The granulation in the low photosphere is directly produced by small-scale convection cells (see, e.g., Nordlund and Dravins 1990), while the reversed granulation in the middle photosphere is a second-order effect. Gas is brought up by convective overshooting in the granule interiors, adiabatically expanding and cooling. It streams down again in the intergranular lanes, where it is compressed and heated. In addition, p-modes, i.e. global oscillations, and local acoustic events are important ingredients of the photospheric dynamics.

Recently, Straus et al. (2008) presented new support for the idea that gravity waves could play an important role, too.

The usually weak magnetic field is brought up from the convection zone below and/or possibly locally generated by small-scale dynamo action close to the surface. In the photosphere, the weak field is more or less passively advected towards the intergranular lanes but also towards the upper photosphere. The resulting field concentrations in the lanes become visible as very small and confined structures, e.g., in G-band images (see, e.g., de Wijn et al. 2008, for a recent example). In general, the internetwork field in the photosphere exhibits significant inclination and mixed polarity (see, e.g., Martínez González et al. 2008; Orozco Suárez et al. 2007). The granule interiors may become virtually field-free if there is no supply of magnetic fields with the warm convective upflows. Such voids are enclosed by small-scale canopies. Over most of the granulation, the horizontal field component is stronger than the vertical. This effect is observed as "horizontal internetwork fields" (HIFs). Magnetic field can emerge also in the form of small loops, which may have footpoints even within a granule (see point B in Fig. 16). This process, which was observed by Centeno et al. (2007), most likely adds to the accumulation of field above granules. In addition to emerging loops, Stein and Nordlund (2006) report on flux that is submerging and thus disappears from the surface.

The Upper Photosphere marks the boundary between the photosphere, which is controlled by the effects related to convective overshooting, and the wave-dominated layer above. This boundary can roughly be placed at the height of the classical temperature minimum. There, the temperature structure appears smoothed out and less structured than above and below; it is here that the average temperature amplitudes are smallest. It is roughly the height where the UV continuum at 160 nm is formed (cf. TRACE passbands). The upper photosphere is the layer, where the small-scale canopies have their top and where stacked (horizontal) current sheet become most obvious. This layer can be seen as a kind of (dynamical) insulation between the internetwork photosphere and fluctosphere. This effect becomes obvious in simulations when starting from an initial condition which feeds in field at the lower domain boundary. The photospheric field is built-up rather quickly but the field above only after a time delay because it only slowly spreads into the strongly subadiabatic stratification of the upper photosphere.

The Fluctosphere The shock-dominated domain in subcanopy internetwork regions (see Sect. 4.2), is referred to as "fluctosphere" by Wedemeyer-Böhm and Wöger (2008), while Rutten (2007) uses the term "clapotisphere". It is located between the photosphere and the part of the chromosphere visible in Hα. It is composed of propagating and interacting shock waves (with weak field only) and intermediate cool post-shock regions. Ideally, the wave fronts would expand spherically, while moving in vertical direction. In reality, they are deformed by running into an inhomogeneous medium of downflowing gas, which was shaped by precursory wave trains. The horizontal expansion of the fronts inevitably causes interaction between them. A visible result is the formation of Ca grains at heights, which traditionally would be assigned to the low chromosphere. The waves are excited in the photosphere below via different processes, which are related to convection (e.g., exploding granules), overshooting, and p-modes. The magnetic field in the fluctosphere is rather weak and is therefore mostly passively shuffled around by the shock waves. The result is a very dynamic and entangled field. The strongly varying surface of plasma $\beta = 1$ or in this context better $c_s = c_A$ is most likely located at heights of the order 1000 km to 1500 km or even higher (see Sect. 4.2). There, the conditions allow wave mode conversion, so the parts of the fluctosphere

below and above can show a somewhat different dynamical behavior. In the upper part, the weak fields become more important and rapidly moving filaments of enhanced field strength are generated. The propagating shock waves nevertheless remain the dominating structuring agent. A consequence, however, is that the current sheets are only stacked at plasma $\beta > 1$. Above, they are less regular as they are formed in the narrow collision zones of shocks, where the magnetic field is occasionally compressed. This shock-induced magnetic field compression might qualify as a (minor) heating process with potential consequences for the chromospheric energy balance.

The fluctosphere is not directly visible in Hα (in the line core at least) and is thus not a part of the chromosphere in a strict sense. It seems advisable to reserve the term chromosphere for the fibrilar canopy domain as visible in Hα (or in the very line cores of the Ca II lines). However, the fluctospheric shock waves could still leave an imprint in chromospheric diagnostics by interacting and penetrating the canopy field. On the other hand, the fluctosphere is also no part of the photosphere, although causally connected via the shock waves that propagate upwards from the low photosphere. The fluctosphere could be regarded as a second-order effect only, in contrast to the granulation and reversed granulation, which are direct consequences of the solar surface convection.

5.4 Shock Waves Meet the "Canopy"

Some details of Fig. 16 concern the interplay of propagating waves and the magnetic canopy. There are certain zones in these magnetic structures that act as mode conversion zone (Bogdan et al. 2003; Cally 2007), e.g., converting incoming acoustic waves into other modes, such as fast and slow magnetoacoustic waves. It is thus possible that such converted waves continue to propagate along the canopy field lines as some kind of "canopy waves". For simplicity, such a zone is marked by "wave conversion" at the "outer" boundary between canopy and subcanopy domain in the figure. Generally, such zones can be located everywhere in the structure where sound speed and Alfvén speed are of equal magnitude.

Furthermore, refraction and even reflection of waves can occur in such zones. As for the mode conversion, details depend on the relative orientation of the incoming wave and the magnetic field (Hasan et al. 2008). A wave can remain barely affected by the field when traveling perpendicular to the field lines, e.g., upwards in a vertical flux concentration. On the other hand, significant (relative) inclination can even result in total internal reflection for some wave modes (Rosenthal et al. 2002). In general, it can be assumed that the (acoustic) shock waves coming from the fluctosphere are guided by the magnetic canopy (e.g., point D in Fig. 16). Consequently, waves might follow the canopy field upwards and compress and heat the gas trapped between chromospheric "funnels" (point E). In closed loop regions, strong waves could push into the canopy from below and compress the magnetic field (location F). Depending on the local field configuration and the properties of the incoming wave, such an event could eventually trigger reconnection events. It could contribute to chromospheric heating. It certainly would not be only limited to the locations indicated in the sketch but occur more often in complex 3D field configurations. On the other hand, a regular closed structure as in the figure could possibly refract the waves from below such that they are "focussed" in the top of the subcanopy domain, amplifying their effect on the canopy field. A possible—although speculative—result could be the triggering of "nanoflares", although they are initiated by other mechanisms at other locations, too. That the upwards propagating waves interact with the canopy field is implied by observations in Hα. The dynamic behavior of the chromospheric fibrils is reminiscent of strings that sway back and forth in reaction to the quasi-continuous impact of

waves from below (point F). Under certain conditions, the shock waves might actively deform the field configuration of the magnetic canopy. Magnetoacoustic waves can already enter network flux concentrations in the photosphere, where the inclined magnetic field lines act as "magnetoacoustic portals" (Jefferies et al. 2006). The observations of so-called "acoustic shadows" provide observational evidence for the interaction of acoustic waves with the field around network footpoints (Krijger et al. 2001; McIntosh et al. 2001; McIntosh and Judge 2001).

5.5 Probing the Upper Atmosphere

With the currently available diagnostics for the chromosphere, observations of the sub-canopy domains are problematic. The Hα line core samples only the "canopy domain", whereas observations in the line wing reveal a background that most likely is dominated by the reversed granulation at much lower heights. It seems questionable if the layer in-between—the fluctosphere—can be observed in the Hα line wing at all in internetwork regions. Polarimetric measurements in the He 10 830 Å line (see Sect. 3) principally allow for the determination of the magnetic structure in a slab located between 1000 km and 2000 km. The formation of this line requires coronal illumination in the UV, resulting in complete absence of any photospheric contamination. However, the main contribution in the He 10 830 Å line comes from layers slightly above the fluctosphere. The Ca H & K and IR lines in principle would allow observations of the fluctosphere if very narrow filters are used. Otherwise, the detected intensity is "contaminated" with radiation from layers below. Very often, Ca observations with too broad filter prominently show the reversed granulation (see Figs. 7d and 11), which is easily mistaken as chromospheric signal. Very narrow filters, on the other hand, make it necessary to properly correct for Doppler shifts. A solution is fast scans through the Ca II IR lines with new imaging polarimeters such as IBIS (Cauzzi et al. 2008; Kleint et al. 2008) or CRISP (Scharmer et al. 2008), or spectro-polarimeters like SPINOR (Socas-Navarro et al. 2006). The extended formation height ranges and the non-equilibrium conditions, under which the inner parts of these lines are formed, complicate the interpretation and the derivation of the atmospheric structure. A promising alternative are the (sub-)millimeter continua, which will become accessible with the Atacama Large Millimeter Array (ALMA) a few years from now. Although technical details of this new type of observation render the construction of brightness temperature maps a certainly very complicated task, the scientific results could significantly contribute to our understanding of the solar atmosphere at chromospheric heights (Loukitcheva et al. 2008; Wedemeyer-Böhm et al. 2007).

6 Conclusions

The solar atmosphere is a very dynamic and inhomogeneous multi-scale system. Its individual components are coupled; some of them even show a kind of hierarchical self-similarity. Examples are the observational imprints of sub-surface convection, with a continuous spectrum of scales from below granulation scales to above supergranulation scales, and magnetic fields, which also exhibit similar features over a large range of spatial scales.

Despite great progress on the theoretical and observational sides, which go hand in hand, we are still missing an ultimate, comprehensive picture of the quiet Sun atmosphere. But at least we can now see what is needed for a corresponding numerical simulation. First, the computational domain should be large enough to encompass a few supergranulation cells

while the spatial resolution must still be high enough to capture important processes that occur on scales smaller than granulation. The vertical couplings make it necessary to consider an extensive height range. The corona and chromosphere can only be treated realistically when including the important driving motions in the layers below, i.e. in the photosphere and (at least) the upper part of the convection zone. While many simplifying assumptions can be made for the lower parts of such a model, the layers above the (middle) photosphere require a numerically complicated and thus computationally expensive non-equilibrium modeling approach, e.g., a realistic treatment of hydrogen ionization etc. The production of such a comprehensive model—and analogous models for, e.g., active regions—is thus very involved and can be regarded as one of the current challenges in (computational) solar physics.

On the observational side, we must continue to push forward the instrumental possibilities towards higher resolution in the spatial, temporal, and spectral domains—all at the same time. In addition, the (further) development and exploitation of advanced diagnostics is needed to derive a seamless tomography of the atmosphere as an integral phenomenon.

Until we succeed to reach these ambitious goals, we are left with a number of open questions. Of particular interest for the quiet Sun are, amongst others:

- How does the weak internetwork field connect with the stronger network field? What does the magnetic field look like just below the "canopy"? And can we talk about a "canopy" even in a wider sense after all?
- How do the propagating fluctospheric shock waves interact with the stronger ("canopy") field? Is it mostly a "passive" refraction/reflection at the "boundaries" of flux concentrations (e.g., Rosenthal et al. 2002; Steiner et al. 2007) or also "active" distortion/displacement/compression of magnetic field? How and where does mode conversion take exactly place under realistic conditions?
- The question of the coupling between the atmospheric layers is closely connected to the heating mechanism question or, better said, to the question of the atmospheric energy balance, not only for the Sun but also for other stars. Amongst other things, the ongoing controversy concerning the heating mechanism of the quiet Sun chromosphere (e.g., Fossum and Carlsson 2005) has important implications for stellar activity in general.

Acknowledgements We would like to thank O. Steiner, K. Schrijver, B. de Pontieu, M. Carlsson, V. Hansteen, R. Rutten, H. Peter, and Ø. Langangen for helpful discussions. SWB was supported with a Marie Curie Intra-European Fellowship of the European Commission (6th Framework Programme, FP6-2005-Mobility-5, Proposal No. 042049). Å.N. acknowledges support from the Danish Natural Science Research Council and from the Danish Center for Scientific Computing.

References

W.P. Abbett, The magnetic connection between the convection zone and corona in the quiet Sun. Astrophys. J. **665**, 1469–1488 (2007)

T. Amari, J.J. Aly, J.F. Luciani, T.Z. Boulmezaoud, Z. Mikic, Reconstructing the solar coronal magnetic field as a force-free magnetic field. Sol. Phys. **174**, 129–149 (1997)

A. Asensio Ramos, J. Trujillo Bueno, E. Landi Degl'Innocenti, Advanced forward modeling and inversion of stokes profiles resulting from the joint action of the Hanle and Zeeman effects. Astrophys. J. **683**, 542–565 (2008)

M. Asplund, Å. Nordlund, R. Trampedach, C. Allende Prieto, R.F. Stein, Line formation in solar granulation. I. Fe line shapes, shifts and asymmetries. Astron. Astrophys. **359**, 729–742 (2000a)

M. Asplund, H.-G. Ludwig, Å. Nordlund, R.F. Stein, The effects of numerical resolution on hydrodynamical surface convection simulations and spectral line formation. Astron. Astrophys. **359**, 669–681 (2000b)

R.G. Athay, Are spicules related to coronal heating? Sol. Phys. **197**, 31–42 (2000)

T.R. Ayres, Does the Sun have a full-time COmosphere? Astrophys. J. **575**, 1104–1115 (2002)

T.E. Berger, B. De Pontieu, L. Fletcher, C.J. Schrijver, T.D. Tarbell, A.M. Title, What is moss? Sol. Phys. **190**, 409–418 (1999)
M. Bianda, J.O. Stenflo, S.K. Solanki, Hanle diagnostics of solar magnetic fields: the SR II 4078 Angstrom line. Astron. Astrophys. **337**, 565–578 (1998)
T.J. Bogdan, M. Carlsson, V.H. Hansteen, A. McMurry, C.S. Rosenthal, M. Johnson, S. Petty-Powell, E.J. Zita, R.F. Stein, S.W. McIntosh, Å. Nordlund, Waves in the magnetized solar atmosphere. II. Waves from localized sources in magnetic flux concentrations. Astrophys. J. **599**, 626–660 (2003)
V. Bommier, Quantum theory of the Hanle effect. II—Effect of level-crossings and anti-level-crossings on the polarization of the D3 helium line of solar prominences. Astron. Astrophys. **87**, 109–120 (1980)
P.S. Cally, What to look for in the seismology of solar active regions. Astron. Nachr. **328**, 286 (2007)
M. Carlsson, A computer program for solving multi-level non-LTE radiative transfer problems in moving or static atmospheres. Report No. 33, Uppsala Astronomical Observatory
M. Carlsson, Modeling the Solar Chromosphere, in *The Physics of Chromospheric Plasmas*, ed. by P. Heinzel, I. Dorotovič, R.J. Rutten. Astronomical Society of the Pacific Conference Series, vol. 368 (ASP, San Francisco, 2007), p. 49
M. Carlsson, R.F. Stein, in *Proc. Mini-Workshop on Chromospheric Dynamics*, ed. by M. Carlsson (Inst. Theor. Astrophys., Oslo, 1994), p. 47
M. Carlsson, R.F. Stein, Does a nonmagnetic solar chromosphere exist? Astrophys. J. Lett. **440**, L29–L32 (1995)
M. Carlsson, R.F. Stein, Formation of solar calcium H and K bright grains. Astrophys. J. **481**, 500 (1997)
M. Carlsson, R.F. Stein, Dynamic hydrogen ionization. Astrophys. J. **572**, 626–635 (2002)
R. Casini, A. López Ariste, S. Tomczyk, B.W. Lites, Magnetic maps of prominences from full stokes analysis of the He I D3 line. Astrophys. J. Lett. **598**, L67–L70 (2003)
G. Cauzzi, K.P. Reardon, H. Uitenbroek, F. Cavallini, A. Falchi, R. Falciani, K. Janssen, T. Rimmele, A. Vecchio, F. Wöger, The solar chromosphere at high resolution with IBIS. I. New insights from the Ca II 854.2 nm line. Astron. Astrophys. **480**, 515–526 (2008)
R. Centeno, H. Socas-Navarro, B. Lites, M. Kubo, Z. Frank, R. Shine, T. Tarbell, A. Title, K. Ichimoto, S. Tsuneta, Y. Katsukawa, Y. Suematsu, T. Shimizu, S. Nagata, Emergence of small-scale magnetic loops in the quiet-Sun internetwork. Astrophys. J. Lett. **666**, L137–L140 (2007)
M.C.M. Cheung, M. Schüssler, F. Moreno-Insertis, Magnetic flux emergence in granular convection: radiative MHD simulations and observational signatures. Astron. Astrophys. **467**, 703–719 (2007)
M. Collados, I. Rodríguez Hidalgo, L. Bellot Rubio, B. Ruiz Cobo, D. Soltau, TIP (Tenerife Infrared Polarimeter): A near IR full Stokes polarimeter for the German solar telescopes at Observatorio del Teide, in *Astronomische Gesellschaft Meeting Abstracts* (1999), p. 13
M. Collados, A. Lagg, J. Díaz García, E. Hernández Suárez, R. López López, E. Páez Mañá, S.K. Solanki, Tenerife Infrared Polarimeter II, in *The Physics of Chromospheric Plasmas*, ed. by P. Heinzel, I. Dorotovič, R.J. Rutten. ASP Conference Series, vol. 368 (ASP, San Francisco, 2007), pp. 611–616
S. Danilovic, A. Gandorfer, A. Lagg, M. Schüssler, S.K. Solanki, A. Vögler, Y. Katsukawa, S. Tsuneta, The intensity contrast of solar granulation: comparing Hinode SP results with MHD simulations. Astron. Astrophys. **484**, L17–L20 (2008)
B. De Pontieu, T. Tarbell, R. Erdélyi, Correlations on arcsecond scales between chromospheric and transition region emission in active regions. Astrophys. J. **590**, 502–518 (2003)
B. De Pontieu, R. Erdélyi, S.P. James, Solar chromospheric spicules from the leakage of photospheric oscillations and flows. Nature **430**, 536–539 (2004)
B. De Pontieu, S. McIntosh, V.H. Hansteen, M. Carlsson, C.J. Schrijver, T.D. Tarbell, A.M. Title, R.A. Shine, Y. Suematsu, S. Tsuneta, Y. Katsukawa, K. Ichimoto, T. Shimizu, S. Nagata, A tale of two spicules: the impact of spicules on the magnetic chromosphere. Publ. Astron. Soc. Jpn. **59**, 655 (2007a)
B. De Pontieu, S.W. McIntosh, M. Carlsson, V.H. Hansteen, T.D. Tarbell, C.J. Schrijver, A.M. Title, R.A. Shine, S. Tsuneta, Y. Katsukawa, K. Ichimoto, Y. Suematsu, T. Shimizu, S. Nagata, Chromospheric alfvénic waves strong enough to power the solar wind. Science **318**, 1574 (2007b)
B. De Pontieu, V.H. Hansteen, L. Rouppe van der Voort, M. van Noort, M. Carlsson, High-resolution observations and numerical simulations of chromospheric fibrils and mottles, in *The Physics of Chromospheric Plasmas*, ed. by P. Heinzel, I. Dorotovič, R.J. Rutten. Astronomical Society of the Pacific Conference Series, vol. 368 (ASP, San Francisco, 2007c), p. 65
A.G. de Wijn, B.W. Lites, T.E. Berger, Z.A. Frank, T.D. Tarbell, R. Ishikawa, Hinode observations of magnetic elements in internetwork areas. Astrophys. J. **684**, 1469–1476 (2008)
P. Dmitruk, D.O. Gómez, Scaling law for the heating of solar coronal loops. Astrophys. J. Lett. **527**, L63–L66 (1999)
S.B.F. Dorch, Å. Nordlund, On the transport of magnetic fields by solar-like stratified convection. Astron. Astrophys. **365**, 562–570 (2001)

D. Dravins, Å. Nordlund, Stellar granulation, Part Four—Line formation in inhomogeneous stellar photospheres. Astron. Astrophys. **228**, 184 (1990)
D. Dravins, L. Lindegren, A. Nordlund, Solar granulation—Influence of convection on spectral line asymmetries and wavelength shifts. Astron. Astrophys. **96**, 345–364 (1981)
W. Finsterle, S.M. Jefferies, A. Cacciani, P. Rapex, S.W. McIntosh, Helioseismic mapping of the magnetic canopy in the solar chromosphere. Astrophys. J. Lett. **613**, L185–L188 (2004)
L. Fletcher, B. De Pontieu, Plasma diagnostics of transition region "moss" using SOHO/CDS and TRACE. Astrophys. J. Lett. **520**, L135–L138 (1999)
J.M. Fontenla, E.H. Avrett, R. Loeser, Energy balance in the solar transition region. III—Helium emission in hydrostatic, constant-abundance models with diffusion. Astrophys. J. **406**, 319–345 (1993)
A. Fossum, M. Carlsson, High-frequency acoustic waves are not sufficient to heat the solar chromosphere. Nature **435**, 919–921 (2005)
B. Freytag, M. Steffen, B. Dorch, Spots on the surface of Betelgeuse—Results from new 3D stellar convection models. Astron. Nachr. **323**, 213–219 (2002)
A.H. Gabriel, A magnetic model of the solar transition region. R. Soc. Lond. Philos. Trans. Ser. A **281**, 339–352 (1976)
D.J. Galloway, N.O. Weiss, Convection and magnetic fields in stars. Astrophys. J. **243**, 945–953 (1981)
K. Galsgaard, Å. Nordlund, Heating and activity of the solar corona 1. Boundary shearing of an initially homogeneous magnetic field. J. Gephys. Res. **101**, 13445–13460 (1996)
D. Georgobiani, J. Zhao, A.G. Kosovichev, D. Benson, R.F. Stein, Å. Nordlund, Local helioseismology and correlation tracking analysis of surface structures in realistic simulations of solar convection. Astrophys. J. **657**, 1157–1161 (2007)
R.G. Giovanelli, H.P. Jones, The three-dimensional structure of atmospheric magnetic fields in two active regions. Sol. Phys. **79**, 267–278 (1982)
B.V. Gudiksen, Connections: Photosphere–chromosphere–corona, in *Solar MHD Theory and Observations: A High Spatial Resolution Perspective*, ed. by J. Leibacher, R.F. Stein, H. Uitenbroek. Astronomical Society of the Pacific Conference Series, vol. 354 (ASP, San Francisco, 2006), p. 331
B.V. Gudiksen, Å. Nordlund, Bulk heating and slender magnetic loops in the solar corona. Astrophys. J. Lett. **572**, L113–L116 (2002)
B.V. Gudiksen, Å. Nordlund, An ab initio approach to solar coronal loops. Astrophys. J. **618**, 1031–1038 (2005a)
B.V. Gudiksen, Å. Nordlund, An ab initio approach to the solar coronal heating problem. Astrophys. J. **618**, 1020–1030 (2005b)
G.E. Hale, On the probable existence of a magnetic field in Sun-spots. Astrophys. J. **28**, 315 (1908)
W. Hanle, Über magnetische beeinflussung der polarisation der resonanzfluoreszenz. Z. Phys. D: At. Mol. Clust. **18**(1), 5–10 (1991). Reprinted from Z. Phys. **30**, 93 (1924). ISSN 0178-7683 (Print) 1431–5866 (Online)
V.H. Hansteen, Waves and shocks in the solar atmosphere, in *New Solar Physics with Solar-B Mission*, ed. by K. Shibata, S. Nagata, T. Sakurai. Astronomical Society of the Pacific Conference Series, vol. 369 (ASP, San Francisco, 2007), p. 193
V.H. Hansteen, B. Gudiksen, 3D numerical models of quiet Sun coronal heating, in *Solar Wind 11/SOHO 16, Connecting Sun and Heliosphere*. ESA Special Publication, vol. 592 (ESA, Noordwijk, 2005)
V.H. Hansteen, B. De Pontieu, L. Rouppe van der Voort, M. van Noort, M. Carlsson, Dynamic fibrils are driven by magnetoacoustic shocks. Astrophys. J. Lett. **647**, L73–L76 (2006)
V.H. Hansteen, M. Carlsson, B. Gudiksen, 3D numerical models of the chromosphere, transition region, and corona, in *The Physics of Chromospheric Plasmas*, ed. by P. Heinzel, I. Dorotovič, R.J. Rutten. Astronomical Society of the Pacific Conference Series, vol. 368 (ASP, San Francisco, 2007), p. 107
R.A. Harrison, A statistical study of the coronal mass ejection phenomenon. Adv. Space Res. **14**, 23 (1994)
S.S. Hasan, O. Steiner, A. van Ballegooijen, Inferring the chromospheric magnetic topology through waves, in *IAU Symposium*, vol. 247 (2008), pp. 78–81
D.L. Hendrix, G. van Hoven, Z. Mikic, D.D. Schnack, The viability of ohmic dissipation as a coronal heating source. Astrophys. J. **470**, 1192 (1996)
R. Ishikawa, S. Tsuneta, K. Ichimoto, H. Isobe, Y. Katsukawa, B.W. Lites, S. Nagata, T. Shimizu, R.A. Shine, Y. Suematsu, T.D. Tarbell, A.M. Title, Transient horizontal magnetic fields in solar plage regions. Astron. Astrophys. **481**, L25–L28 (2008)
S.M. Jefferies, S.W. McIntosh, J.D. Armstrong, T.J. Bogdan, A. Cacciani, B. Fleck, Magnetoacoustic portals and the Basal heating of the solar chromosphere. Astrophys. J. Lett. **648**, L151–L155 (2006)
S. Jendersie, H. Peter, Link between the chromospheric network and magnetic structures of the corona. Astron. Astrophys. **460**, 901–908 (2006)
H.P. Jones, R.G. Giovanelli, Magnetograph response to canopy-type fields. Sol. Phys. **79**, 247–266 (1982)

P. Judge, Observations of the solar chromosphere, in *Solar MHD Theory and Observations: A High Spatial Resolution Perspective*, ed. by J. Leibacher, R.F. Stein, H. Uitenbroek. Astronomical Society of the Pacific Conference Series, vol. 354 (ASP, San Francisco, 2006), p. 259

L. Kleint, K. Reardon, J.O. Stenflo, H. Uitenbroek, A. Tritschler, Spectropolarimetry of Ca II 8542: probing the chromospheric magnetic field, in *Proceedings of the 5th Solar Polarization Workshop*, Ascona, Switzerland, 17–21 September 2007, ed. by N.K.N.S. Berdyugina, R. Ramelli (2008)

J.M. Krijger, R.J. Rutten, B.W. Lites, T. Straus, R.A. Shine, T.D. Tarbell, Dynamics of the solar chromosphere. III. Ultraviolet brightness oscillations from TRACE. Astron. Astrophys. **379**, 1052–1082 (2001)

A. Lagg, L. Merenda, Measurement of the magnetic canopy over supergranulation cell (2008, in preparation)

A. Lagg, J. Woch, N. Krupp, S.K. Solanki, Retrieval of the full magnetic vector with the He I multiplet at 1083 nm. Maps of an emerging flux region. Astron. Astrophys. **414**, 1109–1120 (2004)

A. Lagg, J. Woch, S. Solanki, N. Krupp, Supersonic downflows in the vicinity of a growing pore: Evidence of unresolved magnetic fine structure at chromospheric heights. Astron. Astrophys. **462**, 1147–1155 (2007)

E. Landi Degl'Innocenti, The determination of vector magnetic fields in prominences from the observations of the Stokes profiles in the D3 line of Helium. Sol. Phys. **79**, 291–322 (1982)

Ø. Langangen, L. Rouppe van der Voort, Y. Lin, Measurements of plasma motions in dynamic fibrils. Astrophys. J. **673**, 1201–1208 (2008a)

Ø. Langangen, B. De Pontieu, M. Carlsson, V.H. Hansteen, G. Cauzzi, K. Reardon, Search for high velocities in the disk counterpart of type II spicules. Astrophys. J. Lett. **679**, L167–L170 (2008b)

Ø. Langangen, M. Carlsson, L. Rouppe van der Voort, V. Hansteen, B. De Pontieu, Spectroscopic measurements of dynamic fibrils in the Ca II $\lambda = 8662$ line. Astrophys. J. **673**, 1194–1200 (2008c)

J. Lee, Radio emissions from solar active regions. Space Sci. Rev. **133**, 73–102 (2007)

J. Leenaarts, S. Wedemeyer-Böhm, Time-dependent hydrogen ionisation in 3D simulations of the solar chromosphere. Methods and first results. Astron. Astrophys. **460**, 301–307 (2006)

J. Leenaarts, M. Carlsson, V. Hansteen, R.J. Rutten, Non-equilibrium hydrogen ionization in 2D simulations of the solar atmosphere. Astron. Astrophys. **473**, 625–632 (2007)

B.W. Lites, K.D. Leka, A. Skumanich, V. Martinez Pillet, T. Shimizu, Small-scale horizontal magnetic fields in the solar photosphere. Astrophys. J. **460**, 1019 (1996)

B.W. Lites, R.J. Rutten, T.E. Berger, Dynamics of the solar chromosphere. II. Ca II H_2V and K_2V grains versus internetwork fields. Astrophys. J. **517**, 1013–1033 (1999)

B. Lites, H. Socas-Navarro, M. Kubo, T.E. Berger, Z. Frank, R.A. Shine, T.D. Tarbell, A.M. Title, K. Ichimoto, Y. Katsukawa, S. Tsuneta, Y. Suematsu, T. Shimizu, S. Nagata, Hinode observations of horizontal quiet Sun magnetic flux and the "hidden turbulent magnetic flux". Publ. Astron. Soc. Jpn. **59**, 571 (2007)

B.W. Lites, M. Kubo, H. Socas-Navarro, T. Berger, Z. Frank, R. Shine, T. Tarbell, A. Title, K. Ichimoto, Y. Katsukawa, S. Tsuneta, Y. Suematsu, T. Shimizu, S. Nagata, The horizontal magnetic flux of the quiet-Sun internetwork as observed with the hinode spectro-polarimeter. Astrophys. J. **672**, 1237–1253 (2008)

A. López Ariste, R. Casini, Inference of the magnetic field in spicules from spectropolarimetry of He I D3. Astron. Astrophys. **436**, 325–331 (2005)

M.A. Loukitcheva, S.K. Solanki, S. White, ALMA as the ideal probe of the solar chromosphere. Astrophys. Space Sci. **313**, 197–200 (2008)

M.J. Martínez González, M. Collados, B. Ruiz Cobo, C. Beck, Internetwork magnetic field distribution from simultaneous 1.56 μm and 630 nm observations. Astron. Astrophys. **477**, 953–965 (2008)

J. Martínez-Sykora, V. Hansteen, M. Carlsson, Twisted flux tube emergence from the convection zone to the corona. Astrophys. J. **679**, 871–888 (2008)

S.W. McIntosh, P.G. Judge, On the nature of magnetic shadows in the solar chromosphere. Astrophys. J. **561**, 420–426 (2001)

S.W. McIntosh, T.J. Bogdan, P.S. Cally, M. Carlsson, V.H. Hansteen, P.G. Judge, B.W. Lites, H. Peter, C.S. Rosenthal, T.D. Tarbell, An observational manifestation of magnetoatmospheric waves in internetwork regions of the chromosphere and transition region. Astrophys. J. Lett. **548**, L237–L241 (2001)

L. Merenda, J. Trujillo Bueno, M. Collados, A magnetic map of a solar filament, in *The Physics of Chromospheric Plasmas*, ed. by P. Heinzel, I. Dorotovič. Astronomical Society of the Pacific Conference Series, vol. 368 (ASP, San Francisco, 2007), p. 347

Å. Nordlund, 3-D Model calculations, in *Small Scale Magnetic Flux Concentrations in the Solar Photosphere*, ed. by W. Deinzer, M. Knölker, H.H. Voigt (1986), p. 83

Å. Nordlund, D. Dravins, Stellar granulation. III—Hydrodynamic model atmospheres. IV—Line formation in inhomogeneous stellar photospheres. V—Synthetic spectral lines in disk-integrated starlight. Astron. Astrophys. **228**, 155–217 (1990)

T.J. Okamoto, S. Tsuneta, T.E. Berger, K. Ichimoto, Y. Katsukawa, B.W. Lites, S. Nagata, K. Shibata, T. Shimizu, R.A. Shine, Y. Suematsu, T.D. Tarbell, A.M. Title, Coronal transverse magnetohydrodynamic waves in a solar prominence. Science **318**, 1577 (2007)

D. Orozco Suárez, L.R. Bellot Rubio, J.C. del Toro Iniesta, S. Tsuneta, B.W. Lites, K. Ichimoto, Y. Katsukawa, S. Nagata, T. Shimizu, R.A. Shine, Y. Suematsu, T.D. Tarbell, A.M. Title, Quiet-Sun internetwork magnetic fields from the inversion of hinode measurements. Astrophys. J. Lett. **670**, L61–L64 (2007)

E.N. Parker, Topological dissipation and the small-scale fields in turbulent gases. Astrophys. J. **174**, 499 (1972)

E.N. Parker, The dissipation of inhomogeneous magnetic fields and the problem of coronae. I—Dislocation and flattening of flux tubes. II—The dynamics of dislocated flux. Astrophys. J. **244**, 631–652 (1981)

E.N. Parker, Magnetic neutral sheets in evolving fields—Part two—Formation of the solar corona. Astrophys. J. **264**, 642 (1983)

H. Peter, B.V. Gudiksen, Å. Nordlund, Coronal heating through braiding of magnetic field lines. Astrophys. J. Lett. **617**, L85–L88 (2004)

H. Peter, B.V. Gudiksen, Å. Nordlund, Forward modeling of the corona of the Sun and solar-like Stars: From a three-dimensional magnetohydrodynamic model to synthetic extreme-ultraviolet spectra. Astrophys. J. **638**, 1086–1100 (2006)

A. Pietarila, J. Hirzberger, S. Solanki, Fibrils in cak. Astron. Astrophys. (2008, in preparation)

C.S. Rosenthal, T.J. Bogdan, M. Carlsson, S.B.F. Dorch, V. Hansteen, S.W. McIntosh, A. McMurry, Å. Nordlund, R.F. Stein, Waves in the magnetized solar atmosphere. I. Basic processes and internetwork oscillations. Astrophys. J. **564**, 508–524 (2002)

L.H.M. Rouppe van der Voort, V.H. Hansteen, M. Carlsson, A. Fossum, E. Marthinussen, M.J. van Noort, T.E. Berger, Solar magnetic elements at 0.1 arcsec resolution. II. Dynamical evolution. Astron. Astrophys. **435**, 327–337 (2005)

L.H.M. Rouppe van der Voort, B. De Pontieu, V.H. Hansteen, M. Carlsson, M. van Noort, Magnetoacoustic shocks as a driver of quiet-Sun mottles. Astrophys. J. Lett. **660**, L169–L172 (2007)

R.J. Rutten, On the nature of the solar chromosphere, in *Solar MHD Theory and Observations: A High Spatial Resolution Perspective*, ed. by J. Leibacher, R.F. Stein, H. Uitenbroek. Astronomical Society of the Pacific Conference Series, vol. 354 (ASP, San Francisco, 2006), p. 276

R.J. Rutten, Observing the solar chromosphere, in *The Physics of Chromospheric Plasmas*, ed. by P. Heinzel, I. Dorotovič, R.J. Rutten. Astronomical Society of the Pacific Conference Series, vol. 368 (ASP, San Francisco, 2007), p. 27

T. Sakurai, Calculation of force-free magnetic field with non-constant α. Sol. Phys. **69**, 343–359 (1981)

T. Sakurai, Computational modeling of magnetic fields in solar active regions. Space Sci. Rev. **51**, 11–48 (1989)

C. Sasso, A. Lagg, S.K. Solanki, R. Aznar Cuadrado, M. Collados, Full-stokes observations and analysis of He I 10830 Å in a flaring region, in *The Physics of Chromospheric Plasmas*, ed. by P. Heinzel, I. Dorotovič R.J. Rutten. Astronomical Society of the Pacific Conference Series, vol. 368 (ASP, San Francisco, 2007), p. 467

W. Schaffenberger, S. Wedemeyer-Böhm, O. Steiner, B. Freytag, Magnetohydrodynamic simulation from the convection zone to the chromosphere, in *Chromospheric and Coronal Magnetic Fields*, ed. by D.E. Innes, A. Lagg, S.A. Solanki. ESA Special Publication, vol. 596 (ESA, Noordwijk, 2005)

W. Schaffenberger, S. Wedemeyer-Böhm, O. Steiner, B. Freytag, Holistic MHD-simulation from the convection zone to the chromosphere, in *Solar MHD Theory and Observations: A High Spatial Resolution Perspective*, ed. by J. Leibacher, R.F. Stein, H. Uitenbroek. Astronomical Society of the Pacific Conference Series, vol. 354 (ASP, San Francisco, 2006), p. 345

G.B. Scharmer, G. Narayan, T. Hillberg, J. de la Cruz Rodriguez, M.G. Lofdahl, D. Kiselman, P. Sutterlin, M. van Noort, A. Lagg, CRISP Spectropolarimetric Imaging of Penumbral Fine Structure. ArXiv e-prints, **806** (2008)

C.J. Schrijver, A.M. Title, The magnetic connection between the solar photosphere and the corona. Astrophys. J. Lett. **597**, L165–L168 (2003)

C.J. Schrijver, A.A. van Ballegooijen, Is the quiet-Sun corona a quasi-steady, force-free environment? Astrophys. J. **630**, 552–560 (2005)

M. Schüssler, A. Vögler, Strong horizontal photospheric magnetic field in a surface dynamo simulation. Astron. Astrophys. **481**, L5–L8 (2008)

R. Skartlien, R.F. Stein, Å. Nordlund, Excitation of chromospheric wave transients by collapsing granules. Astrophys. J. **541**, 468–488 (2000)

H. Socas-Navarro, D. Elmore, Physical properties of spicules from simultaneous spectropolarimetric observations of He I and Ca II lines. Astrophys. J. Lett. **619**, L195–L198 (2005)

H. Socas-Navarro, D. Elmore, A. Pietarila, A. Darnell, B.W. Lites, S. Tomczyk, S. Hegwer, Spinor: visible and infrared spectro-polarimetry at the national solar observatory. Sol. Phys. **235**, 55–73 (2006)
S.K. Solanki, A. Lagg, J. Woch, N. Krupp, M. Collados, Three-dimensional magnetic field topology in a region of solar coronal heating. Nature **425**, 692–695 (2003)
R.F. Stein, Å. Nordlund, Simulations of solar granulation. I. General properties. Astrophys. J. **499**, 914 (1998)
R.F. Stein, Å. Nordlund, Solar small-scale magnetoconvection. Astrophys. J. **642**, 1246–1255 (2006)
R.F. Stein, D. Benson, D. Georgobiani, Å. Nordlund, Supergranule scale convection simulations, in *Proceedings of SOHO 18/GONG 2006/HELAS I, Beyond the Spherical Sun*. ESA Special Publication, vol. 624 (ESA, Noordwijk, 2006)
O. Steiner, Distribution of magnetic flux density at the solar surface. Formulation and results from simulations. Astron. Astrophys. **406**, 1083–1088 (2003)
O. Steiner, Recent progresses in the simulation of small-scale magnetic fields, in *Modern Solar Facilities—Advanced Solar Science*, ed. by F. Kneer, K.G. Puschmann, A.D. Wittmann (2007), p. 321
O. Steiner, P. Murdin, Chromosphere: Magnetic canopy, in *Encyclopedia of Astronomy and Astrophysics* (2000)
O. Steiner, U. Grossmann-Doerth, M. Knoelker, M. Schuessler, Dynamical interaction of solar magnetic elements and granular convection: Results of a numerical simulation. Astrophys. J. **495**, 468 (1998)
O. Steiner, G. Vigeesh, L. Krieger, S. Wedemeyer-Böhm, W. Schaffenberger, B. Freytag, First local helioseismic experiments with CO^5 BOLD. Astron. Nachr. **328**, 323 (2007)
O. Steiner, R. Rezaei, W. Schaffenberger, S. Wedemeyer-Böhm, The horizontal internetwork magnetic field: Numerical simulations in comparison to observations with hinode. Astrophys. J. Lett. **680**, L85–L88 (2008)
J.O. Stenflo, R. Holzreuter, Empirical view of magnetoconvection, in *SOLMAG 2002. Proceedings of the Magnetic Coupling of the Solar Atmosphere Euroconference*, ed. by H. Sawaya-Lacoste. ESA Special Publication, vol. 505 (ESA, Noordwijk, 2002), pp. 101–104
J.O. Stenflo, R. Holzreuter, Flux tubes or fractal distributions—on the nature of photospheric magnetic fields. Astron. Nachr. **324**, 397 (2003)
J.O. Stenflo, C.U. Keller, The second solar spectrum. A new window for diagnostics of the Sun. Astron. Astrophys. **321**, 927–934 (1997)
J.O. Stenflo, A. Gandorfer, R. Holzreuter, D. Gisler, C.U. Keller, M. Bianda, Spatial mapping of the Hanle and Zeeman effects on the Sun. Astron. Astrophys. **389**, 314–324 (2002)
T. Straus, B. Fleck, S. Jefferies, G. Cauzzi, C. G., S. McIntosh, K. Reardon, G. Severino, M. Steffen, The energy flux of internal gravity waves in the lower solar atmosphere. Astrophys. J. Lett. **681**, L125–L128 (2008)
Y. Suematsu, Y. Katsukawa, K. Ichimoto, S. Tsuneta, T. Okamoto, S. Nagata, T. Shimizu, T. Tarbell, R. Shine, A. Title, High resolution observation of spicules in Ca II H with Hinode/SOT, in *American Astronomical Society Meeting Abstracts*, vol. 210 (2007), p. 94.11
S.M. Tobias, N.H. Brummell, T.L. Clune, J. Toomre, Pumping of magnetic fields by turbulent penetrative convection. Astrophys. J. Lett. **502**, L177 (1998)
S.M. Tobias, N.H. Brummell, T.L. Clune, J. Toomre, Transport and storage of magnetic field by overshooting turbulent compressible convection. Astrophys. J. **549**, 1183–1203 (2001)
J. Trujillo Bueno, E. Landi Degl'Innocenti, M. Collados, L. Merenda, R. Manso Sainz, Selective absorption processes as the origin of puzzling spectral line polarization from the Sun. Nature **415**, 403–406 (2002)
J. Trujillo Bueno, N. Shchukina, A. Asensio Ramos, A substantial amount of hidden magnetic energy in the quiet Sun. Nature **430**, 326–329 (2004)
J. Trujillo Bueno, L. Merenda, R. Centeno, M. Collados, E. Landi Degl'Innocenti, The Hanle and Zeeman effects in solar spicules: A novel diagnostic window on chromospheric magnetism. Astrophys. J. Lett. **619**, L191–L194 (2005)
A. Vögler, M. Schüssler, A solar surface dynamo. Astron. Astrophys. **465**, L43–L46 (2007)
A. Vögler, S. Shelyag, M. Schüssler, F. Cattaneo, T. Emonet, T. Linde, Simulations of magneto-convection in the solar photosphere. Equations, methods, and results of the MURaM code. Astron. Astrophys. **429**, 335–351 (2005)
S. Wedemeyer, Multi-dimensional radiation hydrodynamic simulations of the non-magnetic solar atmosphere. Ph.D. thesis, University of Kiel, 2003. http://e-diss.uni-kiel.de/diss_764/
S. Wedemeyer, B. Freytag, M. Steffen, H.-G. Ludwig, H. Holweger, Numerical simulation of the three-dimensional structure and dynamics of the non-magnetic solar chromosphere. Astron. Astrophys. **414**, 1121–1137 (2004)
S. Wedemeyer-Böhm, Point spread functions for the Solar optical telescope onboard Hinode. Astron. Astrophys. **487**, 399–412 (2008)
S. Wedemeyer-Böhm, M. Steffen, Carbon monoxide in the solar atmosphere. II. Radiative cooling by CO lines. Astron. Astrophys. **462**, L31–L35 (2007)

S. Wedemeyer-Böhm, F. Wöger, Small-scale structure and dynamics of the lower solar atmosphere, in *IAU Symposium*, vol. 247 (2008), pp. 66–73

S. Wedemeyer-Böhm, I. Kamp, J. Bruls, B. Freytag, Carbon monoxide in the solar atmosphere. I. Numerical method and two-dimensional models. Astron. Astrophys. **438**, 1043–1057 (2005a)

S. Wedemeyer-Böhm, W. Schaffenberger, O. Steiner, M. Steffen, B. Freytag, I. Kamp, Simulations of magnetohydrodynamics and CO formation from the convection zone to the chromosphere, in *Chromospheric and Coronal Magnetic Fields*, ed. by D.E. Innes, A. Lagg, S.A. Solanki. ESA Special Publication, vol. 596 (ESA, Noordwijk, 2005b)

S. Wedemeyer-Böhm, I. Kamp, B. Freytag, J. Bruls, M. Steffen, A first three-dimensional model for the carbon monoxide concentration in the solar atmosphere, in *Solar MHD Theory and Observations: A High Spatial Resolution Perspective*, ed. by J. Leibacher, R.F. Stein, H. Uitenbroek. Astronomical Society of the Pacific Conference Series, vol. 354 (ASP, San Francisco, 2006), p. 301

S. Wedemeyer-Böhm, H.G. Ludwig, M. Steffen, J. Leenaarts, B. Freytag, Inter-network regions of the Sun at millimetre wavelengths. Astron. Astrophys. **471**, 977–991 (2007)

N.O. Weiss, D.P. Brownjohn, P.C. Matthews, M.R.E. Proctor, Photospheric convection in strong magnetic fields. Mon. Not. R. Astron. Soc. **283**, 1153–1164 (1996)

T. Wiegelmann, Nonlinear force-free modeling of the solar coronal magnetic field. J. Geophys. Res. (Space Phys.) **113**(A12), 3 (2008)

T. Wiegelmann, J.K. Thalmann, C.J. Schrijver, M.L. Derosa, T.R. Metcalf, Can we improve the preprocessing of photospheric vector magnetograms by the inclusion of chromospheric observations? Sol. Phys. **247**, 249–267 (2008)

F. Wöger, S. Wedemeyer-Böhm, W. Schmidt, O. von der Lühe, Observation of a short-lived pattern in the solar chromosphere. Astron. Astrophys. **459**, L9–L12 (2006)

J. Zhao, D. Georgobiani, A.G. Kosovichev, D. Benson, R.F. Stein, Å. Nordlund, Validation of time-distance helioseismology by use of realistic simulations of solar convection. Astrophys. J. **659**, 848–857 (2007)

Magnetic Flux Emergence, Activity, Eruptions and Magnetic Clouds: Following Magnetic Field from the Sun to the Heliosphere

L. van Driel-Gesztelyi · J.L. Culhane

Originally published in the journal Space Science Reviews, Volume 144, Nos 1–4, 351–381.
DOI: 10.1007/s11214-008-9461-x

Abstract We present an overview of how the principal physical properties of magnetic flux which emerges from the toroidal fields in the tachocline through the turbulent convection zone to the solar surface are linked to solar activity events, emphasizing the effects of magnetic field evolution and interaction with other magnetic structures on the latter. We compare the results of different approaches using various magnetic observables to evaluate the probability of flare and coronal mass ejection (CME) activity and forecast eruptive activity on the short term (i.e. days). Then, after a brief overview of the observed properties of CMEs and their theoretical models, we discuss the ejecta properties and describe some typical magnetic and composition characteristics of magnetic clouds (MCs) and interplanetary CMEs (ICMEs). We review some individual examples to clarify the link between eruptions from the Sun and the properties of the resulting ejecta. The importance of a synthetic approach to solar and interplanetary magnetic fields and activity is emphasized.

Keywords Magnetic flux emergence · Magnetic observables · Flare · Coronal mass ejection · Magnetic cloud · ICME

1 Introduction

Day after day enormous amounts of magnetic flux emerge on the Sun: $\Phi \leq 10^{24}$ Mx not accounting for the hidden turbulent magnetic flux (Lites et al. 2007). The observable flux appears on different scales ($10^{18} \leq \Phi \leq 10^{23}$ Mx) forming active regions (ARs), ephemeral

L. van Driel-Gesztelyi (✉) · J.L. Culhane
University College London, Mullard Space Science Laboratory, Holmbury St. Mary, Dorking, Surrey, RH5 6NT, UK
e-mail: Lidia.vanDriel@obspm.fr

L. van Driel-Gesztelyi
Observatoire de Paris, LESIA, FRE 2461(CNRS), 92195 Meudon Principal Cedex, France

L. van Driel-Gesztelyi
Konkoly Observatory of Hungarian Academy of Sciences, Budapest, Hungary

M.J. Thompson et al. (eds.), *The Origin and Dynamics of Solar Magnetism*.
DOI: 10.1007/978-1-4419-0239-9_18

regions (ERs), and on as small a scale as the inter-network field (INF). The frequency distribution of flux emergence over its scale spectrum is smooth and continuous spanning almost five orders of magnitude in flux and eight orders of magnitude in frequency (10^{4}–10^{-4} day^{-1}; Hagenaar et al. 2003; Meunier 2003). The daily magnetic flux emergence rate is highest on the smallest scale, dominating the flux budget at any given moment. However, the overturn time of small-scale flux is only a few minutes (Lites et al. 1996), therefore the long-surviving large-scale flux determines the magnetic properties of the Sun. Furthermore, it is the large-scale flux, which is responsible for the most energetic activity events. Thus in this review we concentrate on the magnetic characteristics of large-scale flux emergence forming ARs.

Large-scale flux emergence reveals physical processes related to magnetic field generation and transport in the sub-photospheric layers. Emergent flux carries clues about the (i) characteristics of dynamo, but also about (ii) conditions in the convection zone with which it interacted during its ascent to the surface and in which the subsurface part of the flux-tube is still embedded. Furthermore, magnetic characteristics of ARs (iii) determine their eruptive activity leading to flares and coronal mass ejections, where the latter expel huge amount of plasma and magnetic field into interplanetary space. Magnetic flux tubes observed *in-situ* close to the Earth and beyond have a direct continuity to their solar source: they represent flux, which has been amplified by the global dynamo at the bottom of the convection zone (Parker 1993 and for a recent review see Gilman 2005), became buoyant, emerged to the surface and was eventually launched by an MHD instability into the interplanetary space. Keeping this continuity in mind we review the principal characteristics of emerging flux (Sect. 2) and active region decay (Sect. 3) focussing on how much we know about the link between these characteristics and the occurrence of solar eruptive events (i.e. our ability to predict flares and coronal mass ejections; Sect. 4). Then, after a brief overview of CME models (Sect. 5), and ICME and magnetic cloud characteristics (Sect. 6) we illustrate by describing a few case studies how well we presently understand the link between solar eruptions and their interplanetary consequences (Sect. 7). We conclude in Sect. 8, emphasizing the importance of a synthetic view.

2 Flux Emergence

2.1 The Three Main Rules of Magnetic Flux Emergence

Since helioseismology is presently unable to ‘detect’ magnetic field in the solar interior deeper than a few Mm, dynamo models must rely on boundary conditions provided by direct observations of the magnetic fields in the solar atmosphere. The three main observationally established rules of solar activity related to the orientation and emergence patterns of sunspot groups (bipoles) during the 11/22-year solar cycle, namely Hale’s law (Hale and Nicholson 1925), the butterfly diagram (or Spörer’s law; Carrington 1858) and Joy’s law (Hale et al. 1919), are the pillars of all successful dynamo models.

Hale’s law states that bipolar active regions (ARs) that are aligned roughly in the east-west direction, on opposite hemispheres have opposite leading magnetic polarities (leading in the sense of solar rotation). The magnetic polarities are alternating between successive sunspot cycles. Spörer’s law (i.e. the butterfly diagram) expresses that the latitudes of flux emergence show a dependence on the solar cycle. When the cycle begins ARs first emerge at high latitudes then tend to emerge at progressively lower latitudes as the cycle progresses. Joy’s law recognizes that there is a systematic deviation from the east-west alignment of bipolar ARs with the leading spots being closer to the equator on both solar hemispheres.

These basic rules were recognized not long after (or even before!) the discovery of solar magnetism a hundred years ago (Hale 1908).

2.2 Additional Characteristics: Asymmetries and Tilt

More recently, some additional physical characteristics of emerging flux were recognized providing further clues to the flux generation by the dynamo and flux transport in the convection zone, as well as for understanding eruptive activity.

The asymmetries in bipolar ARs, namely that (i) the leading sunspots are larger and longer-lived than following spots and (ii) in the divergent motions during emergence the leading sunspots move much faster westward than the following spots eastward, were explained as being due to a systematic eastward tilt of emerging flux tubes (van Driel-Gesztelyi and Petrovay 1990). MHD simulations in the thin flux tube approximation showed that buoyant rising flux tubes become inclined to the vertical while emerging through the convection zone due to the conservation of the angular momentum (Moreno-Insertis et al. 1994; Caligari et al. 1995; Abbett et al. 2001). Conservation of angular momentum induces an eastward (retrograde) plasma flow in the flux tube decreasing the plasma pressure in the leading, while increasing it in the following leg (Fan et al. 1993). Pressure equilibrium requires an inverse change in magnetic pressure leading to an asymmetry in stability between the leading and following spots in ARs. However recent 3-D spherical shell inelastic MHD simulations of the buoyant rise of magnetic flux tubes through the convection zone by Fan (2008) presented a very different picture on the origin of these asymmetries. She showed that due to asymmetric stretching of the rising flux tube by the Coriolis force, a field strength asymmetry develops with the field strength in the leading leg being stronger than the field in the following leg, which results in larger and more stable leading spots. Another consequence is that the leading legs of Ω-loops become more buoyant, producing an asymmetry in the Ω-loops' shape which is *opposite* to that of the simulations in the thin flux tube approximation. Therefore the asymmetry in the divergent motions between the leading and following spots of emerging bipoles cannot be explained by the sub-photospheric shape of the emerging Ω-loop as proposed by van Driel-Gesztelyi and Petrovay (1990). Instead, based on the results of these 3-D simulations, we suggest that the asymmetry in sunspot proper motions is caused by the faster rise of the leading than that of the following leg of the Ω-loop.

The tilt of bipolar ARs relative to the E–W direction, which increases with latitude and is described by Joy's law was shown to be caused by the Coriolis force (Schmidt 1968; Fisher et al. 1995). However, tilt can also be caused by large-scale vortices in the convective zone deforming the rising flux tube (López-Fuentes et al., 2000, 2003). The effect of turbulent buffeting of rising flux tubes is well demonstrated by departures from Joy's law which increase with decreasing flux content of the emerging bipoles (Harvey 1993; Longcope and Fisher 1996). Such turbulent perturbations, if created in the topmost layer of the convection zone, should relax rapidly (Longcope and Choudhuri 2002) turning the flux tube to conform with Joy's law.

2.3 Inherent Twist and Its Implications

The potentially widest-ranging impact came from the recognition, that emerging flux is inherently twisted. Leka et al. (1996) were the first to provide observational evidence for flux emergence in a non-potential state, inspiring research contributing to a revival of interest in helicity. Non-potential magnetic flux emergence has a very important relevance for solar activity: such emerging flux carries free magnetic energy 'ready' to be released. Photospheric

shearing motions, which have been long thought to be the generators of magnetic stresses, may simply reflect the emergence of a twisted structure as successive cross-sections of a helical structure can easily be mis-interpreted as shearing flows Démoulin and Berger (2003). Nevertheless, plasma flows do exist on the Sun, therefore their effects on emerged fields should not be dismissed. Rather, twisted flux emergence and large-scale flows are both responsible for the free energy level of the magnetic field structures we see on the Sun.

Prior to the observational evidence by Leka et al. (1996) theoretical arguments have been raised in favour of non-potential flux emergence from considerations of the energy available for flaring (McClymont and Fisher 1989; Melrose 1992). Furthermore, Schüssler (1979) and later Longcope et al. (1996), through MHD simulations, showed that non-twisted flux cannot even make it through the convection zone due to a strong tendency for fragmentation. However, the flux tube cannot be fragmented by eddies forming in its wake but can remain coherent if it is sufficiently twisted (Moreno-Insertis and Emonet 1996). Many other simulations have been carried out since, verifying this result while probing deeper into details of inherent twist in emerging flux tubes (see e.g. Murray and Hood 2008). These simulation results imply that inherent twist is a general property of flux emergence on the Sun, i.e. that all the large-scale flux that has crossed the convection zone must be twisted and must therefore possess magnetic helicity.

Magnetic helicity is a quantitative, mathematical measure of the chiral properties of magnetic structures. Chirality patterns discovered in active regions, coronal loops, filaments, coronal arcades and interplanetary magnetic clouds (Pevtsov and Balasubramaniam 2003, and references therein) indicate that the Sun preferentially exhibits left-handed features in its northern hemisphere and right-handed features in the south. A right-handed twist and a clockwise rotation of the loops when viewed from above implies positive helicity, and *vice versa* for negative helicity. Exceptions to these helicity rules occur in most categories of solar activity at a significant level (20–35%). Nevertheless, the Sun's preference for features adhering to these rules is suggestive of underlying mechanisms related to the working of the dynamo and differential rotation that are, evidently, global in scope.

However, it must be noted that observations indicate a relatively low level of twist in emerging flux regions as deduced from both current helicity (Longcope et al. 1999) and photospheric magnetic helicity flux measurements (Démoulin and Pariat 2008, and references therein), the latter being compatible with a 0.01–0.2 end-to-end twist of field lines between their photospheric footpoints in emergent flux ropes.

Recent 3-D spherical shell inelastic MHD simulations of the buoyant rise of magnetic flux tubes through the convection zone by Fan (2008), cited above, also, indicate that the initial level of twist must be lower than indicated by previous simulations. Fan's 3-D simulations show that for tubes with the twist rate that is necessary for a cohesive rise, the twist-induced tilt (deformation of the flux tube at its apex) dominates that caused by the Coriolis force, and furthermore, the twist-induced tilt is of the wrong direction (opposite to the observational Joy's law) if the twist is left-handed (right-handed) in the northern (southern) hemisphere, following the observed hemispheric preference of the sign of the active region twist. In order for the emerging tube to show the correct tilt direction (consistent with observations), the initial twist rate of the flux tube needs to be less than half of that needed for a cohesive rise. Under such conditions, however, severe flux loss was found during the rise, with less than 50% of the initial flux remaining in the Ω-loop by the time it reaches the surface.

The emergence of even a mildly twisted flux rope has its caveats as dense plasma accumulating in its concave-up parts located below the axis of the flux rope practically anchor its U-loop sections in and below the photosphere. Furthermore, due to fast changes in physical conditions, the flux rope has great difficulties in crossing the photosphere (e.g., Magara

2004; Manchester et al. 2004) leading to its fragmentation. Emergence of a flux rope therefore must involve many episodes of magnetic reconnection to succeed (Pariat et al. 2004). Nevertheless, characteristic magnetic polarity distribution patterns in longitudinal magnetic maps of emerging flux regions dubbed "magnetic tongues", first identified and interpreted as the signature of the azimuthal field component in an emerging flux rope by López Fuentes et al. (2000) do indicate that there is an overall organization in the emerging flux tube, which is compatible with a global twist. These "magnetic tongues" are present as long as the top part of the twisted flux rope crosses the photosphere and they can be used as a proxy for determination of the helicity sign of an active region (Green et al. 2007).

2.4 Magnetic Helicity

Magnetic helicity quantifies how the magnetic field is sheared and twisted compared to its lowest-energy state; the potential (or current-free) field. However, unlike other physical quantities of magnetic stress (e.g. shear) magnetic helicity can be precisely *quantified* in a given magnetic configuration and possesses the unique property of being almost completely conserved even in resistive MHD on time-scales involved in solar activity events and during magnetic reconnection (Berger 1984). Helicity generation is a natural product of dynamo processes and potentially saturates the dynamo (α-effect quenching); for a recent review, see Brandenburg and Subramanian (2005).

Magnetic helicity is the volume integral of the product of magnetic vector potential $\mathbf{A}$ and magnetic field $\mathbf{B}$ ($\mathbf{B} = \nabla \times \mathbf{A}$). Since $\mathbf{A}$ is not a measurable quantity and has a gauge freedom, magnetic helicity remained a theoretical concept for decades. It was only very recently, that theoretical developments allowed observational applications, making helicity studies the most dynamically developing field of solar physics. As a more easily computable quantity, current helicity ($H = \int \mathbf{B} \cdot \mathbf{j}\, d^3x$, with $\mu_0 \mathbf{j} = \nabla \times \mathbf{B}$) was widely used. Current helicity measures the curl of $\mathbf{B}$ along the magnetic field quantifying local twist. Magnetic and current helicity usually have the same sign, but they also have basic differences, e.g. current helicity is not a conserved MHD quantity. Observational studies of magnetic helicity and current helicity both helped to quantify twisted flux emergence and enhance our understanding of the build-up of eruptive activity on the Sun.

Based on tracking photospheric flows, methods have been developed to measure the magnetic helicity flux (or rate) through the photosphere ranging from the first estimation by Wang (1996) and the first measurements by Chae (2001) to more recent developments involving a new method to measure helicity flux density, or helicity flux through the photosphere per unit surface (Pariat et al. 2005, 2006). The bulk of the helicity is clearly injected during the main phase of flux emergence with at first (for about two days) a lower, followed by a higher rate, increasing in tandem with the magnetic flux (Jeong and Chae 2007; Tian and Alexander 2008). The temporal profile of magnetic helicity flux is indicative of helicity brought up by a twisted flux tube (c.f. Cheung et al. 2005; Chae et al. 2004; Pariat et al. 2005, and for an assessment see a review by Démoulin 2007). Coronal helicity content of ARs can be computed from magnetic extrapolations (e.g. Démoulin et al. 2002; Green et al. 2002a; Mandrini et al. 2005). The coronal helicity content of ARs, like that of emerging flux, appears to be modest, being equivalent to that of a twisted flux tube having 0.2 turn with $\mathbf{H}_{\max}(\mathrm{AR}) \approx 0.2\text{–}0.01\Phi^2$, where Φ is the total magnetic flux of the AR (Démoulin 2007; Démoulin and Pariat 2008). From coronal helicity estimates before and after a CME, the loss of magnetic helicity from an AR was assessed (Bleybel et al. 2002). Methods for helicity calculations in magnetic clouds (MCs) have been developed and compared with the decrease of helicity in the CME source region, the two being in satisfactory agreement (Mandrini et al. 2005; Luoni et al. 2005). For an insightful review on recent theoretical

and observational results on magnetic helicity in the Sun and the interplanetary space see Démoulin (2007).

The cycle-invariant relentless accumulation of helicity in the solar corona being brought up by emerging flux and generated by differential rotation combined with its well-conserved nature also poses a problem. Though some helicity can be canceled between the two hemispheres through magnetic reconnection of opposite helicity structures (Pevtsov 2000), such reconnections involve only a small fraction of the flux present on the Sun, therefore this mechanism is probably insufficient to relieve the buildup. Rust (1994) and Low (1997) suggested that the Sun only avoids endless accumulation of helicity in the solar atmosphere by ejecting helicity via CMEs.

The well-conserved nature of magnetic helicity provides us with a quantitative measure to be traced and compared as buoyant magnetic flux travels from the tachocline through the convection zone, emerges through the photosphere to the corona and is ejected into interplanetary space during CME events reaching the vicinity of the Earth and beyond as a magnetic cloud or ICME (Démoulin 2008).

2.5 Nesting Tendency of Flux Emergence

Harvey and Zwaan (1993) found a 22-fold higher emergence rate within existing ARs than elsewhere. Furthermore, there is a tendency for ARs to emerge in the immediate vicinity of an existing AR, or at the site of a previous AR, forming 'activity nests', which may exist as long as 6–7 months (Brouwer and Zwaan 1990). The nested nature of flux emergence is very strong, nearly 50% of all emergent bipoles being part of an active nest or activity complex (Schrijver and Zwaan 2000). The recurrent nature of flux emergence ('active longitudes', first noted by Carrington in 1858), has been linked to longitudinal wave numbers of magnetic instabilities in a concentrated toroidal field (Gilman and Dikpati 2000) and more recently to shallow-water instability of differential rotation and toroidal field bands in the solar tachocline (Dikpati and Gilman 2005).

This nested nature is reflected in the formation of some of the large, magnetically complex ARs as several bipolar ARs emerge separated, but in close proximity and in close succession within a few days (Schrijver and Zwaan 2000). Magnetic complexity and activity level are closely linked, as we will discuss in Sect. 3.

3 Decay of Active Regions

Once all the flux has emerged, or possibly even before that (Wang et al. 1991), active regions start to decay. After sunspots reach maximum area partially through coalescence of smaller umbrae, spots start shrinking and breaking up. Vigorous moving magnetic feature (MMF) activity is seen around spots carrying flux away (Harvey and Harvey 1973; Hagenaar and Shine 2005; Sainz Dalda and Martínez-Pillet 2005; Ryutova and Hagenaar 2007). There is a notable asymmetry in the time spent by an active region in emergence and decay: emergence lasts for hours to days (≤ 5 days; Harvey 1993), while the decay of spots may last from days to several weeks (e.g. Hathaway and Choudhary 2008) and even in some cases months (van Driel-Gesztelyi et al. 1999). Active regions, even after the disappearance of their spots, remain distinguishable from their magnetic environment for up to seven months while their magnetic flux in a magnetically undisturbed environment during solar minimum spreads over an ever-increasing area (Fig. 1b; see also van Driel-Gesztelyi 1998) forming large bipolar regions shaped by an interplay of convective flows and differential rotation,

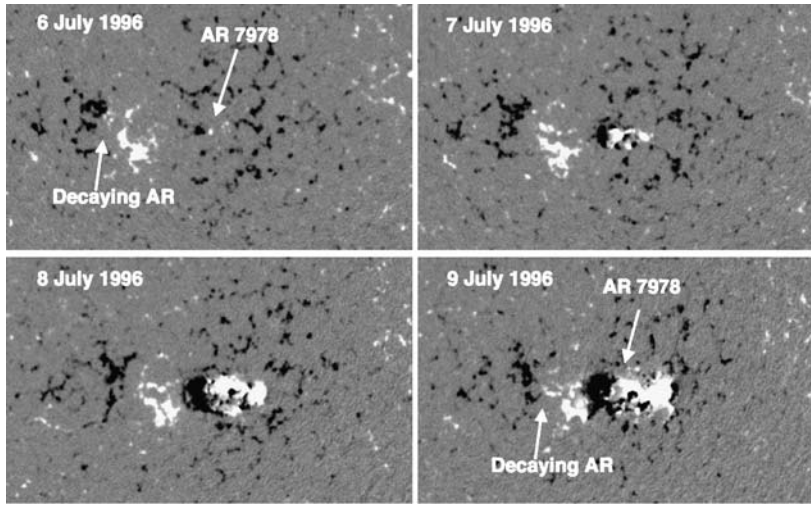


Fig. 1a SOHO/MDI magnetograms showing the emergence of NOAA AR 7978 in July 1996. The boundaries of the emerging bipole are outlined with a white contour. Note the decaying AR East to the growing AR 7978, which totally disappeared by the next rotation shown in Fig. 1b, presumably due to effective magnetic cancellation processes

which slowly become part of the 'background field'. During the decay process of active regions large-scale magnetic complexity is decreasing then disappears due to effective cancellation processes, unless the active region is part of an active nest and thus a place of repeated large-scale flux emergence. In the latter case, however, the evolution of the individual bipoles can be significantly shortened by magnetic cancellation. While the decay phase is marked by a decrease of magnetic flux density accompanied by a decrease of all plasma parameters (temperature, emission measure, pressure; van Driel-Gesztelyi et al. 2003) there is a remarkable growing feature in the AR. Filaments, which are absent, or if present, are short and variable when the AR is young, become stable and can reach a length of 10^5 km or more, becoming increasingly parallel to the equator. Though flare activity is fast disappearing with the decrease of magnetic flux density, coronal mass ejections may well occur during the decay stage due to the repeated eruption of the long filament.

3.1 The Effect of Magnetic Evolution on Activity

An exceptional opportunity for observing active region emergence and decay as well as the accompanying flare and CME activity arose during the previous solar minimum, when solar activity was dominated by a single isolated active region NOAA AR 7978 in the period of July–November 1996. The number of flares and CMEs which originated from this AR is shown in Table 1. During the emergence and the two following rotations, the AR produced numerous flares (including an X2.6 flare and CME event on 9 July, see Dryer et al. 1998) until the disappearance of the main spots after its third rotation. On the other hand, CME activity, which was at first mainly related to flare events, continued at a surprisingly high level for the next three rotations (van Driel-Gesztelyi et al. 1999), while the magnetic helicity content of the dispersed active region remained reasonably high (Démoulin et al. 2002). However, none of the late CMEs were related to flare events above the GOES B1 level. Table 1 lists the number of flares in different GOES classes and of

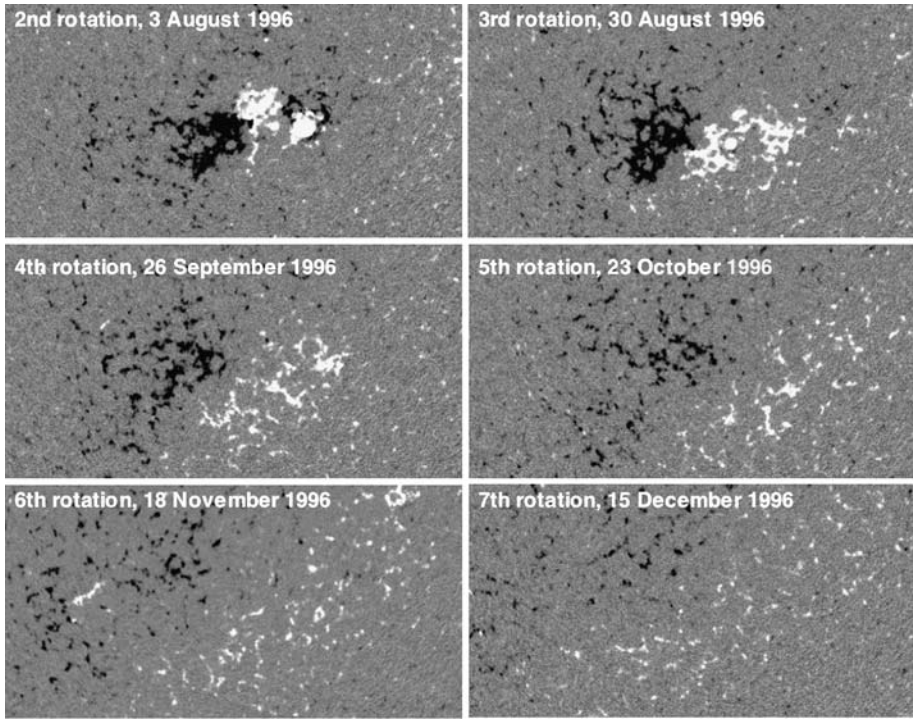


Fig. 1b SOHO/MDI magnetograms showing the decay phase (Rotations 2–7) of NOAA AR 7978 (cf. Fig. 1a) Note the simplification of the magnetic structure with time. Figure adapted from van Driel-Gesztelyi (1998)

the observed CMEs. The flare data in Table 1 was taken from GOES X-ray and optical event catalog (http://www.lmsal.com/SXT/). CMEs have been identified in SoHO/EIT data (Delaboudinière et al. 1995) and SoHO/LASCO (Brueckner et al. 1995) observations by Démoulin et al. (2002). The low level of activity during the lifetime of AR 7978 allowed the identification of even back-side CMEs that originated from this AR when it was on the far side of the Sun. The number of CMEs has been corrected for data gaps assuming that the frequency of the CMEs was the same during the gaps as during observing times (Table 1). This doubles the sampling of CME relative to that of the flares which could only be observed when the AR was on the visible hemisphere.

It is clear from the examples shown in Table 1 that the highest activity occurs during the emergence phase. High magnetic flux density in an AR increases the probability of high reconnection rate in activity events and thus the appearance of bright flare ribbons. CMEs occurring during the decay phase due to filament eruption may well have the same underlying physics, but the accompanying activity manifestation (two-ribbon flare) will be weaker and beyond a certain point into the decay phase even below detection level.

There is a clear consensus linking the energy source of eruptive activity to free energy in the magnetic field. Interesting recent results by Règnier and Priest (2007) compute the altitudes and magnitudes of free energy in an emerging flux region with low magnetic current density and a decaying active region with high current density. The decaying active region has a simple magnetic configuration in which the distribution of strong currents indicate a twisted flux tube with a free energy of 2.6×10^{31} erg (40% of the total energy) stored as

Table 1 Evolution of Flare and Coronal Mass Ejection (CME) activity during the lifetime of an isolated active region NOAA 7978 (July–November 1996; based on Démoulin et al. 2002)

Rotation No.	Day of CMP	Number of flares					Number of CMEs	
		X	M	C	B	Subflare[1]	Observed	Corrected[2]
1st[3]	07 July	01	02	14	11	11	08	11
2nd[4]	02 August	–	–	–	16	01	05	05
3rd	30 August	–	–	01	08	–	02	03
4th[5]	25 September	–	–	–	–	1(?)[6]	05	05
5th	23 October	–	–	–	–	1(?)	03	04
6th	19 November	–	–	–	–	2(?)	03	03

[1]GOES flux is in the B-flare range, but there is no GOES-class given in the list

[2]Corrected for data gaps

[3]Emergence

[4]Peak magnetic flux

[5]Sunspots have disappeared

[6]GOES flux reaches B1 level, but source region is uncertain

high as ∼50 Mm. The weak currents in the newly emerged complex active region do not dramatically modify the connectivity of the magnetic field lines and the magnetic topology of the configuration i.e. the departure from a potential field is small, but the excess magnetic energy of 2.4×10^{31} erg, which represents only 2.5% of the total energy of the AR, is stored in the low corona and is still enough to power flares. Corroborating evidence for low-lying free energy storage in an emerging flux region is provided using non-linear force-free field extrapolations from vector magnetic field measurements. Schrijver et al. (2008) find evidence for filamentary coronal currents located $\leq$20 Mm above the photosphere in an emerging AR 10930 prior to the X-class flare and CME event on 13 December 2006.

4 Relationship of Magnetic Properties to Activity

Less than 10% of the ARs which emerge on the Sun will ever produce a major (M, X) flare (Georgoulis and Rust 2007). However, do we understand what makes one active region produce energetic flares and fast CMEs and another quiet? Which are the distinguishing features of regions of highly activity?

Based on decades-long observations the most frequently mentioned magnetic characteristics of ARs in which large solar flares/CMEs occur are:

- fast evolution (flux emergence)
- large size (high magnetic flux)
- complex magnetic field topology—δ-spots
- long magnetic inversion lines
- high magnetic shear
- strong field gradients
- high helicity and/or high free energy content.

However, which one of these is the most important? Is there one single determining characteristic or is it perhaps a combination of different factors which leads to important

eruptions? Seeking answers to these questions, several groups published series of papers carrying out parametric studies of the photospheric magnetic field in an attempt to find the link with flare and/or CME. Since the proof of the pudding is in the eating, most of the groups expressed the results in probability (success) of forecasting activity events.

4.1 Magnetic Parametric Studies and Short-Term Activity Forecast

K.D. Leka and G. Barnes published a series of papers between 2003 and 2007 in pursuit of finding the best magnetic field parameters for predicting imminent flare activity. In Paper I, Leka and Barnes (2003a) using time-series of photospheric vector magnetic data for three ARs, derived 30 (!) magnetic parameters. The evolution of these was studied in pre-flare vs. flare-quiet periods. No obvious flare-unique signature was found. In Paper II, Leka and Barnes (2003b) took a statistical approach based on discriminant analysis (DA) for 7 ARs in 10 flaring and 14 quiet periods. The conclusions were disappointing: no single parameter appeared to separate reliably the samples of these two populations without producing false alarms. However, when multiple parameters were considered simultaneously, the samples separated in some cases.

In Paper III (Barnes and Leka 2006) coronal topology or complexity was analyzed in a parametric approach using the magnetic charge topology model (Barnes et al. 2005), which separated the two samples more successfully. However, the small sample size prevented them from reaching definite conclusions. In Paper IV, Leka and Barnes (2007) analyzed daily samples of the two populations on a much larger dataset (1200 magnetograms, 496 ARs), and at the small-flare (C1) level the most powerful predictors were found to be two strongly correlated variables: total magnetic flux $\Phi_{\rm tot}$ and total electric current $I_{\rm tot}$. The best discriminant functions resulted from a combination of $\Phi_{\rm tot}$ or $I_{\rm tot}$ with another uncorrelated variable, e.g. magnetic shear (80% success *vs* 70% for flare quiet case). On the larger (M1) flare level excess photospheric energy outperformed other variables (93% success *vs* 90% for flare quiet case). However, they concluded, that "The state of the photospheric magnetic field at any given time has limited bearing on whether that region will be flare productive." Are these negative summed-up results too pessimistic? Perhaps these authors have put the stakes too high, or rather, the flare importance level they wished to predict, too low and were drowned with a large number of photospheric magnetic parameters. Another possible approach is to go only for the big activity events trying to spot some distinguishing differences in the hopefully more important photospheric signatures.

In a series of papers the MSFC group, Falconer, Moore and Gary (Falconer 2001; Falconer et al. 2002, 2003, 2006) explored the significance and correlation of three to six magnetic parameters in an increasingly large sample (4–31) of bipolar ARs and their forecasting power for the occurrence of CMEs. Three of the parameters represented measures of the total non-potentiality of the AR: (i) total length of magnetic inversion lines with high shear (L_{SSM}), and (ii) high gradient (L_{SGM}), and (iii) the net vertical current I_N. Two parameters represented measures of the degree of overall twist: (iv) the 'best' α, $\alpha_{\rm BC}$, and (v) the 'magnetic twist' parameter $\alpha_{IN} = \mu I_N/\Phi$, where Φ is (vi) the total magnetic flux, an independent parameter. They found the best predictive power of parameters for magnetic twist and flux content, but remarked that the total magnetic free energy in an AR is stronger determinant of its CME productivity than is the field's overall twist (helicity) alone. However, recall that a *proxy* was used for magnetic free energy $\alpha_{IN}\Phi$ ($\alpha_{IN} = \mu I_N/\Phi$) and for helicity the "twist" parameter α (while helicity, in fact, rather is $\alpha\Phi^2$). Furthermore, the α parameters used are global ones, disregarding that the twist may be localized in ARs.

Wang et al. (2006) carried out case studies for five ARs which produced six $\geq$ 5X flares. They found that locations of high shear (derived from vector magnetograms) and gradient (from LOS magnetograms) were well correlated ($\sim$90%). Magnetic gradient appeared to be a better proxy than shear for predicting where a major flare might occur. However, it is noteworthy that the mean gradient for these extreme flaring neutral lines was between 0.14–0.50 G km^{-1}, which is 2.3–8 times higher than the usual magnetic gradient values. Horizontal and vertical shearing flows in the vicinity the neutral line prior to and during an X10 flare (Deng et al. 2006) confirm the concentration of free energy on small spatial scales.

The NAO (Beijing) group, Cui et al. (2006, 2007) also analyzed these parameters and their predicting power for flares using 1353 vector magnetograms from Huairou. Using this broader and less extreme sample they found that high-shear and high-gradient neutral lines as defined by the MSFC group for CME prediction appear at about the same time, but they do not overlap much in space. However, the length of their overlap both in space and time gave the best correlation with flare productivity.

The Lockheed group tried to quantify the direct cause of non-potentiality in the active region corona instead of flare forecast. Schrijver et al. (2005) extrapolated the photospheric field to study the deviation of the coronal field in 95 ARs from the potential configuration comparing field lines from potential extrapolations to observed coronal loops (TRACE). They concluded that significant deviation from non-potentiality occurs when (i) new flux has emerged within or very near a region within the last $\sim$30 hr, creating complex polarity separation lines, and (ii) rapidly-evolving opposite-polarity concentrations are in contact at 4″ resolution. As for flare frequency, they found that flares occur 2.4 times more frequently and are 3.3 times brighter (in SXRs) in non-potential ARs, which provides another evidence for the role of free magnetic energy plays in flares.

However, providing flare forecast using a photospheric parameter which can easily be derived from SOHO/MDI LOS magnetograms remained too tempting. Schrijver (2007) took the challenge, and defined a new metric (R) for this purpose quantifying high-gradient strong-field polarity inversion lines. First, in an MDI magnetic map strong positive and negative magnetic areas ($\geq$150 Mx cm^2) are identified using 6″ $\times$ 6″ kernels (2.2×10^{16} cm^2 area). The parameter R is defined as the summed-up flux of the *overlap* between the positive and negative strong-flux areas. Figure 2 shows an example: MDI magnetogram of AR 10720 on 18 January 2005 (left panel; note that in the original paper this AR is mis-identified) and the location of high-gradient, strong-field, polarity-separation lines, which, after summing their absolute values, yields R (right panel). Forecast success of a major flare (M or X GOES class) within 24 hours had a probability of almost 1 when $R \geq 2 \times 10^{21}$ Mx ($\log R \geq 4.8$), while the probability was almost zero when $R \leq 10^{19}$ Mx ($\log R \leq 2.8$). A great advantage of this method is that determination of R is readily automated, making it an effective tool for flare forecasting. The apparent importance of high-gradient strong inversion lines, which are considered as characteristics of emergence of compact electrical currents, provide further evidence for flux emergence in a strongly non-potential state or with twist (helicity) and its importance for eruptive activity.

4.2 Metrics and Effects of Magnetic Complexity

In order to produce a metric for magnetic complexity, Georgoulis and Rust (2007) introduced the effective connected magnetic field of active regions. Building on the magnetic charge topology model developed by Barnes et al. (2005), they resolve an AR having N $(m + l)$ flux concentrations each with a flux Φ_k and centroid position r_k, there are $m \times l$ magnetic connectivities, each having flux of Φ_{ij} with $L_{ij} = |r_i - r_j|$ separation length. The

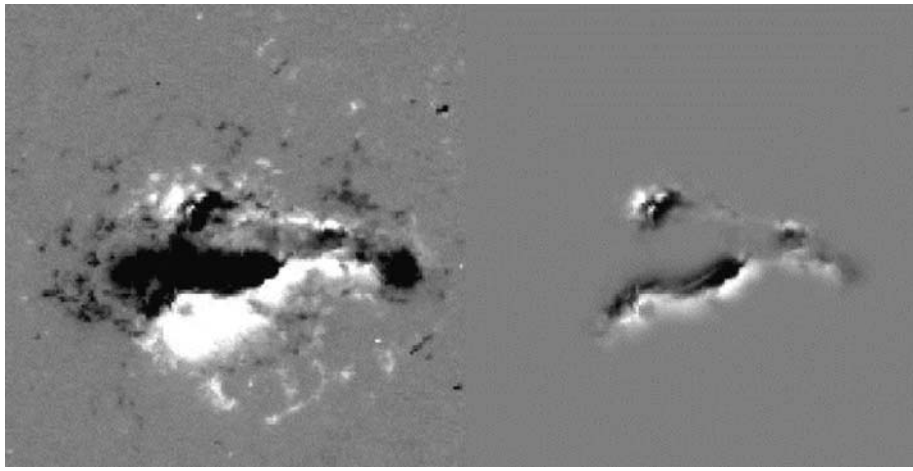


Fig. 2 SOHO/MDI magnetogram of NOAA AR 10720 on 18 January 2005 (*left panel*) and the location of high-gradient, strong-field, polarity-separation lines, which, after summing their absolute values, yields the metric R as proposed by Schrijver (2007) for flare forecast (*right panel*). Figure adopted from Schrijver (2007)

effective connected magnetic flux, $B_{\rm eff} = \sum_{i-l}^{m} \sum_{j-l}^{l} \Phi_{ij} L_{ij}^2$ for $\Phi_{ij} \neq 0$. Calculating $B_{\rm eff}$ for 298 ARs observed between 1996 and 2005, which fell into three groups: 47 and 46 were X-flare and M-flare productive, respectively, and 205 had no major flares, it was found that $B_{\rm eff}$ exceeded 1600 and 2100 G for M and X-class flares respectively, at 95% probability.

Continuing this work, Georgoulis (2008) calculated $B_{\rm eff}$ for 23 CME source regions and studied its correlations with flare magnitude, CME velocity and CME acceleration magnitude, which all seemed to increase with increasing effective connected magnetic field.

4.3 Photospheric Fields Are Relevant, but Are They Sufficient?

Another clue for the role of photospheric magnetic fields in flares is provided by a clear sign that the photospheric magnetic field changes abruptly and non-reversibly during the flare impulsive phase (see e.g. Sudol and Harvey 2005, for a recent survey). Rapid changes in sunspot structure have also been detected by Chen et al. (2007) in 40% of X-class flares, 17% of M flares and 10% C flares.

Free magnetic energy for eruptions is stored in the corona. A typical pre-eruption configuration is a stressed core field (around a high-gradient and high-shear magnetic inversion line) held down by an overlying stabilizing (arcade) field. The latter underlines the importance of a larger-scale magnetic environment in the eruption process.

Whether an eruption fails or succeeds depend on the strength and profile of the overlying fields: rapid decrease, which is typical for complex active regions, being more favorable for full and fast eruption (Török and Kliem 2005, 2007; Liu 2008). Therefore, besides the characteristics of the photospheric magnetic field in an AR one has to assess the coronal conditions as well in order to understand and thus be able to forecast eruptions.

4.4 Helicity Injection, Content and Eruptive Activity

When measurements of the helicity content of active regions became possible, they seemed to bring a crucial factor to understanding the initiation of CMEs.

Andrews (2003) analyzed X and M flares in the period of 1996–1999. The sample consisted of 229 flares with good LASCO coverage. About 40% of M-class flares had no associated CME, while all of the X-class flares were found to be associated with CMEs (however, see Green et al. 2002b for a counter-example). Nindos and Andrews (2004) studied the same data set to find out what the difference is between eruptive and confined big (M) flares, asking the question: is it helicity that makes the difference? They computed coronal helicity content of the ARs using magnetic extrapolations in the linear force-free field approximation fitting computed field lines to observed coronal loops. The two samples appeared well separated: active regions which produced flares with accompanying CMEs had, on average, about a factor of four times more magnetic helicity than the ARs which produced M-class flares without CME, an impressive result.

Barry LaBonte had started a comprehensive project to nail down the role of helicity in eruptive activity events. Unfortunately, his untimely death left the project half-finished (LaBonte et al. 2007; Georgoulis and LaBonte 2007). Using automated processing the photospheric magnetic helicity flux was computed for 48 X-flare producing and 345 non-flaring ARs observed in the period 1996–2005. It was found that most regions grow or decay 10% day^{-1}, except for EFRs, with most of the X-flaring ARs being in growth phase. Causal links were demonstrated between both peak helicity injection rate and 4–7 day helicity changes, X-flaring and CME production. Peak helicity flux prior to X-class flares producing a CME exceeded 6×10^{36} $\mathrm{Mx}^2\,\mathrm{s}^{-1}$.

To the questions ‘is large helicity necessary condition for big flare/CME?’ or ‘What is more important, large free magnetic energy or large magnetic helicity?’ We still cannot give a confident answer. A systematic study is still lacking.

5 Eruptions

While there are quasi-steady outflows of matter from the Sun e.g. fast and slow solar wind, and also comparatively low mass transient outflows such as coronal jets, coronal mass ejections (CMEs) represent the principal solar eruptive phenomenon. In these large-scale matter expulsions, around 10^{13} kg of solar material embedded in 10^{20}–10^{22} Mx magnetic flux is involved with a typical total energy of $\approx 10^{25}$ J. The material drives a shock through the corona and into the interplanetary medium (IPM) with velocity of typically 1000 $\mathrm{km\,s}^{-1}$ but ranging up to three times this value. Systematic CME studies began through the use of space-borne coronagraphs with early work being carried out by instruments on the Skylab and Solar Maximum Missions (MacQueen et al. 1974, 1980). Coronagraphs respond to photospheric white light scattered by the expanding ejected material. A typical CME structure is indicated in Fig. 3 while a schematic diagram of the related shock, erupting material, cavity and the underlying prominence is also shown. More recently a substantial body of CME observations, undertaken with the Large Angle Spectrometric Coronagraph on SOHO (Brueckner et al. 1995), has provided considerable new information about these events.

A number of on-disc and low coronal signatures are now recognized as being associated with CMEs. These include prominence or filament eruptions, post-eruption arcades and Moreton waves (Moreton and Ramsey 1960). In addition type II radio bursts (Wild and McCready 1950) have been associated with the propagating CME shock. More recently coronal EUV or X-ray dimmings (Sterling and Hudson 1997) and EIT waves (Thompson et al. 1998) were recognized as being related to CMEs. In addition there have been efforts to characterize particular pre-eruption active region structures e.g. sigmoidal loops (Rust and Kumar 1996; Canfield et al. 1999), as being likely to originate CMEs. An example of CME-related dimmings associated with the flare and CME events of 12 May 1997 (Attrill et al. 2006) is given

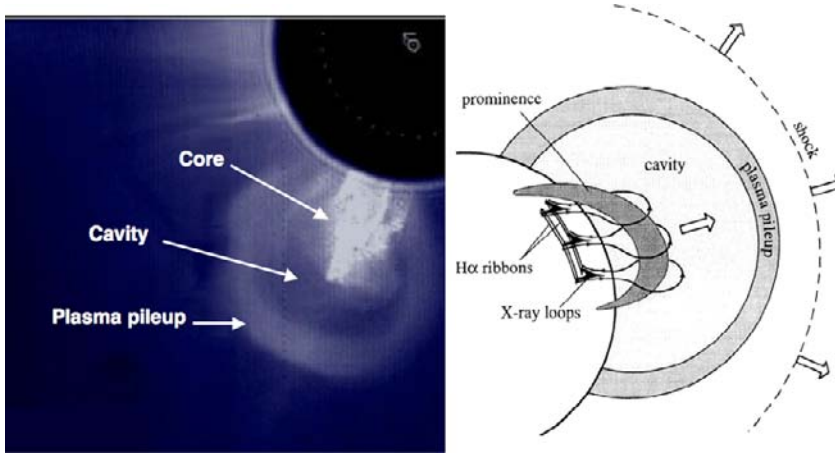


Fig. 3 (*Left panel*) A Solar Maximum Mission archive image showing the principal features of a CME (Hundhausen 1999). (*Right panel*) Schematic view of the CME features (Forbes 2000)

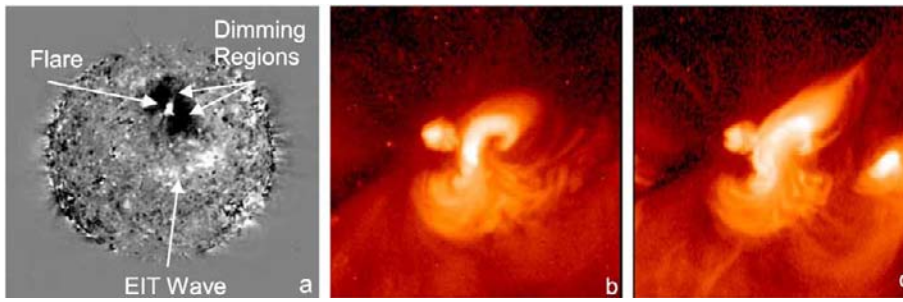


Fig. 4 **a** An EIT difference image showing the flare site, twin dimming regions and the propagating global wave (Attrill et al. 2006). A sigmoid shaped structure associated with the eruptive event of 8/9 June 1998 showing the structure **b** before and **c** after eruption (Glover et al. 2001)

in Fig. 4a. The related global EIT wave is also apparent. A coronal sigmoidal structure eruption is also shown (Figs. 4b, 4c; Glover et al. 2001).

It is generally believed that free energy stored in the magnetic field provides the most likely energy source for these eruptions and that they originate from initially closed non-potential magnetic field that is forced open. Non-potentiality is required as a consequence of the Aly-Sturrock conjecture (Aly 1984, 1991; Sturrock 1991) which asserts that a closed force-free field configuration will always have less energy than the corresponding open field. Magnetic helicity, a quantity that describes the non-potentiality and topological complexity of the magnetic field (Sect. 2.4), is generated in the solar interior and transported to the surface. Magnetic helicity (Berger 1984) is a conserved quantity that emerges with consistent sign; negative in the Northern hemisphere and positive in the Southern, and this pattern does not change with the solar cycle (Pevtsov et al. 2001). Thus helicity accumulates in closed magnetic structures but cannot be eliminated by e.g. flare reconnection, so is probably removed to the IPM by CMEs. The latter also originate in closed structures e.g. active regions, streamers, and are often associated with prominence eruptions.

Eruption models concentrate on explaining how the required energy is stored and how its ultimate release is triggered. Among several reviews of this topic e.g. Forbes (2000), Klimchuk (2001), Zhang and Low (2005), we follow the classification of models suggested by Klimchuk as being i) directly driven or ii) storage and release. Models in category i) include thermal blast where the eruption energy is available from solar flares e.g. Dryer (1982), Wu (1982), and dynamo models e.g. Chen (1989), where rapidly injected magnetic flux further stresses or shears existing field structures. In general, these models cannot easily reproduce the observed features of CMEs or they require unreasonably rapid rates of magnetic flux injection into the corona.

The storage and release models involve energy build up through the stressing of the magnetic field which provides energy to drive the eruption. The involvement of prominences and overlying streamers led to the development of mass loading models e.g. Low and Smith (1993), Low (1996, 1999). Here the already non force-free field of a flux rope is further stressed by the cool prominence mass and the mass of the streamer. Removal of the mass can then lead to the eruption. However not all CMEs have associated prominence mass involved while for those that do, some or all of this mass is often seen to rise as part of the eruption rather than drain away as would be required to unload the stressed configuration. It is also difficult to envisage the conditions of high plasma β and specific coronal mass distribution with high density material overlying low density cavities as are required in the model of Wolfson and Saran (1998). Here again not all CMEs are seen to involve helmet streamers. Thus mass loading models can at best explain only a subset of all CMEs.

Most recently developed models have tended to focus on changes in magnetic structures that are non-potential and therefore have associated free energy that can become available to drive an eruption. Common to all of these models is the progressive build up of free energy and its eventual eruptive release. The structural changes usually involve magnetic reconnection. From the large number of such models that have been developed, we will describe three examples. A model that requires quadrupolar magnetic topology and uses energy stored in sheared arcades—the magnetic breakout model, has been proposed by Antiochos et al. (1999). The progress of energy storage through stressing of a central magnetic arcade and the approach to final eruption are shown in Fig. 5 where four flux systems are involved. Footpoint motions shear the central closed arcade which is inflated by increasing magnetic stress. Reconnection occurs between the expanding arcade field (blue) and the overlying field lines (red). Removal of the overlying field allows further expansion of

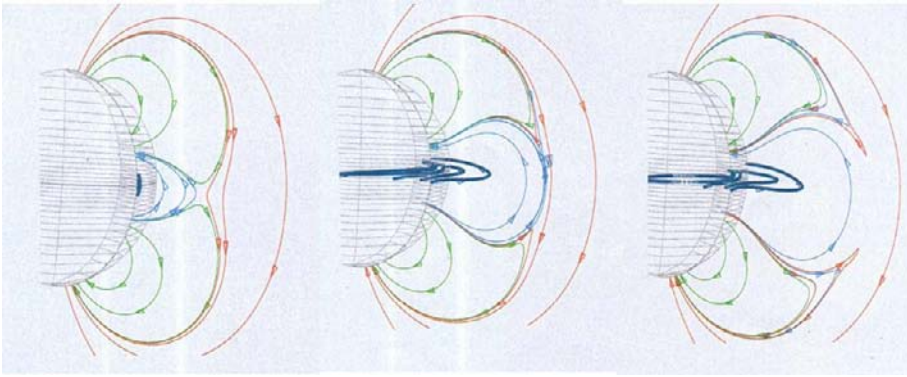


Fig. 5 Stages in the development of a magnetic breakout eruption simulation (Antiochos et al. 1999)

the central sheared arcade which leads rapidly to an eruption driven by the sheared arcade magnetic energy.

Flux cancellation or the mutual disappearance of magnetic fields of opposite polarity at the neutral line separating them (Martin et al. 1985), can lead to sheared field and the creation of a flux rope with associated free energy (van Ballegooijen and Martens 1989). More recent calculations and simulations e.g. Forbes and Isenberg (1991), Forbes et al. (1994), Lin et al. (1998), Linker et al. (2003), have shown that, following the formation of a fluxrope, continued flux cancellations will result in an increasing fluxrope height and a loss of equilibrium. Towards the end of the flux cancellation phase, a vertical current sheet is formed that stretches downwards from the elevated fluxrope. Reconnection at the current sheet allows the eruption to proceed rapidly to completion. This reconnection leads to the formation of a closed loop arcade underneath the erupting fluxrope that grows with time.

The third kind of model for eruptions involves the operation of ideal MHD instabilities e.g. the kink and torus instabilities. The kink instability occurs in a flux rope when the twist exceeds a critical value leading to a helical deformation of the flux rope's axis (Hood and Priest 1981). Using the loop model of Titov and Démoulin (1999) as a starting point, Török et al. (2004) and Török and Kliem (2005) have simulated the kink instability. They use their simulations to reproduce both confined (27 May 2002) and completed (15 May 2001) eruptions and show that a steeper decrease of magnetic field with height in the corona above the flux rope can allow the full eruption to proceed. Fan (2005) has also simulated the kink instability and showed that transient S- or sigmoid-shaped structures can develop during eruption onset similar to those observed in some flares and CMEs e.g. Sterling and Hudson (1997). Williams et al. (2005) have compared an observation of a filament eruption observed by TRACE on 2004 November 10 with the simulation of Török and Kliem (2005). Stages of the eruption are compared with the simulation in Fig. 6 where the qualitative agreement is apparent. However for this event, which takes place in a quadrupolar magnetic configuration, it is likely that elements of both the flux cancellation and breakout models may also have been involved in weakening the restraining fields before the kink instability finally became the principal driver of the event.

A current carrying ring situated in an external poloidal magnetic field (B_{ex}) is unstable against radial expansion when the Lorentz self-force (hoop force) decreases more slowly than the stabilizing Lorentz force due to B_{ex}. Known as the torus instability, its possible role in solar eruptions has been examined by Kliem and Török (2006) while an MHD simulation based on a line tied flux rope (Titov and Démoulin 1999) was done by Török and Kliem (2007). With $B_{ex} \approx R^{-n}$, they establish that $n > 3/2$ is the threshold for instability onset and that with an appropriate starting height for the curved flux rope, the eruption acceleration depends on the steepness of the radial field gradient. Thus CMEs from complex active regions with steep field gradients in the corona are more likely to give rise to fast CMEs—something that is indeed observed.

While there is often a compulsion to establish a single energy storage and release model as being the cause of eruptions or CMEs, given the complexity of the magnetic topologies it would not be surprising if elements of several models were involved. Thus we have seen that separate means for weakening the fields that restrain fluxropes may still play a role in situations where the kink instability emerges as the principal driver of the eruption. For the MHD instabilities, the kink process which is commonly thought more likely to cause confined eruptions, may establish the initial conditions where a complete eruption can continue through the operation of the torus instability. Line tied fluxropes clearly play an important role both for containment of cool prominence material and as erupting flux configurations even in cases where no prominence material is present. For both the magnetic breakout and

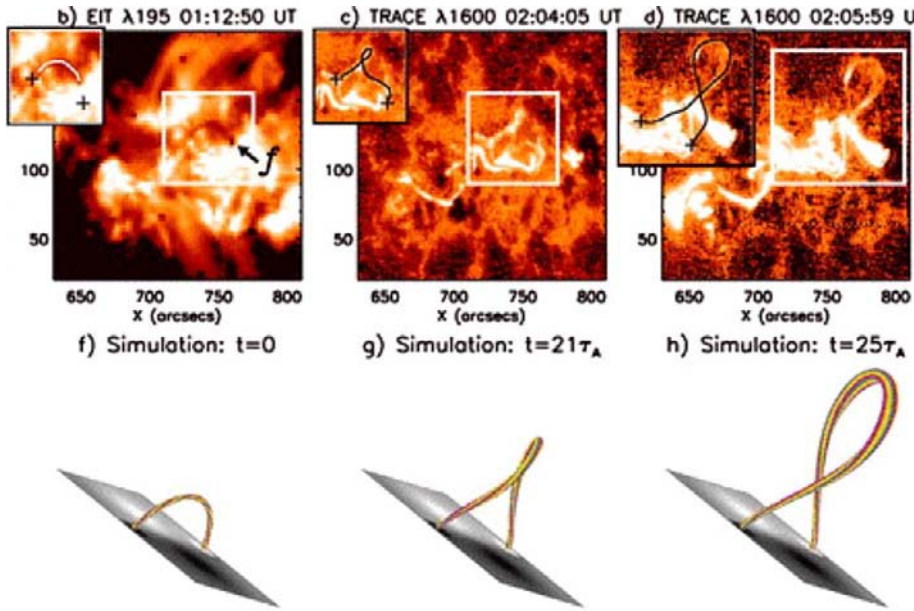


Fig. 6 Panel **b** shows a pre-eruption EIT image while panels **c** and **d** are TRACE images that show the eruption of a heated kinked filament (Williams et al. 2005). Panels **f**, **g** and **h** illustrate the progress of the kink instability from a numerical simulation by Török and Kliem (2005)

flux cancellation models, fluxropes are formed during the energy build up phase or in the course of the eruption. The MHD instabilities could be initiated by adding twist or curvature to a pre-existing fluxrope or to one that had previously emerged. However it is difficult at present to observationally establish the origins of fluxropes later seen in the IPM.

6 Interplanetary Coronal Mass Ejections and Magnetic Cloud Characteristics

Following an eruption or CME on the Sun, plasma and magnetic field expand into interplanetary space behind a propagating shock. The resulting structures are called Interplanetary CMEs (ICMEs). The term Magnetic Cloud (MC), where a more restricted set of characteristics is present, is used for a subset of these cases. Thus for in-situ observations at ≈ 1 AU distance from the Sun, the cloud will exhibit a stronger magnetic field than the surroundings (Hirshberg and Colburn 1969), lower temperature and plasma β values (Gosling et al. 1973) and a smooth rotation of the magnetic field (Klein and Burlaga 1982).

It is important to relate the properties of the MC to those of the original eruption. Relevant parameters for comparison include i) magnetic field direction, ii) magnetic flux, iii) magnetic helicity and iv) plasma composition. At present coronal *magnetic field direction* is usually inferred from extrapolations of photospheric field measurements but in an eruption, field strength and direction can change rapidly (see Fig. 6). A CME is typically not identified in a coronagraph image until ≈ 20 min after its launch. Related surface phenomena e.g. coronal dimming outflows, can however be identified and associated with the footpoints of the erupting structure. This can in turn allow the *magnetic flux* associated with the dimming regions to be estimated but we will see below that the relationship with magnetic cloud flux is not

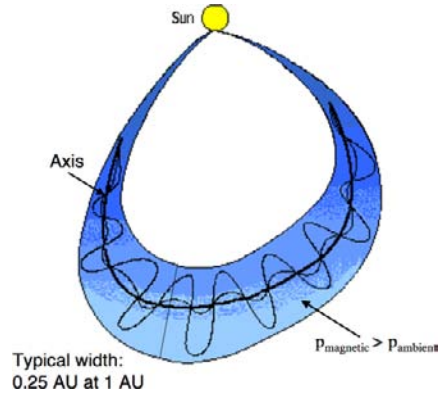


Fig. 7 Schematic diagram of a magnetic cloud showing the geometry and a typical scale. The helical magnetic field lines of the flux rope configuration are indicated. Both ends of the flux rope are connected to the Sun (adapted from Webb et al. 2000)

always simple. *Magnetic helicity* as a conserved quantity preserves its sign and magnitude. Sign can be inferred from vector field measurements at the eruption site or from the orientation of observable structures in the chromosphere and corona. Following the definition of relative helicity (Berger 1984) the change in helicity in the corona following an eruption can be calculated and compared with the value in a MC. It is difficult to determine *plasma composition* in the erupting material for later comparison with that detected at around 1 AU since there are many possibilities for changes to occur in transit. Thus such comparisons tend at present to be made in a qualitative manner.

A schematic diagram of a MC is shown in Fig. 7. The observed rotation suggests that the cloud field configuration is that of a fluxrope which has expanded with the original CME (see Figs. 1a and 1b). In-situ measurement of the field is typically made by magnetometers on a single spacecraft which, for best advantage would pass close to the cloud axis. In such an encounter the direction of the fluxrope axis can be established by a minimum variance analysis. Somewhat better estimates may be obtained by fitting different fluxrope models to the magnetometer data and comparing the results (Dasso et al. 2005). For encounters with high impact parameter, it is necessary to proceed by applying different models. A typical sample of in-situ data from the WIND spacecraft is shown in Fig. 8. The upper panel indicates the sudden increase in plasma velocity that accompanies the arrival of the shock. This is followed by an interval of swept-up solar wind or sheath plasma. Within the cloud, a significantly reduced density is observed in the second panel of the figure while the next three panels show the characteristic magnetic field rotation that characterizes the fluxrope structure. Comparison of cloud axis directions, measured in-situ, with that of the original erupting filament channel or prominence reveals a wide range of behavior. In many cases there is good agreement between these directions (Marubashi 1997; Bothmer and Schwenn 1998) but in others, rotations range from a few tens of degrees (Marubashi 1997; Zhao and Hoeksema 1998) to 130–160 degrees (Rust et al. 2005; Harra et al. 2007). Development of a full kink instability, where magnetic helicity is transformed from twist to writhe may be responsible for the extreme values. The event shown in Fig. 6 provides a possible example of this behavior.

The schematic of Fig. 7 shows both ends of the cloud connected to the Sun. In-situ observation of counter-streaming supra-thermal electrons in a cloud is usually taken to indicate that the cloud or fluxrope is connected at both ends (Richardson et al. 1991; Richardson 1997). The bottom panel of Fig. 8 shows an in-situ electron analyzer spectrogram where the electrons are widely distributed in pitch angle indicating the presence of bi-directional electron streams. Conversely the absence of such electron streams suggests

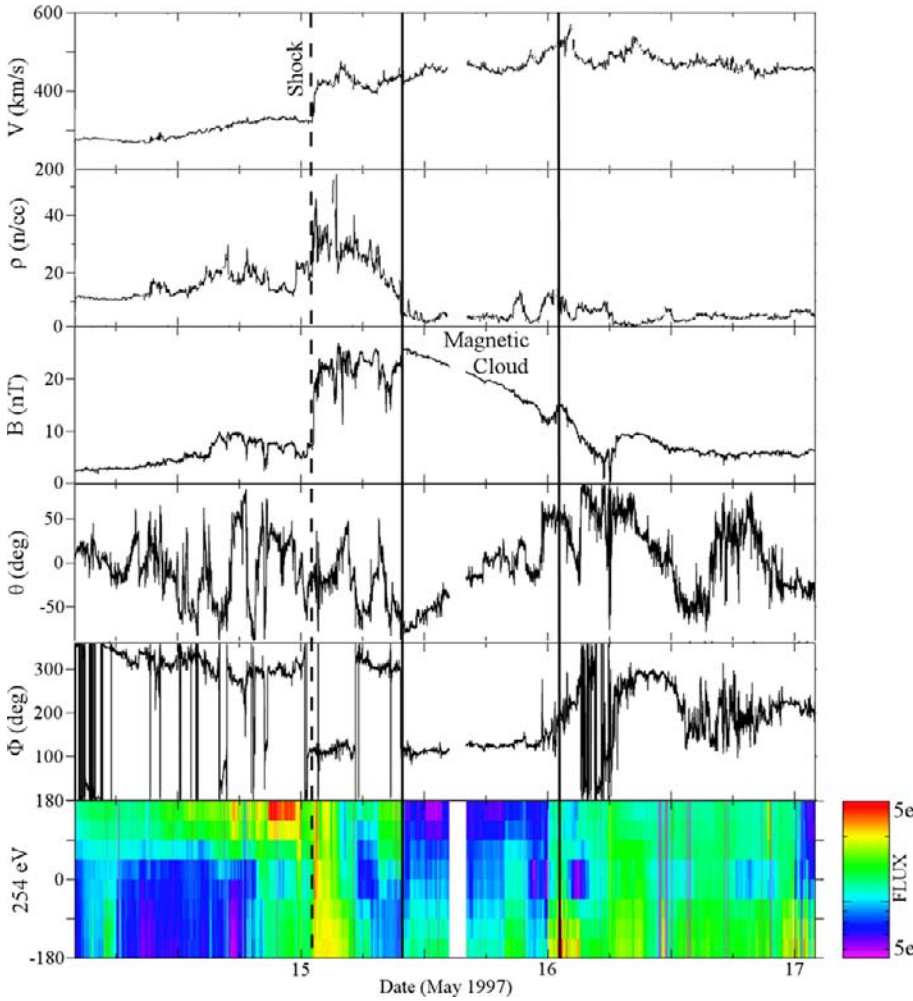


Fig. 8 In-situ observations of a magnetic cloud from the WIND spacecraft. *Panels from the top* give solar wind velocity (V), plasma density (ρ), magnetic field strength (B), elevation (θ) and azimuth (ϕ) of the field direction in solar ecliptic coordinates. The *bottom panel* shows the electron intensity distribution in pitch angle. The broad distribution indicates the presence of bi-directional electron flows. The arrival of the shock and the passage of the cloud are shown by *vertical lines* on the plots

complete disconnection while a uni-directional electron flow points to the cloud being connected to the Sun at one end only. While the flux rope may be initially connected at both ends, its topology may be modified due to reconnection in the corona with e.g. streamer structures, or by reconnection with interplanetary solar wind magnetic field. The connection topology will clearly impact the associated magnetic flux.

Since the usual interpretation of the coronal dimming regions is as sites of material outflow (Hudson et al. 1996; Harra and Sterling 2001) and two prominent regions are often located on both sides of the eruption site, it has frequently been assumed e.g. Webb et al. (2000), that the dimming regions are at the ejected flux rope footpoints. Magnetic flux at the dimming regions may be measured from magnetogram images and compared with the flux

associated with the related cloud. Axial and azimuthal magnetic flux determination usually requires deduction from in-situ magnetometer data by fitting a magnetic model of the cloud (Dasso et al. 2005) where knowledge of MC axis direction and an assumption of the cloud length are also required. Comparisons often show rough agreement between dimming region and cloud axial fluxes where the latter are estimated from near-Earth in-situ data e.g. Lepping et al. (1997), Webb et al. (2000). However structures observed in interplanetary space are often highly twisted and many clouds have substantial azimuthal flux. Twist may be added following reconnection in a sheared arcade overlying an expanding flux rope (Mandrini et al. 2005; Attrill et al. 2006) in a manner that allows the open flux from the dimming regions to contribute to the cloud azimuthal flux component. It is nevertheless not always possible to relate the dimming region flux to that observed in the MC. Mandrini et al. (2007) have studied an eruption where they find no agreement between the MC flux and that of the multiple dimming regions involved. Thus when comparing solar fluxes with those observed in MCs, the magnetic context of dimming regions and their relation to the eruption involved must be considered carefully.

As described in Sect. 2.4, magnetic helicity, H, quantifies how the magnetic field is sheared and twisted compared to its lowest energy state of a potential or current-free field. The value of the helicity content of active regions as a pointer to their activity is discussed in Sect. 4.4. As a conserved quantity it has an important role in comparisons between eruptions and their related MCs. Thus it is increasingly believed that source region helicity is removed from the Sun in CMEs and is found as a measurable quantity in MCs. Using the methods outlined in Sect. 2.4, the helicity content of an AR can be calculated based on magnetic field extrapolation and the change in helicity from before to after the eruption of a CME can be estimated. As was the case for magnetic flux, the helicity of a magnetic cloud can also be estimated by fitting a model to the MC in-situ magnetometer data or in cases of low spacecraft to cloud impact parameter, H may be derived directly from the data (Dasso et al. 2006). Comparison of the helicity change in the source region with the value measured for the cloud provides insight into the eruption physics and offers a useful aid in matching eruption and cloud identities. These methods are increasingly being used in CME/MC studies and we will discuss an example in Sect. 7.

Although an apparently simple picture of the eruption near the Sun is usually presented (see Fig. 3), the determination of the composition and temperature distribution of the plasma involved is by no means straightforward. Assuming that the eruption involves a flux rope topology, with or without contained filament material, the expanding shock can sweep up a range of possible plasmas in the overlying corona. These may include material from active region loops ($T \sim$ 2–5 MK), flare heated plasma ($T \sim$ 20–30 MK), along with streamers and other quiet Sun structures ($T \sim 1$ MK). In cases where filament or prominence material is involved, the temperature would typically be ~ 0.1 MK. However dark filaments, seen in Hα can also "disappear" before eruption with the plasma becoming visible in emission lines with maximum abundance temperature ranging from He II (0.18 MK) to Fe X (1.0 MK) and above (Tripathi et al. 2008). The situation is further complicated by the partial filament eruptions that are frequently observed. CME magnetic structures may also reconnect with structures in the upper solar atmosphere or in the solar wind that have oppositely directed magnetic fields. This will lead to mixing of the original source region plasma with material that did not participate in the eruption. When this happens below the threshold height of a few $R_{\odot}$ at which temperature "freeze-in" occurs (Hundhausen et al. 1968), the plasma ion composition will be altered. The plasma may undergo further selective modification by diffusion across field lines during its passage from the Sun which leads to e.g. an enhanced heavy ion concentration (Wurz et al. 2000).

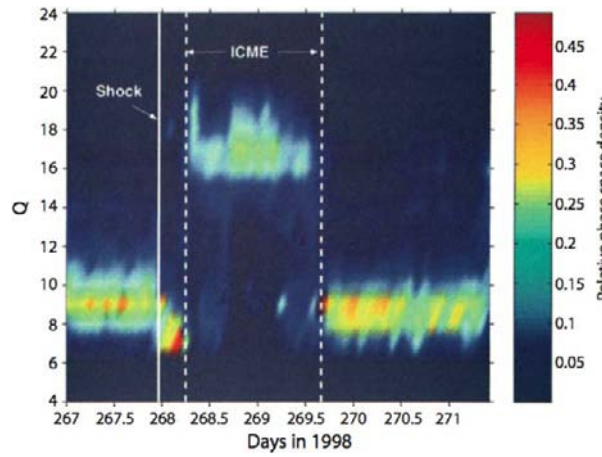


Fig. 9 A broad range of Fe charge states observed from ACE in the period 24–28 September 1998 by Lepri et al. (2001). The shock arrival and ICME passage are indicated by *vertical lines*

Although MCs are identified in only 50% of ICMEs, their magnetic characteristics make it easier to relate in-situ observations to remote observations of the original eruptions. However both ICMEs and MCs exhibit similar composition characteristics. Ion composition and inferred element abundance measurements are made by ion mass spectrometers e.g. the SWICS instruments on the Ulysses and ACE spacecraft and CELIAS on SOHO. In-situ signatures that are of relevance for eruptions at the Sun have been summarized in a review by Zurbuchen and Richardson (2006) for ICMEs and include a) elevated oxygen charge states, $O^{7+}/O^{6+} > 1$; b) average Fe charge state (Q^{Fe}) > 12 or $Fe^{\cdot 16+}/Fe_{Total} > 0.1$; c) detection of He^{+}; $He^{+}/He^{2+} > 0.01$ and d) high $^3He/^4He$; $(^3He/^4He)_{ICME}/(^3He/^4He)_{photosphere} > 2$. Such signatures are taken to indicate the passage of an ICME or MC and they can at least be qualitatively related to the plasma in the original eruption. Thus Richardson and Cane (2004) in an extensive study have found a) to be a reliable ICME indicator. The elevated Oxygen charge states indicate plasma with $T > 2$ MK and since this value is frozen in below 1 $R_{\odot}$, it probably reflects an origin in coronal active region structures overlying the eruption site that have been swept up in the expansion of the CME shock.

For indicator b), Richardson and Cane have also demonstrated an association with ICMEs in approaching 70% of the cases they studied. Lepri et al. (2001), have observed a range of Fe charge states from Fe^{+15} up to Fe^{+19} for an ICME seen at ACE during September, 1998 (Fig. 9). This observation shows clearly how the ionizations stage varies from the front (highest) to the back (lowest) of the MC and indicates freeze-in temperatures in the range 2–8 MK. Given that for Fe ions, freeze-in typically occurs at heights up to 4 $R_{\odot}$, it is clear that the higher temperature material in particular probably originated as 10–20 MK heated plasma from a solar flare associated with the original CME. Assuming a magnetic connection to the CME flux rope or that hot plasma was swept up by the expanding shock, the high Fe stages clearly show a flare-CME association in these cases (Lepri and Zurbuchen 2004).

Indicator c) denotes an enhanced presence of singly charged He. This is seen in a comparatively small number of MCs and suggests the presence of filament material with $T \sim 0.1$ MK (Gosling et al. 1980; Gloeckler et al. 1999). Enhancement of indicator d), suggestive of chromospheric material which would likewise form part of a filament, is observed along with enhancement of He^{+}. Burlaga et al. (1998) observed a MC for which the rear part of the cloud showed a high density along with enhancements of He^{++}, He^{+} and the presence of O^{5+} and Fe^{5+}. This composition suggests freeze-in temperatures of ~0.1–0.4 MK

characteristic of a filament that may have experienced some heating just before or during the eruption. Given the broad range of plasma compositions and temperatures likely to be involved in the original eruption, the in-situ determination of MC composition and ion stage distribution can provide valuable information on the CME process at the Sun and help to constrain the range of eruption models.

7 Coronal Mass Ejections: From Sun to Earth

As discussed in the two sections above, comparison of the original eruption at the Sun with the behavior and properties of the ejecta in interplanetary space can usefully clarify understanding for both phases. Since a significant minority of CMEs reach the Earth and interact with its magnetic and plasma environments, such studies—often described under the general heading of "space weather", are also valuable in clarifying the CME-Earth interaction and the impact on the near-Earth environment.

In seeking to associate the arrival of a shock front and its associated magnetic cloud at Earth with an eruption that would have occurred ∼ two days previously at the Sun, the shock propagation speed is clearly an important parameter. The dynamic spectra of slow-drift type II radio bursts, generally attributed to shock-accelerated electrons (Wild and Smerd 1972), can provide estimates of shock velocity. Since the radio emission is due to local plasma oscillations excited by the passage of the shock where the oscillation frequency is $f_p = 9000\sqrt{n_e}$, the drift rate of the burst to lower frequency can give the shock velocity provided that the electron density of the medium through which the shock propagates is known as a function of distance from the Sun. Frequently used n_e–h models for the corona are those of Newkirk (1961) and Saito (1970) but they predict high n_e values at 1 AU. Since it is important to track interplanetary type II bursts to the neighbourhood of Earth, hybrid density models e.g. Vršnak et al. (2004), are used for connecting bursts in the corona and in interplanetary space. Observations of bursts are made at decimeter–meter wavelengths in the solar atmosphere using ground-based radio telescopes and at decametric–hectometric wavelengths in the interplanetary medium where space-based antennae are required. Uncertainties in the n_e values make it important to assess carefully any speed estimates based on type II burst observations.

Coronagraphs allow measurement of CME speeds in the plane of the sky but in cases where the CME is directed towards Earth—the so-called halo events, determination of the radial velocity is required for estimates of the transit time. It has been pointed out by Dal Lago et al. (2004), that the expansion speed, $V_{\rm exp}$ or the CME lateral growth speed may be determined uniquely for all types of CME. Based on these ideas, Schwenn et al. (2005) have established an empirical relation between the radial and expansion velocities or $V_{\rm rad} = 0.88 V_{\rm exp}$ with a correlation coefficient of 0.86. However the resulting radial velocities apply comparatively close to the Sun. For slightly asymmetric halo CMEs where the launch site is not at sun center, the cone model of Michałek et al. (2003) may be used to estimate radial velocity. The model assumes constant velocity and relies on the time difference between first and last appearance of the CME edges in the coronagraph. The need for asymmetry limits the number of events for which the cone model may be used while the comparatively poor time cadence of current coronagraphs e.g. SOHO LASCO, limits its accuracy.

At distances of ∼20–220 $R_\odot$ from the Sun, interplanetary radio scintillation (IPS) observations which exploit the scattering of radiation from distant radio sources e.g. Quasars, by density irregularities in the solar wind, can be used to assess the density-turbulence condition of the ambient solar wind (Tappin 1986; Manoharan 1993). For moving ICMEs, IPS

measurements can easily detect the excess turbulence produced in the ICME sheath or region of compressed solar wind plasma between the shock and the driving cloud (Manoharan et al. 1995, 2000). When coupled with LASCO observations, the IPS measurements are extremely valuable in establishing speed–distance profiles out to beyond 1 AU. They observe in particular the deceleration of faster CMEs by interaction with the ambient solar wind plasma. However comparatively infrequent sampling—three to four velocity measurements during a CME Sun–Earth transit are typical, limits the applicability of the method.

The in-situ observations by near-Earth spacecraft e.g. ACE, Cluster, can register the arrival of CME-related shocks and their associated material. Arriving shocks are detected as sharp increases in solar wind speed as measured by ion spectrometers (see Fig. 8, top panel). Immediately behind the shock, turbulent sheath or swept-up solar wind material is detected. Finally some five to ten or more hours later, the driving coronal ejecta arrive. The time of shock arrival may be related to the CME launch time in order to deduce an average transit speed assuming that a correct association can be made between shock and CME.

As will be clear from the above comments, the estimation of CME transit speed in the interplanetary medium is not a straightforward matter. Thus a combination of the above approaches will usually be required to achieve a reliable outcome. While obtaining a reliable estimate of CME transit time from Sun to Earth is a necessary part of relating observations of the original eruption to the in-situ identification of the associated magnetic cloud or ICME near-Earth, progress in understanding requires a detailed comparison of the parameters of the eruption e.g. magnetic flux, helicity, as described in Sect. 5 with those later measured for the magnetic cloud in the neighbourhood of Earth. We will now seek to clarify this relationship with reference to some sample events.

Mandrini et al. 2005, have observed an unusual eruption at $\sim 09:00$ UT on 11 May 1998 associated with a small bipolar X-ray bright point. An overview of the launch of the fluxrope and the detection of a small magnetic cloud at Earth over four days later is given in Fig. 10. The MDI magnetogram (Fig. 10a) shows the small bipole located near disc center. Its evolution was followed from 9 May and an apparent rotation of magnetic polarities was probably due to the emergence of a strongly twisted flux tube. This is supported by the observation of a sigmoidal appearance in the coronal structure above the bipole as seen in EIT 284 Å images. The modelled magnetic field in the corona also showed an unusually high degree of non-potentiality. X-ray emission from this small region was observed with Yohkoh SXT (Fig. 10b). Three impulsive events were seen and the third of these, which lasted for three hours, had the largest time-integrated X-ray flux. During this latter event, significant changes occurred in the small coronal structures seen with EIT and a cusp formation was also observed. Dimming regions associated with the third event are shown as contours overlaid on an MDI magnetogram in Fig. 10c. In addition to two concentrated regions close to the bipole, there are extended regions that cover a larger area of quiet Sun. From careful measurement of the net magnetic flux associated with the dimming regions, Mandrini et al. obtained a value of $13 \pm 2 \times 10^{19}$ Mx of which about 8% was contributed by the extended quiet region dimming. They also calculated the change in coronal relative magnetic helicity before and after the event. This was based on a linear force free field model where $\nabla \times B = \alpha B$ and best-fitting α values were found by comparing the extrapolated field with the coronal loop structures as seen in the TRACE images. The resulting helicity change was in the range -3.3×10^{39} Mx$^2 \leq \Delta H_{\text{corona}} \leq -2.3 \times 10^{39}$ Mx^2.

In order to search for a matching in-situ signature, WIND data were examined by Mandrini et al. for an interval two to five days after the small event on 11 May. A small magnetic cloud—probably the smallest ever observed, was registered by the spacecraft in the interval 22:00 UT to 01:50 UT on 16 May. The characteristic smooth rotation of the magnetic field,

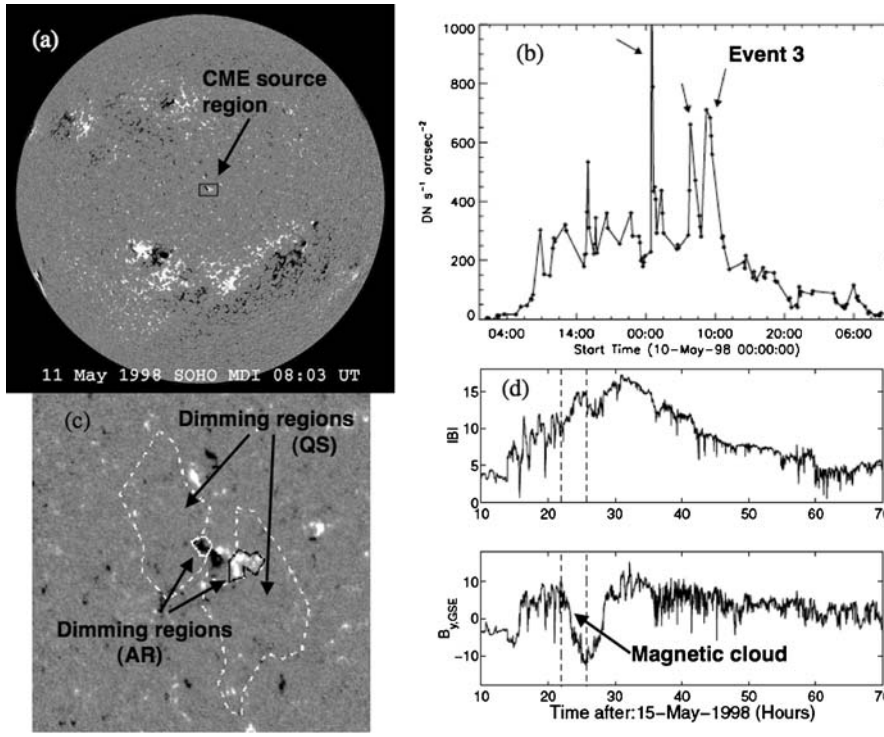


Fig. 10 Observation of a small eruption on 11 May 1998 by Mandrini et al. (2005). The MDI magnetogram (**a**) shows a small bipole near disc center. A Yohkoh SXT X-ray *light curve* (**b**) indicates three small events the last of which relates to the eruption. Associated dimming regions deduced from EIT 195 Å observations are shown in (**c**) while WIND observations on 15 May 1998 of magnetic field for the associated magnetic cloud are in (**d**)

B_{yGSE} shown in the lower panel of Fig. 10d, is consistent with a cylindrical fluxrope crossing the spacecraft. High magnetic intensity and low proton temperature, good indicators of a cloud (Burlaga et al. 1981) were also present. Though no associated CME was observed at the Sun on 11 May, it was presumed that the cloud progressed at the current solar wind speed of 350 ± 50 km s^{-1}. The resulting transit time is 119 ± 17 h while the elapsed time from the relevant solar event to cloud arrival was 110 h. Using the methods described by Dasso et al. (2003), the cloud magnetic field axis direction was determined along with the cloud field and the sign of the twist and helicity. All were consistent with those of the pre-eruption coronal sigmoid structure. Finally the value of magnetic flux ($10–20 \times 10^{19}$ Mx) associated with the cloud was consistent with the net flux associated with the coronal dimming regions while the cloud helicity value (-1.5×10^{39} Mx^2 to -3.0×10^{39} Mx^2) agreed with the pre- to post-eruption helicity change in the corona deduced from the magnetic field extrapolations at the Sun. Thus an unusually good agreement was established between the magnetic properties at the eruption site and those of the corresponding magnetic cloud that was observed near-Earth.

Although the above small event conforms well to the idea that erupting fluxrope footpoints are associated with dimming regions symmetrically located on opposite sides of an active region, this is by no means always the case. The well studied C 1.3 flare event of 12 May 1997 (Webb et al. 2000), with a full-halo CME and an associated magnetic cloud at

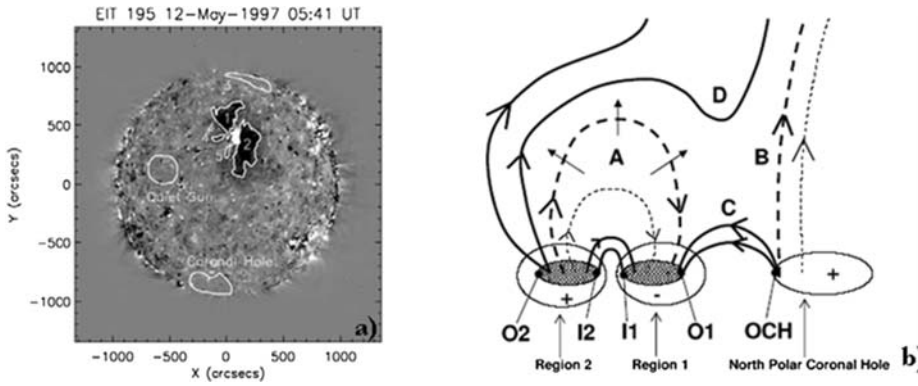


Fig. 11 **a**) EIT 195 Å base difference image at maximum dimming extent. The principal regions are labeled *1* (north) and *2* (south). The North-polar coronal hole is also indicated (Attrill et al. 2006). **b**) Sketch of the evolution of the global magnetic topology for the 12 May 1997 CME and its reconnection with open field lines of the North-polar coronal hole. *Dashed lines* show pre-reconnection magnetic structures, *solid lines* show the post-reconnection fields and *hashed areas* show the main dimming regions. Other symbols are discussed in the text

Earth, was at first thought to conform fully to the twin dimming region scenario. While the magnetic flux associated with the dimming regions was twice that for the related magnetic cloud, only the axial flux component was considered. However the bi-directional electron streaming signature, usually found when both fluxrope legs are connected to the Sun, was not observed.

More recently Attrill et al. (2006), have re-examined this event. The principal dimming regions, labeled 1 and 2 are shown in the EIT 195 Å image of Fig. 11a. From a careful analysis of the time evolution of the two principal dimming regions, Attrill et al have reconstructed the global magnetic topology of the event. A schematic diagram is given in Fig. 11b. While the initial eruption did indeed involve fluxrope connection to regions 1 and 2, the expanding field labeled A reconnected with the open field B of the north polar coronal hole (OCH) to form the magnetic field systems C (closed) by connection to the outer boundary of region 1 (O1) and D (open) which originates from the southern boundary of region 2 (O2). The hot loops that comprise system C are in fact visible in an SXT X-ray image. Progressive reconnection from O1 closes down region 1 which is observed as a shrinking of the region in a plot of dimming recovery. However because of the newly forming open field system D, the southern region 2 remains open for longer. The post-eruption flare loops form between I1 and I2. From the MDI magnetograms, the net magnetic flux from region 2 is $(21 \pm 7) \times 10^{20}$ Mx.

In-situ observations of the related magnetic cloud near-Earth were again obtained by the WIND spacecraft. The observation of uni-directional electron flows suggests that at the time of these observations, the cloud was connected only to the southern region 2. Taking account of the probable time of disconnection from region 1 and of additional path length introduced in the newly opened field from region 2, the cloud length was estimated as 1.3 AU. Fitting magnetic models to the in-situ data as the WIND spacecraft encounters the cloud yields the cloud axis direction which is consistent with the above magnetic topology at the Sun. The total magnetic flux (axial and azimuthal) associated with the cloud is estimated from the best fitting model as $(22 \pm 9) \times 10^{20}$ Mx for the assumed length of 1.3 AU. This figure is in excellent agreement with the southern dimming region flux.

Although the magnetic topology of the 12 May 1997 eruption differs from that of the small eruption of 11 May 1998 that was discussed previously, it was nevertheless possible to relate the magnetic flux associated with a single dimming region with that of the related magnetic cloud. However for the X17 flare and eruption of 28 October 2003, studied by Mandrini et al. (2007), it was not possible to establish any correspondence between the magnetic flux observed in multiple dimming and in the related magnetic cloud. Here the main dimming regions were probably masked by the high flare brightness. In addition the strong lateral expansion of the erupting field reorganized magnetic connectivities which caused the spread of dimming regions over a large part of the Sun (Attrill et al. 2007). Thus the magnetic topology and evolution of each eruption must be studied carefully before any attempt is made to relate the dimming flux and other magnetic properties of the eruption with those of the resulting magnetic cloud. It is clear that such comparisons will be more easily achieved for smaller events than for the eruptions associated with very large flares.

8 Conclusions

The most important characteristic of emerging flux relevant for eruptive activity is that it appears from the solar interior in a non-potential state. Emergence of major concentrated current-carrying (twisted) flux and high photospheric helicity flux show the strongest correlation with major flares and fast CMEs and has therefore the best predictive power. Concentrated current-carrying (high-helicity) flux is characteristic of a flux rope, so big flare and fast CME forecasts provide circumstantial evidence that flux rope emergence plays important role in these activity events. Helicity injection curves in emerging flux regions (e.g. Chae et al. 2004) show a conspicuous peak during the first few days, which greatly resemble the behavior of helicity flux in 3-D MHD simulations of emergence of a twisted flux tube (Cheung et al. 2005). Emergence of a flux rope has many caveats, e.g. dense plasma accumulation in its field lines located under the axis of the flux rope, and a steep gradient in physical parameters leading to strong fragmentation just under the photosphere. A successful emergence must involve many episodes of magnetic reconnection (Pariat et al. 2004). However, characteristic polarity distribution patterns of longitudinal magnetic field in emerging flux regions, the so-called magnetic tongues, indicate that there is an overall organization of the emerging flux tube, which is compatible with a (modest) global twist (Démoulin and Pariat 2008). There are some doubts as to whether or not a flux rope can possibly emerge as an entity. However a weak flux rope emergence may have been seen in Hinode SOT (Tsuneta et al. 2008) vector magnetic data (Okamoto et al. 2008).

During the decay phase of ARs CME activity is maintained (slow CMEs accompanied by small flares, e.g. Démoulin et al. 2002). The AR assumes a simple magnetic configuration, but relatively high current densities indicate an overall flux-rope structure with free magnetic energy stored higher than in an emerging active region (Règnier and Priest 2007). Forecasting methods developed for young ARs would not work for these slow CMEs, however.

It is remarkable that flux ropes are involved in all currently favoured models of CME. The models differ on the nature of the trigger only. Flux emergence and/or flows are implicated in the increase of shear, twist, and complexity, while tether cutting and breakout in changing the overlying field strength. Kink instability can lift the flux rope, facilitating torus instability. There is no consensus among the modellers over the origin of flux rope, whether it pre-exists or forms during the eruption. It seems, however, that during the eruption magnetic reconnections increase its twist (Démoulin 2008; Gibson and Fan 2008). Photospheric

magnetic field typical of pre-eruption conditions seem to favour the pre-existence of a flux rope in the AR.

The CME eruption—and any associated solar flares, represent the last stage in the progression of magnetic flux from the Sun's interior to its outer atmosphere. The need to occult the solar disk poses a problem in searching for associated photospheric, chromospheric and lower coronal phenomena. Nevertheless considerable progress has been made in this area, principally through observations by the Yohkoh, SOHO, TRACE and more recently Hinode and STEREO missions. Use of the SOHO EIT and LASCO instruments has been most relevant but their comparatively low time cadence has led to difficulty. While models involving the storage of free magnetic energy and its later release through a triggered instability are becoming generally accepted as providing the basis for understanding eruptions, the observational difficulties mentioned above make it difficult to establish valid and complete explanations. The situation is further complicated by evidence that different models may be appropriate for different events and, in some cases elements of several models may be involved in a single eruption. While there is good evidence that magnetic flux ropes represent a preferred eruption topology, the manner of their formation for particular events—emergence from below the photosphere, in the corona prior to the eruption or in the later stages of the eruption itself, remains uncertain.

Availability of near-Earth (SOHO, ACE, WIND, Cluster, STEREO) and interplanetary (Ulysses) missions equipped with magnetometers, plasma and particle energy and composition analyzers has allowed the intensive study of the interplanetary consequences of solar eruptions (ICMEs) and of the more tightly defined entities known as Magnetic Clouds (MCs). In many cases a particular event has been registered as it encountered a single near-Earth spacecraft. This requires that a model of the cloud magnetic structure be fit to the data so that the associated magnetic flux and helicity may be deduced. This is done with reasonable reliability except in cases where the spacecraft to cloud impact parameter is large. However the quality of observations is much enhanced when a single structure is observed in-situ with multiple spacecraft. The availability of missions deploying several spacecraft e.g. Cluster, Double Star, STEREO, is making this increasingly possible. Comparisons of the cloud magnetic properties with those of the original erupting material can provide valuable insight into the original process though the situation is complicated by possible magnetic interactions by the cloud during its passage through the solar atmosphere to the in-situ spacecraft. Estimates of cloud transit speed can help to provide verification for associating a particular cloud with its original eruption which will typically have occurred two days earlier. In-situ composition measurements provide another useful basis for comparison between the cloud and the parent eruption and can help in understanding the origins of the latter. Here again interactions of the expanding shock with material in the solar atmosphere can complicate the picture. There is a growing realization that relating in-situ observations to remotely sensed views of the eruption site is valuable for understanding the interaction of solar eruptions with the near-Earth environment. This area, often described by the term "space weather", is assuming increased importance given the possibility that Sun–Earth interactions can result in damage to near-Earth space assets in general and to astronauts outside the Earth's magnetosphere in particular.

While much valuable work has been done by considering the properties of large numbers of events on a statistical basis, ultimate understanding of the physics involved, both at the Sun and in the interplanetary environment requires the detailed examination of individual eruptions both at the Sun and in the near-Earth environment. The complexities involved render this approach a difficult one as the examples addressed in Sect. 7 of this review have demonstrated. The existence of several possible mechanisms that give rise to eruptions at

the Sun presents particular challenges, especially in the case of large events where much of the solar atmosphere may be involved. As the history of solar flare studies has shown there is a danger that preoccupation with individual cases, both at the Sun and near the Earth, can obscure important features of events at both locations. Thus it is essential that as far as possible a common and broadly based approach is pursued for studies at both locations and in the interplanetary medium.

Acknowledgements We thank Kimberley Steed, MSSL Solar and Plasma Groups, for making available Fig. 8 in the paper. LvDG acknowledges Hungarian government grant OTKA T048961. JLC acknowledges the Leverhulme Trust for the award of an Emeritus Fellowship. We are also grateful for the help of the referee who suggested significant improvements to the paper.

References

W.P. Abbett, G.H. Fisher, Y. Fan, Astrophys. J. **546**, 1194–1203 (2001)
J.J. Aly, Astrophys. J. **283**, 349–362 (1984)
J.J. Aly, Astrophys. J. Lett. **375**, L61–L64 (1991)
M.D. Andrews, Sol. Phys. **218**, 261–279 (2003)
S.K. Antiochos, C.R. Devore, J.A. Klimchuk, Astrophys. J. **510**, 485–493 (1999)
G. Attrill, M.S. Nakwacki, L.K. Harra, L. van Driel-Gesztelyi, C.H. Mandrini, S. Dasso, J. Wang, Sol. Phys. **238**, 117–139 (2006)
G.D.R. Attrill, L.K. Harra, L. van Driel-Gesztelyi, P. Démoulin, Astrophys. J. **656**, L101–L104 (2007)
G. Barnes, K.D. Leka, Astrophys. J. **646**, 1303–1318 (2006)
G. Barnes, D.W. Longcope, K.D. Leka, Astrophys. J. **629**, 561–571 (2005)
M.A. Berger, Geophys. Astrophys. Fluid Dyn. **30**, 79–104 (1984)
A. Bleybel, T. Amari, L. van Driel-Gesztelyi, K.D. Leka, Astron. Astrophys. **395**, 685–695 (2002)
V. Bothmer, R. Schwenn, Ann. Geophys. **16**, 1–24 (1998)
A. Brandenburg, K. Subramanian, Phys. Rep. **417**, 1–4 (2005)
M.P. Brouwer, C. Zwaan, Sol. Phys. **129**, 221–246 (1990)
G.E. Brueckner et al., Sol. Phys. **162**, 357–402 (1995)
L.F. Burlaga et al., J. Geophys. Res. **10**, 277–295 (1998)
L. Burlaga, E. Sittler, F. Mariani, R. Schwenn, J. Geophys. Res. **86**, 6673–6684 (1981)
P. Caligari, F. Moreno-Insertis, M. Schüssler, Astrophys. J. **441**, 886–902 (1995)
R.C. Canfield, H.S. Hudson, D.E. McKenzie, Geophys. Res. Lett. **26**, 627–630 (1999)
R.C. Carrington, Mon. Not. R. Astron. Soc. **19**, 1–3 (1858)
J. Chae, Astrophys. J. **560**, L95–L98 (2001)
J. Chae, Y.-J. Moon, Y.D. Park, Sol. Phys. **223**, 39–55 (2004)
J. Chen, Astrophys. J. **338**, 453–470 (1989)
W.-Z. Chen, C. Liu, H. Song, N. Deng, C.-Y. Tan, H. Wang, Chin. J. Astron. Astrophys. **7**(5), 733–742 (2007)
M. Cheung, M. Schüssler, F. Moreno-Insertis, in *Chromospheric and Coronal Magnetic Fields*, ed. by D.E. Innes, A. Lagg, S.A. Solanki. ESA SP, vol. 596 (2005), pp. 541–545
Y. Cui, R. Li, L. Zhang, Y. He, H. Wang, Sol. Phys. **237**, 45–59 (2006)
Y. Cui, R. Li, H. Wang, H. He, Sol. Phys. **242**, 1–8 (2007)
A. Dal Lago et al., Sol. Phys. **222**, 323–328 (2004)
S. Dasso, C.H. Mandrini, P. Démoulin, C.J. Farrugia, J. Geophys. Res. **108**, 1362–1369 (2003)
S. Dasso, C.H. Mandrini, P. Démoulin, M.L. Luoni, A.M. Gulisano, Adv. Space Res. **35**, 711–724 (2005)
S. Dasso, C.H. Mandrini, P. Démoulin, M.L. Luoni, Astron. Astrophys. **455**, 349–359 (2006)
J.-P. Delaboudinière et al., Sol. Phys. **162**, 291–312 (1995)
P. Démoulin, Adv. Space Res. **39**, 1674–1693 (2007)
P. Démoulin, Ann. Geophys. **26**, 3113–3125 (2008)
P. Démoulin, M.A. Berger, Sol. Phys. **215**, 203–215 (2003)
P. Démoulin, E. Pariat, Adv. Space Res. (2008, in press)
P. Démoulin, C.H. Mandrini, L. van Driel-Gesztelyi, B.J. Thompson, S. Plunkett, Zs. Kövári, G. Aulanier, A. Young, Astron. Astrophys. **382**, 650–665 (2002)
N. Deng, Y. Xu, G. Yang, W. Cao, C. Liu et al., Astrophys. J. **644**, 1278–1291 (2006)
M. Dikpati, P.A. Gilman, Astrophys. J. **635**, L193–L196 (2005)
M. Dryer, Space Sci. Rev. **33**, 233–275 (1982)
M. Dryer et al., Sol. Phys. **181**, 159–183 (1998)

D.A. Falconer, J. Geophys. Res. **106**, 25185–25190 (2001)
D.A. Falconer, R.L. Moore, G.A. Gary, Astrophys. J. **569**, 1016–1025 (2002)
D.A. Falconer, R.L. Moore, G.A. Gary, J. Geophys. Res. **108**, 1380 (2003). CiteID 1380
D.A. Falconer, R.L. Moore, G.A. Gary, Astrophys. J. **644**, 1258–1272 (2006)
Y. Fan, Astrophys. J. **630**, 543–551 (2005)
Y. Fan, Astrophys. J. **676**, 680–697 (2008)
Y. Fan, G.H. Fisher, E.E. DeLuca, Astrophys. J. **405**, 390–401 (1993)
G.H. Fisher, Y. Fan, R.F. Howard, Astrophys. J. **438**, 463–471 (1995)
T.G. Forbes, J. Geophys. Res. **105**, 23153–23166 (2000)
T.G. Forbes, P.A. Isenberg, Astrophys. J. **373**, 294–307 (1991)
T.G. Forbes, E.R. Priest, P.A. Isenberg, Sol. Phys. **150**, 245–266 (1994)
M.K. Georgoulis, Geophys. Res. Lett. **35**, 6 (2008). CiteID L06S02
M.K. Georgoulis, B.J. LaBonte, Astrophys. J. **671**, 1034–1050 (2007)
M.K. Georgoulis, D.M. Rust, Astrophys. J. **661**, L109–L112 (2007)
S.E. Gibson, Y. Fan, J. Geophys. Res. **113**, A09103 (2008)
P.A. Gilman, Astron. Nachr. **326**, 208–217 (2005)
P.A. Gilman, M. Dikpati, Astrophys. J. **528**, 552–572 (2000)
A. Glover, L.K. Harra, S.A. Matthews, K. Hori, J.L. Culhane, Astron. Astrophys. **378**, 239–246 (2001)
G. Gloeckler et al., Geophys. Res. Lett. **26**, 157–160 (1999)
J.T. Gosling, V. Pizzo, S.J. Bame, J. Geophys. Res. **78**, 2001–2009 (1973)
J.T. Gosling, J.R. Asbridge, S.J. Bame, W.C. Feldman, G. Paschmann, N. Sckopke, J. Gepohys. Res. **85**, 744–752 (1980)
L.M. Green, M.C. López-Fuentes, C.H. Mandrini, P. Démoulin, L. van Driel-Gesztelyi, J.L. Culhane, Sol. Phys. **208**, 43–68 (2002a)
L.M. Green, S.A. Matthews, L. van Driel-Gesztelyi, L.K. Harra, J.L. Culhane, Sol. Phys. **205**, 325 (2002b)
L.M. Green, B. Kliem, T. Török, L. van Driel-Gesztelyi, G.D.R. Attrill, Sol. Phys. **246**, 365–391 (2007)
H.J. Hagenaar, R.A. Shine, Astrophys. J. **635**, 659–669 (2005)
H.J. Hagenaar, C.J. Schrijver, A.M. Title, Astrophys. J. **584**, 1107–1119 (2003)
G.E. Hale, Astrophys. J. **28**, 315–343 (1908)
G.E. Hale, S.B. Nicholson, Astrophys. J. **62**, 270–300 (1925)
G.E. Hale, F. Ellerman, S.B. Nicholson, A.H. Joy, Astrophys. J. **49**, 153–178 (1919)
L. Harra, A. Sterling, Astrophys. J. **561**, L215–L218 (2001)
L. Harra et al., Sol. Phys. **244**, 95–114 (2007)
K.L. Harvey, PhD Thesis, University of Utrecht (1993)
K. Harvey, J. Harvey, Sol. Phys. **28**, 61–71 (1973)
K.L. Harvey, C. Zwaan, Sol. Phys. **148**, 85–118 (1993)
D.H. Hathaway, D.P. Choudhary, Sol. Phys. **250**, 269–278 (2008)
J. Hirshberg, D.S. Colburn, Planet. Space Sci. **17**, 1183–1206 (1969)
A.W. Hood, E.R. Priest, Geophys. Astrophys. Fluid Dyn. **17**, 297–318 (1981)
H.S. Hudson, L.W. Acton, S.L. Freeland, Astrophys. J. **470**, 629–635 (1996)
A.J. Hundhausen, H.E. Gilbert, S.J. Bame, Astrophys. J. **152**, L3–L7 (1968)
A.J. Hundhausen, in *The Many Faces of the Sun*, ed. by K.T. Strong, J.L.R. Saba, B.M. Haisch, J.T. Schmelz (Springer, New York, 1999), pp. 143–200
H. Jeong, J. Chae, Astrophys. J. **671**, 1022–1033 (2007)
L.W. Klein, L.F. Burlaga, J. Geophys. Res. **87**, 613–624 (1982)
B. Kliem, T. Török, Phys. Rev. Lett. **96**, 255002–255006 (2006)
J.A. Klimchuk, in *Space Weather*, ed. by P. Song, H.J. Singer, J.L. Siscoe (2001), pp. 143–157
B.J. LaBonte, M.K. Georgoulis, D.M. Rust, Astrophys. J. **671**, 955–963 (2007)
K.D. Leka, G. Barnes, Astrophys. J. **595**, 1277–1295 (2003a)
K.D. Leka, G. Barnes, Astrophys. J. **595**, 1296–1306 (2003b)
K.D. Leka, G. Barnes, Astrophys. J. **656**, 1173–1186 (2007)
K.D. Leka, R.C. Canfield, A.N. McClymont, L. van Driel-Gesztelyi, Astrophys. J. **462**, 547–560 (1996)
R.P. Lepping et al., J. Geophys. Res. **102**, 14049–14064 (1997)
S.T. Lepri, T.H. Zurbuchen, J. Geophys. Res. **109**, A01112–A01124 (2004)
S.T. Lepri, T.H. Zurbuchen, L.A. Fisk, I.G. Richardson, H.V. Cane, G. Gloeckler, J. Geophys. Res. **106**, 29231–29238 (2001)
J. Lin, T.G. Forbes, P.A. Isenberg, P. Démoulin, Sol. Phys. **150**, 245–266 (1998)
J.A. Linker, Z. Mikić, P. Riley, R. Lionello, D. Odstrcil, Phys. Plasmas **10**, 1971–1978 (2003)
B.W. Lites, K.D. Leka, A. Skumanich, V. Martínez Pillet, T. Shimizu, Astrophys. J. **460**, 1019–1026 (1996)
B. Lites, H. Socas-Navarro, M. Kubo, T.E. Berger, Z. Frank, R.A. Shine, T.D. Tarbell, A.M. Title, K. Ichimoto, Y. Katsukawa, S. Tsuneta, Y. Suematsu, T. Shimizu, S. Nagata, Publ. Astron. Soc. Jpn. **59**, S571–S576 (2007)

Y. Liu, Astrophys. J. **679**, L151–L154 (2008)
D.W. Longcope, A.R. Choudhuri, Sol. Phys. **205**, 63–92 (2002)
D.W. Longcope, G.H. Fisher, Astrophys. J. **458**, 380–390 (1996)
D.W. Longcope, G.H. Fisher, S. Arendt, Astrophys. J. **464**, 999–1011 (1996)
D.W. Longcope, M.G. Linton, A.A. Pevtsov, G.H. Fisher, I. Klapper, in *Magnetic Helicity in Space and Laboratory Plasmas*, ed. by M.R. Brown, R.C. Canfield, A.A. Pevtsov. Geophys. Monograph, vol. 111 (AGU, Washington, 1999), p. 93
M. López-Fuentes, P. Démoulin, C.H. Mandrini, L. van Driel-Gesztelyi, Astrophys. J. **544**, 540–549 (2000)
M. López-Fuentes, P. Démoulin, C.H. Mandrini, A.A. Pevtsov, L. van Driel-Gesztelyi, Astron. Astrophys. **397**, 305–318 (2003)
B.C. Low, Sol. Phys. **167**, 216–265 (1996)
B.C. Low, Geophys. Monograph. **99**, 39–48 (1997)
B.C. Low, in *Solar Wind Nine*, *Proceedings of the Ninth International Solar Wind Conference*, ed. by S.R. Habbal, R. Esser, J.V. Hollweg, P.A. Isenberg. AIP Conf. Proc., vol. 471 (1999), p. 109
B.C. Low, D.F. Smith, Astrophys. J. **410**, 412–415 (1993)
M.L. Luoni, C.H. Mandrini, S. Dasso, L. van Driel-Gesztelyi, P. Démoulin, Tracing magnetic helicity from the solar corona to the interplanetary space. J. Atmos. Sol.–Terr. Phys. **67**, 1734–1743 (2005)
R.M. MacQueen et al., Astrophys. J. **187**, L85–L89 (1974)
R.M. MacQueen et al., Sol. Phys. **65**, 91–107 (1980)
T. Magara, A model for dynamic evolution of emerging magnetic fields in the Sun. Astrophys. J. **605**, 480–492 (2004)
W. Manchester, T. Gombosi IV, D. DeZeeuw, Y. Fan, Eruption of a buoyantly emerging magnetic flux rope 2004. Astrophys. J. **610**, 588–596 (2004)
A.N. McClymont, G.H. Fisher, in *Solar System Plasma Physics*, ed. by J.H. Waite, Jr., J.L. Burch, R.L. Moore (1989), pp. 219–228
C.H. Mandrini, S. Pohjolainen, S. Dasso, L.M. Green, P. Démoulin, L. van Driel-Gesztelyi, C. Copperwheat, C. Foley, Astron. Astrophys. **434**, 725–740 (2005)
C.H. Mandrini, M.S. Nakwacki, G. Attrill, L. van Driel-Gesztelyi, P. Démoulin, S. Dasso, H. Elliott, Sol. Phys. **244**, 25–43 (2007)
P.K. Manoharan, Sol. Phys. **148**, 153–167 (1993)
P.K. Manoharan, M. Kojima, N. Gopalswamy, T. Kondo, Z. Smith, Astrophys. J. **530**, 1061–1070 (2000)
P.K. Manoharan, S. Ananthakrishnan, M. Dryer, T.R. Detman, H. Leinbach, M. Kojima, T. Watanabe, J. Kahn, Sol. Phys. **156**, 377–393 (1995)
S.F. Martin, S.H.B. Livi, J. Wang, Aust. J. Phys. **38**, 929–959 (1985)
K. Marubashi, in *Coronal Mass Ejections*, ed. by N. Crooker, J.A. Joselyn, J. Feynmann, AGU Geophys. Monograph, vol. 99 (1997), p. 147
D.B. Melrose, Astrophys. J. **387**, 403–413 (1992)
N. Meunier, Astron. Astrophys. **405**, 1107–1120 (2003)
G. Michałek, N. Gopalswamy, S. Yashiro, Astrophys. J. **584**, 472–478 (2003)
F. Moreno-Insertis, T. Emonet, Astrophys. J. **472**, L53–L56 (1996)
F. Moreno-Insertis, P. Caligari, M. Schüssler, Sol. Phys. **153**, 449–452 (1994)
G.E. Moreton, H.E. Ramsey, Publ. Astron. Soc. Pac. **72**, 357–359 (1960)
M.J. Murray, A.W. Hood, Astron. Astrophys. **479**, 567–577 (2008)
G. Newkirk, Astrophys. J. **133**, 983–1013 (1961)
A. Nindos, M.D. Andrews, Astrophys. J. **616**, L175–L178 (2004)
T.J. Okamoto et al., Astrophys. J. **673**, L215–L218 (2008)
E. Pariat, G. Aulanier, B. Schmieder, M.K. Georgoulis, D.M. Rust, P.N. Bernasconi, Astrophys. J. **614**, 1099–1112 (2004)
E. Pariat, P. Démoulin, M.A. Berger, Astron. Astrophys. **439**, 1191–1203 (2005)
E. Pariat, A. Nindos, P. Démoulin, M.A. Berger, Astron. Astrophys. **452**, 623–630 (2006)
E.N. Parker, Astrophys. J. **408**, 707–719 (1993)
A.A. Pevtsov, Astrophys. J. **531**, 553–560 (2000)
A.A. Pevtsov, K.S. Balasubramaniam, Adv. Space Res. **32**, 1867–1874 (2003)
A.A. Pevtsov, R.C. Canfield, S.M. Latushko, Astrophys. J. **549**, L261–L263 (2001)
S. Règnier, E.R. Priest, Astron. Astrophys. **468**, 701–709 (2007)
I.G. Richardson, in *Coronal Mass Ejections*, ed. by N. Crooker, J.A. Joselyn, J. Feynmann, AGU Geophys. Monograph, vol. 99 (1997), p. 189
I.G. Richardson, H.V. Cane, J. Geophys. Res. **109**, A09104–A09120 (2004)
I.G. Richardson, H.V. Cane, T.T. von Rosenvinge, J. Geophys. Res. **96**, 7853–7860 (1991)
D.M. Rust, Geophys. Res. Lett. **21**, 241–244 (1994)
D.M. Rust, A. Kumar, Astrophys. J. Lett. **464**, L199–L203 (1996)

D.M. Rust, B.J. Anderson, M.D. Andrews, M.H. Acuña, C.T. Russell, P.W. Schuck, T. Mulligan, Astrophys. J. **621**, 524–536 (2005)
M. Ryutova, H. Hagenaar, Sol. Phys. **246**, 281–294 (2007)
A. Sainz Dalda, V. Martínez Pillet, Astrophys. J. **632**, 1176–1183 (2005)
K. Saito, Ann. Tokyo Astron. Obs. **12**, 53–120 (1970)
H.U. Schmidt, in *Structure and Development of Solar Active Regions*, ed. by K.O. Keipenheuer. IAU Symp., vol. 35 (1968), pp. 95–107
C.J. Schrijver, Astrophys. J. **655**, L117–L120 (2007)
C.J. Schrijver, C. Zwaan, *Solar and Stellar Magnetic Activity*. Cambridge Astrophysics Series, vol. 34 (Cambridge University Press, New York, 2000)
C.J. Schrijver, M.L. DeRosa, A.M. Title, T.R. Metcalf, Astrophys. J. **628**, 501–513 (2005)
C.J. Schrijver, M.L. DeRosa, T. Metcalf, G. Barnes, B. Lites et al., Astrophys. J. **675**, 1637–1644 (2008)
M. Schüssler, Astron. Astrophys. **71**, 79–91 (1979)
R. Schwenn, A. Dal Lago, E. Huttunen, W.D. Gonzalez, Ann. Geophys. **23**, 1033–1059 (2005)
A. Sterling, H.S. Hudson, Astrophys. J. Lett. **491**, L55–L59 (1997)
P.A. Sturrock, Astrophys. J. **380**, 655–659 (1991)
J.J. Sudol, J.W. Harvey, Astrophys. J. **635**, 647–658 (2005)
S.J. Tappin, Planet. Space Sci. **34**, 93–97 (1986)
B.J. Thompson, S.P. Plunkett, J.B. Gurman, J.S. Newmark, O.C. StCyr, D.J. Michels, Geophys. Res. Lett. **25**, 2465–2468 (1998)
L. Tian, D. Alexander, Astrophys. J. **673**, 532–543 (2008)
V.S. Titov, P. Démoulin, Astron. Astrophys. **351**, 707–720 (1999)
T. Török, B. Kliem, Astrophys. J. **630**, L97–L100 (2005)
T. Török, B. Kliem, Astron. Nachr. **328**, 743–746 (2007)
T. Török, B. Kliem, V.S. Titov, Astron. Astrophys. **413**, L27–L30 (2004)
D. Tripathi, S.E. Gibson, J. Qiu, L. Fletcher, R. Liu, H. Gilbert, H.E. Mason, Astron. Astrophys. (2008, submitted)
S. Tsuneta et al., Sol. Phys. **249**, 167–196 (2008)
A.A. van Ballegooijen, P.C.H. Martens, Astrophys. J. **343**, 971–984 (1989)
L. van Driel-Gesztelyi, in *Three-Dimensional Structure of Solar Active Regions*, ed. by C.E. Alissandrakis, B. Schmieder. ASP Conf. Ser., vol. 155 (1998), pp. 202–223
L. van Driel-Gesztelyi, K. Petrovay, Sol. Phys. **126**, 285–298 (1990)
L. van Driel-Gesztelyi, C.H. Mandrini, B. Thompson, S. Plunkett, G. Aulanier, P. Démoulin, B. Schmieder, C. de Forest, in *Third Advances in Solar Physics Euroconference: Magnetic Fields and Oscillations*, ed. by B. Schmieder, A. Hofmann, J. Staude, vol. 184 (1999), pp. 302–306
L. van Driel-Gesztelyi, P. Démoulin, C.H. Mandrini, L. Harra, J.A. Klimchuk, Astrophys. J. **586**, 579–591 (2003)
B. Vršnak, J. Magdalenić, P. Zlobec, Astron. Astrophys. **413**, 753–763 (2004)
J. Wang, Sol. Phys. **163**, 319–325 (1996)
H. Wang, H. Zirin, G. Ai, Sol. Phys. **131**, 53–68 (1991)
H. Wang, H. Sung, J. Jing, V. Yurchishyn, Y.-Y. Deng, H.-Q. Zhang, D. Falconer, J. Li, Chin. J. Astron. Astrophys. **6**, 477–488 (2006)
D.F. Webb et al., J. Geophys. Res. **105**, 27251–27260 (2000)
J.P. Wild, L.L. McCready, Aust. J. Sci. Res. A **3**, 387–398 (1950)
J.P. Wild, S.F. Smerd, Ann. Rev. Astron. Astrophys. **10**, 159–156 (1972)
D.R. Williams, T. Török, P. Démoulin, L. van Driel-Gesztelyi, B. Kliem, Astrophys. J. **628**, L163–L166 (2005)
R. Wolfson, S. Saran, Astrophys. J. **499**, 496–503 (1998)
S.T. Wu, Space Sci. Rev. **32**, 115–129 (1982)
P. Wurz, P. Bochsler, M.A. Lee, J. Geophys. Res. **105**, 27239–27250 (2000)
M. Zhang, B.C. Low, Ann. Rev. Astron. Astrophys. **43**, 103–137 (2005)
X.P. Zhao, J.T. Hoeksema, J. Geophys. Res. **103**, 2077–2083 (1998)
T.H. Zurbuchen, I.G. Richardson, Space Sci. Rev. **123**, 31–43 (2006)

Coronal Holes and Open Magnetic Flux

Y.-M. Wang

Originally published in the journal Space Science Reviews, Volume 144, Nos 1–4, 383–399.
DOI: 10.1007/s11214-008-9434-0

Abstract Coronal holes are low-density regions of the corona which appear dark in X-rays and which contain "open" magnetic flux, along which plasma escapes into the heliosphere. Like the rest of the Sun's large-scale field, the open flux originates in active regions but is subsequently redistributed over the solar surface by transport processes, eventually forming the polar coronal holes. The total open flux and radial interplanetary field component vary roughly as the Sun's total dipole strength, which tends to peak a few years after sunspot maximum. An inverse correlation exists between the rate of flux-tube expansion in coronal holes and the solar wind speed at 1 AU. In the rapidly diverging fields present at the polar hole boundaries and near active regions, the bulk of the heating occurs at low heights, leading to an increase in the mass flux density at the Sun and a decrease in the asymptotic wind speed. The quasi-rigid rotation of coronal holes is maintained by continual footpoint exchanges between open and closed field lines, with the reconnection taking place at the streamer cusps. At much lower heights within the hole interiors, "interchange reconnection" between small bipoles and the overlying open flux also gives rise to coronal jets and polar plumes.

Keywords Coronal holes · Open magnetic flux · Solar wind · Photospheric flux transport · Coronal flux-tube expansion · Rigid rotation · Magnetic reconnection

1 Basic Concepts

Coronal holes are predominantly unipolar areas of the Sun where the magnetic field extends outward to form the interplanetary magnetic field (IMF) and plasma escapes to form the solar wind (see Zirker 1977; Cranmer 2002; McComas et al. 2007; Zurbuchen 2007). The approximate locations of these low-density coronal regions can be determined from the observed photospheric field by means of a potential-field source-surface (PFSS) extrapolation. In this empirically well-tested model, the corona is assumed to remain current-free

Y.-M. Wang (✉)
Code 7672W, Space Science Division, Naval Research Laboratory, Washington, DC 20375-5352, USA
e-mail: ywang@yucca.nrl.navy.mil

M.J. Thompson et al. (eds.), *The Origin and Dynamics of Solar Magnetism*.
DOI: 10.1007/978-1-4419-0239-9_19

out to a spherical "source surface" at heliocentric distance $r = R_{ss} \sim 2.5\ R_\odot$, where the tangential field components are set to zero, mimicking the magnetohydrodynamic effect of the plasma pressure as it overcomes the restraining magnetic tension (Schatten et al. 1969). All field lines that cross this surface are considered to be open. At the inner boundary, the radial component of the potential field is matched to the photospheric field, which is taken to be radially oriented at the depth where it is measured. [Note that it is incorrect to match the line-of-sight components directly to each other, as done by Altschuler and Newkirk (1969), because the photospheric field, unlike the coronal field, is highly nonpotential.] That this simple prescription is able to reproduce the global configuration of X-ray and He I 1083.0 nm coronal holes at all phases of the sunspot cycle (Levine 1982; Wang et al. 1996; Neugebauer et al. 1998; Luhmann et al. 2002; Schrijver and DeRosa 2003) has important physical implications:

(1) Because higher order multipoles fall off rapidly with height, the main contribution to the source surface field and the open flux comes from the $l = 1$ and $l = 2$ (i.e., the dipole and, at sunspot maximum, the quadrupole) components of the photospheric field.
(2) Being composed of low-order multipoles, the coronal field (including coronal holes) must rotate more rigidly than the photospheric field, which is dominated by high-order multipoles.
(3) That the observed coronal-hole areas can be reproduced with R_{ss} fixed at $\sim 2.5\ R_\odot$ throughout the solar cycle implies that the plasma pressure p, and hence the coronal heating rate, must be a function of the magnetic field strength B. If the heating were instead described by a simple adiabatic law $p \propto \rho^\gamma$ (where ρ is the plasma density and γ the ratio of specific heats), the point where $p \sim B^2/8\pi$ would move outward and the polar holes would contract as the Sun's dipole strength increases, contrary to observations.

The PFSS model breaks down beyond $r \sim 2.5\ R_\odot$ because it omits the effect of transverse pressure gradients (Suess and Nerney 1975). The induced heliospheric sheet currents act to redistribute the open flux until it becomes independent of heliographic latitude L and longitude ϕ (Schatten 1971), in agreement with *Ulysses* magnetometer measurements (Balogh et al. 1995; Smith et al. 2001). To derive the radial IMF strength, we simply divide the total open flux Φ_{open} by $4\pi r^2$, where Φ_{open} is obtained by integrating $|B_r|$ over the source surface (Wang and Sheeley 1995; Lockwood et al. 1999). Thus, at $r = r_E = 1$ AU,

$$B_E = \frac{\Phi_{open}}{4\pi r_E^2}. \tag{1}$$

Because coronal holes occupy only a small fraction of the solar surface, the open flux initially tends to diverge very superradially, even when averaged over a supergranular area. The factor by which a given coronal flux tube expands in solid angle between its footpoint (taken to be just above the chromospheric canopy) and the source surface is given by

$$f_{ss} = \left(\frac{R_\odot}{R_{ss}}\right)^2 \frac{B_0}{B_{ss}}, \tag{2}$$

where B_0 and B_{ss} denote the field strengths at $r \simeq R_\odot$ and $r = 2.5\ R_\odot$, respectively. Empirically, it is found that the solar wind speed at 1 AU is inversely correlated with f_{ss} (Levine et al. 1977; Wang and Sheeley 1990a; Arge and Pizzo 2000): the more slowly the flux tube diverges in the corona, the higher the final wind speed [contrary to what is sometimes inferred from studies such as those of Munro and Jackson (1977) and Kopp and Holzer (1976)]. The

expansion factor is *relatively* small (but greater than unity) near the centers of large coronal holes, but diverges rapidly near the boundaries between opposite-polarity holes, where $B_{ss} \to 0$. It should be emphasized that, because the field lines do not actually become radial until $r \sim 5$–$10\ R_\odot$, f_{ss} does not represent the net expansion undergone by a flux tube between the Sun and 1 AU (which shows a quite different behavior), but only the expansion out to $r \sim 2.5\ R_\odot$.

2 Solar Cycle Variation of the Open Flux

We now derive some general properties of open field regions over the solar cycle by applying the PFSS model to the photospheric field, in the form of 27.3 day synoptic maps from the Mount Wilson Observatory (MWO) and the Wilcox Solar Observatory (WSO). To correct for the saturation of the Fe I 525.0 nm line profile, we have multiplied the magnetograph measurements by the latitude-dependent factor $(4.5 - 2.5 \sin^2 L)$ (Wang and Sheeley 1995; Ulrich et al. 2002; Arge et al. 2002).

Figure 1 compares the evolution of the total open flux during 1967–2008 with spacecraft measurements of the radial IMF strength at Earth. The two curves are in rough agreement, both varying by a factor of order 2 during each of the last three sunspot cycles. The open flux and IMF strength tend to peak ∼2–3 yr after sunspot maximum, when the Sun's total dipole strength is greatest. Also noteworthy is the weakness of the field during 2007–2008 (∼1.5 nT) compared to its average level during previous sunspot minima (∼2 nT).

As shown in Fig. 2(a), the total area occupied by open flux decreases from ∼20% of the solar surface near sunspot minimum to only ∼5% at sunspot maximum; at the same time, however, the average footpoint field strength in coronal holes increases from ∼5 G at

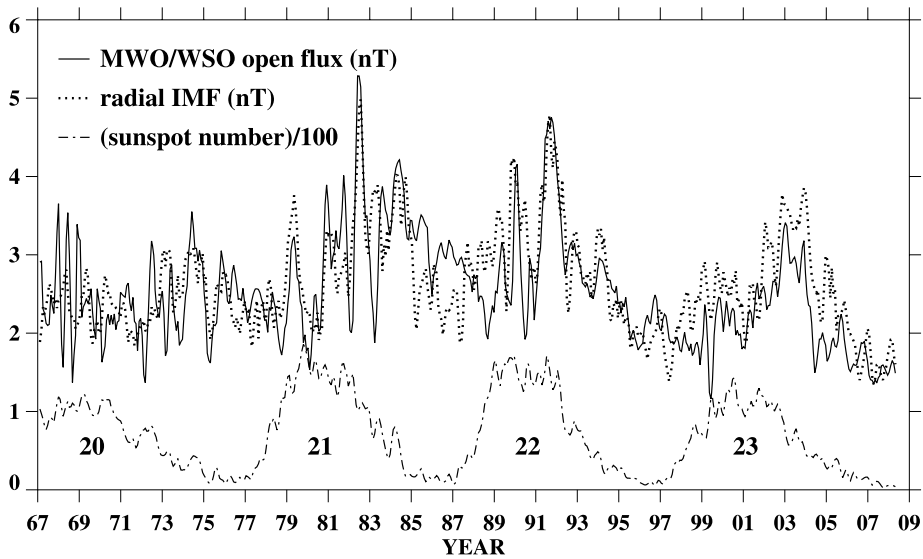


Fig. 1 Comparison between the near-Earth radial IMF strength (NSSDC OMNI 2 data) and the total open flux Φ_{open}, derived from a PFSS extrapolation of MWO and WSO magnetograph measurements and expressed as an equivalent field strength (nT) at 1 AU. Also plotted is the sunspot number. Here and in Fig. 2, three-month running means have been taken

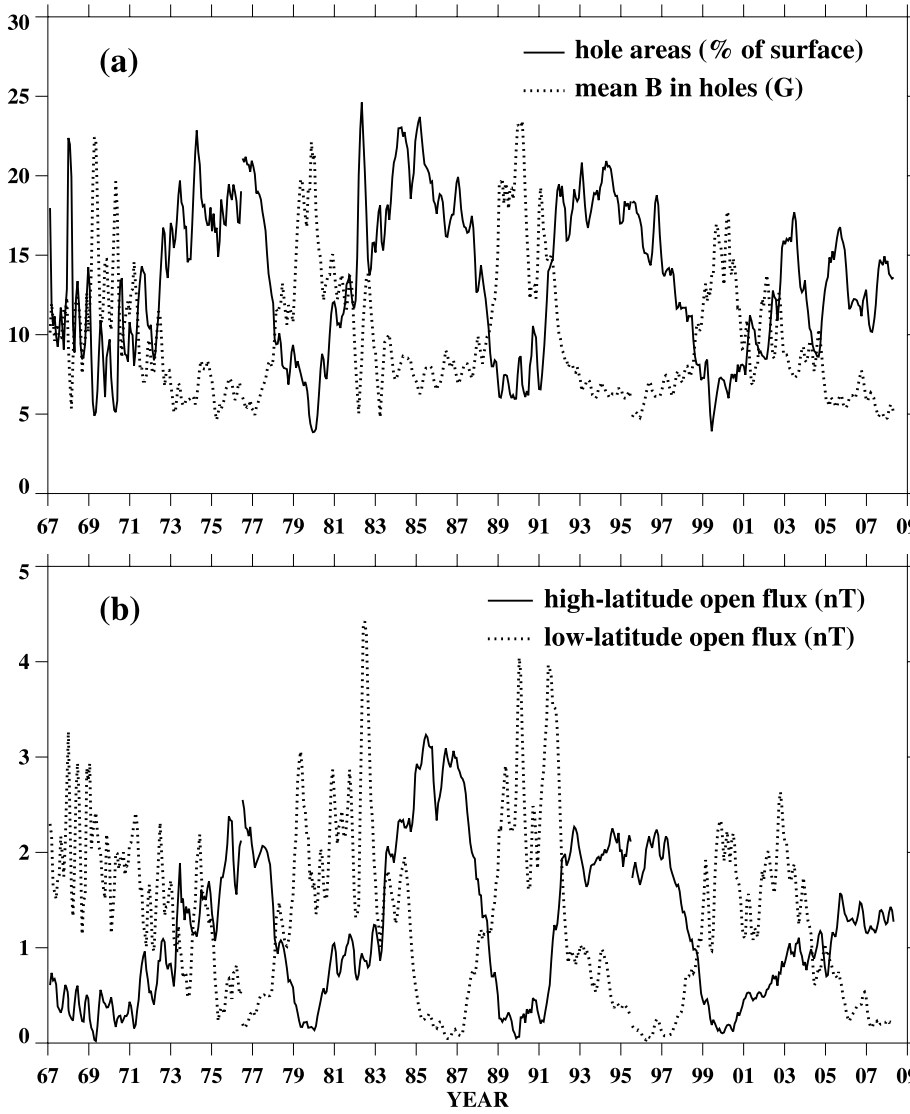


Fig. 2 (**a**) Time variation of the total surface area occupied by open flux (% of the solar surface) and of the average footpoint field strength in open regions (G). (**b**) Time variation of the open flux originating from high latitudes ($|L| > 45°$) and from low latitudes ($|L| < 45°$), expressed as field strengths (nT) at 1 AU

sunspot minimum to ∼20 G at sunspot maximum. The total open flux, which is the product of these two quantities, thus remains constant to within a factor of 2. From Fig. 2(b), we see that most of the open flux resides at high latitudes near sunspot minimum but at low latitudes near sunspot maximum. The low-latitude open flux is characterized by large footpoint field strengths because it is rooted in and around active regions. The high-latitude open flux is seen to be much weaker during the present activity minimum than during the 1976, 1986, and 1996 minima, reflecting the unusually weak polar fields at the end of cycle 23.

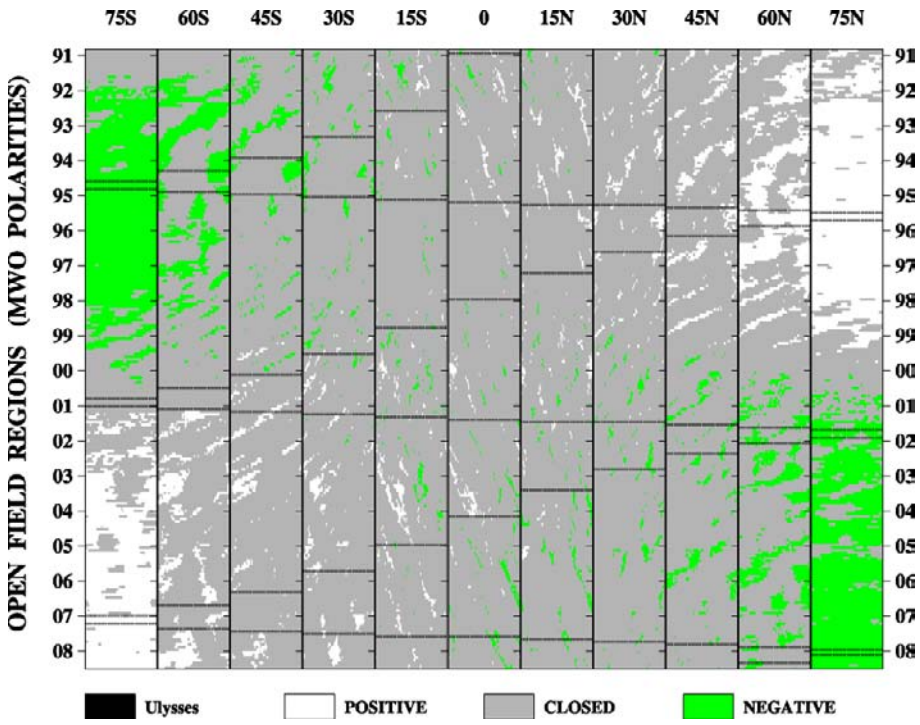


Fig. 3 Stackplot array showing the latitude-by-latitude evolution of coronal holes during 1990–2008. These footpoint areas of open flux were determined from a PFSS extrapolation of MWO magnetograph measurements. In each latitude panel, successive rows of pixels represent successive 27.3 day Carrington rotations, with longitude increasing from 0° at the left to 360° at the right. *White*: outward-pointing open flux. *Green*: inward-pointing open flux. *Gray*: closed field regions. *Horizontal lines*: latitude trajectory of *Ulysses*

The anticorrelation between coronal hole area and field strength suggested by Fig. 2(a) is mainly a consequence of photospheric flux transport (Sect. 3): the open flux is initially concentrated near the edges of active regions, but occupies a progressively larger area and decreases in strength as it diffuses over the solar surface.

The stackplot array in Fig. 3 shows, at a series of latitudes between $L = -75°$ and $L = +75°$, the evolution of open field regions during 1990–2008. In each latitude panel, successive rows of pixels represent successive Carrington rotations, with longitude ϕ increasing from left to right and time running in the opposite direction. The global picture is dominated by the waxing and waning of the polar holes, which disappear when the polar fields reverse near sunspot maximum. At lower latitudes, the coronal holes of a given polarity tend to form coherent patterns lasting up to ∼1–2 yr, with the patterns slanting downward and to the right (left) if they rotate faster (slower) than the 27.3 day Carrington rate.

That the polar holes are confined to latitudes above 60° is significant: it implies that the polar fields have a highly concentrated distribution near sunspot minimum, varying with latitude roughly as $\sin^7 L$ (Svalgaard et al. 1978; Wang and Sheeley 1995). If the polar fields instead had a simple dipole ($\sin L$) distribution, the polar hole boundaries would extend all the way down to latitude 40°. The bunched "topknot" form of the polar fields is also evident from observations of polar plumes, which are much more steeply inclined toward the equator than expected for a dipole field (Saito 1965). The extreme poleward concentration of the photospheric flux is due to the presence of the surface meridional flow.

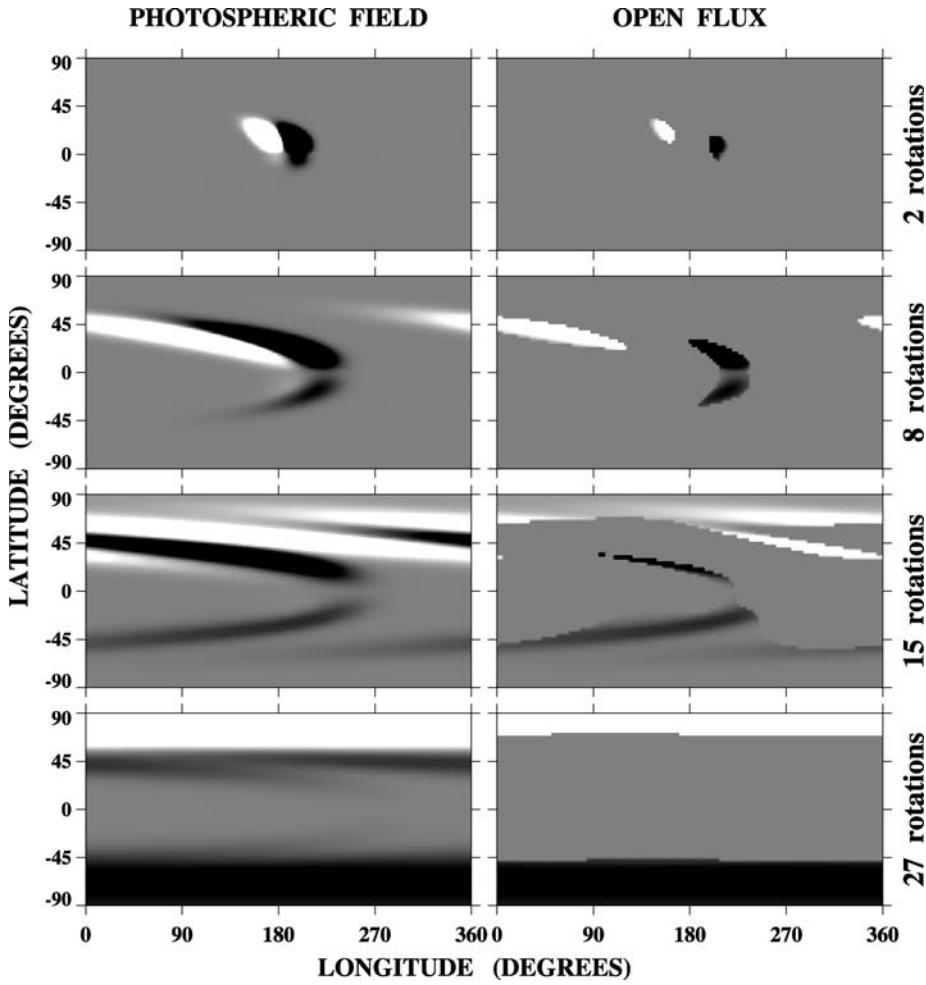


Fig. 4 Large BMR and associated open flux evolving under the influence of differential rotation, diffusion, and poleward flow. The bipole was deposited at latitude $L = +10^\circ$ with a longitudinal (latitudinal) pole separation of 20° (4°). The Carrington-format maps display the distribution of the photospheric field (*left panels*) and of the open flux (*right panels*) after 2, 8, 15, and 27 rotations. *White* (*black*) denotes $B_r > 0$ ($B_r < 0$)

3 Flux Transport and the Formation of the Polar Coronal Holes

The relation between coronal holes and sunspot activity can be understood using a transport model for the photospheric field. We start with a single bipolar magnetic region (BMR) and allow it to evolve under the influence of the photospheric differential rotation, supergranular convection (turbulent diffusion) at a rate of 600 $\mathrm{km^2\,s^{-1}}$, and a poleward bulk flow of amplitude $V_\mathrm{m} = 15\ \mathrm{m\,s^{-1}}$. The BMR, representing a large, idealized active region, is deposited at latitude $L = +10^\circ$, with its poles separated by 20° in longitude and 4° in latitude. The left panels in Fig. 4 show the distribution of the photospheric field after 2, 8, 15, and 27 (27.3 day) rotations, while the corresponding PFSS-derived open field regions are displayed at the right. A pair of small, opposite-polarity holes forms at time $t = 0$ at the far

corners of the BMR: these are the footpoint areas of the high loops that reach the source surface. The nonaxisymmetric (ϕ-dependent) component of the photospheric flux distribution $B_r(R_\odot, L, \phi, t)$ is sheared by the differential rotation, and diffusively annihilated as the latitudinal gradient $|\partial B_r/\partial L|$ progressively steepens. The meridional flow accelerates this process by carrying the BMR flux to midlatitudes, where the rotational gradients are largest. At the same time, a small amount of leading-polarity flux diffuses across the equator. After a time $\tau_{\rm flow} \sim R_\odot/V_{\rm m} \sim 1.5$ yr, the nonaxisymmetric component of the field (including the equatorial dipole moment) has decayed away and the remaining axisymmetric flux has been transported to the poles, forming the polar fields with their embedded holes. Note that the axis of the BMR must be tilted with respect to the east-west line in order for any net flux to reach the poles. Also, it is evident that our newly created polar fields could be canceled again by depositing another BMR with reversed east-west polarity orientation and continuing the simulation.

As remarked in relation to the stackplots of Fig. 3, low-latitude coronal holes and the equatorward extensions of the polar holes tend to form coherent patterns lasting ∼1–2 yr. Likewise, an inspection of Figs. 1 and 2 suggests that the peaks in the radial IMF strength and in the low-latitude open flux are typically on the order of a year wide. This characteristic width corresponds to the decay time for the equatorial dipole field, which in turn is determined by $\tau_{\rm flow}$, the timescale for the surface meridional flow to carry the active region fields to midlatitudes. The same process may be responsible for some of the ∼1.2–1.7 yr quasi-periodicities detected intermittently in the IMF, solar wind, and geomagnetic activity (Silverman and Shapiro 1983; Richardson et al. 1994; Mursula et al. 2003).

4 Solar Wind Speed, Coronal Heating, and Flux-Tube Expansion

Figure 5 compares the proton flow speeds recorded at *Ulysses* during 1990–2008 with the flux-tube expansion factors $f_{\rm ss}$ derived from MWO photospheric field measurements. We conclude from the similar appearance of the two stackplots that fast wind is associated with relatively small expansion factors, and slow wind with very large expansion factors. The three main bands of very fast wind (low expansion) correspond to the *Ulysses* polar passes of 1994–1995, 2001–2002, and 2006–2008.

Now using $f_{\rm ss}$ as an (inverse) proxy for wind speed, we display in Fig. 6 the global patterns of wind speed during 1990–2008 (compare Fig. 3, which shows the underlying distribution of coronal holes). The global picture is dominated by the high-speed wind from the polar holes, whose large interiors are characterized by relatively slow flux-tube expansion. As the polar holes recede and disappear at sunspot maximum, low-speed wind spreads from low latitudes all the way to the poles. The sources of this wind are the many small holes located around active regions and containing strong, rapidly diverging fields (Levine 1982; Wang and Sheeley 1990b; Kojima et al. 1999; Neugebauer et al. 2002; Schrijver and DeRosa 2003; Liewer et al. 2004; Sakao et al. 2007). In contrast, the bulk of the low-speed wind near sunspot minimum comes from the rapidly diverging flux tubes at the polar hole boundaries.

The physical basis for the wind speed–expansion factor relationship can be understood as follows. Let us assume the existence of a heating source in coronal holes which varies with radial distance. If the bulk of the energy is deposited close to the coronal base, the downward heat conduction will act to drive a large mass flux, and the energy available per solar wind proton will be reduced. In contrast, if the energy is deposited over a larger distance extending toward the sonic point, more of it will go into accelerating the wind and

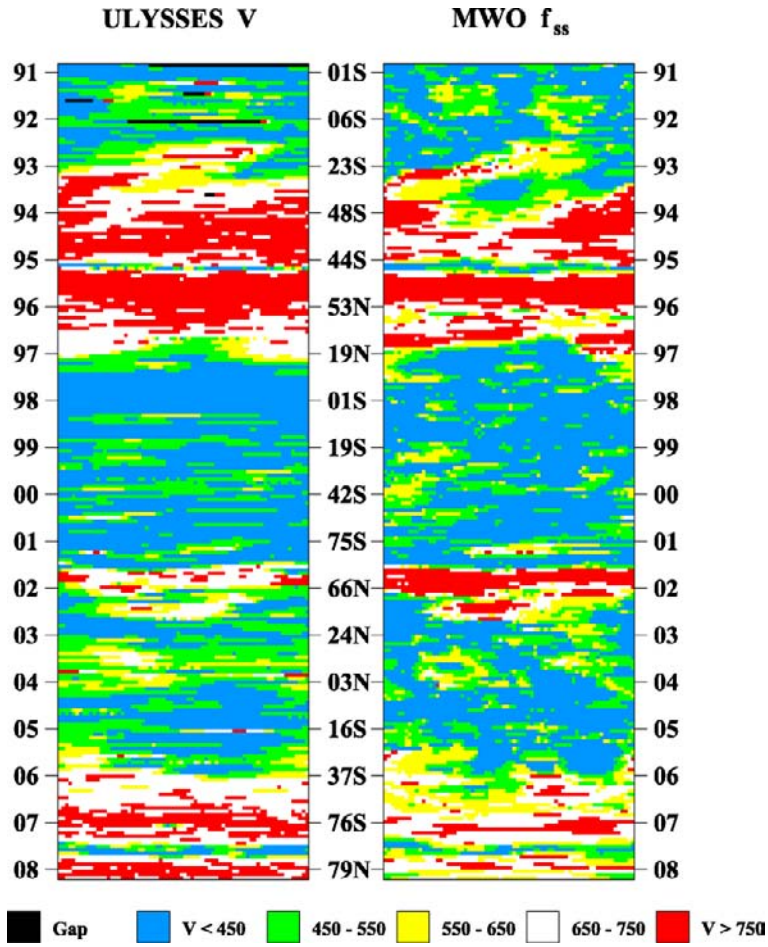


Fig. 5 Stackplots comparing the daily wind speeds V observed at *Ulysses* during 1990–2008 (*left*) with the values predicted by applying the expansion factor model to MWO photospheric field measurements (*right*). Carrington longitude runs from left to right; the heliographic latitude and heliocentric distance (in AU) of the spacecraft are given alongside the horizontal ticks marking the start of each year. *Red*: $V > 750$ km s^{-1} ($f_{ss} < 4.5$). *White*: $V = 650$–750 km s^{-1} ($4.5 < f_{ss} < 7$). *Yellow*: $V = 550$–650 km s^{-1} ($7 < f_{ss} < 10$). *Green*: $V = 450$–550 km s^{-1} ($10 < f_{ss} < 20$). *Blue*: $V < 450$ km s^{-1} ($f_{ss} > 20$). *Black*: data gap

less into increasing the mass flux (Leer and Holzer 1980). If the source of the heating is the coronal magnetic field, we would then expect a rapidly diverging field to be characterized by a shorter damping length and produce slower wind than a more gradually diverging field.

We illustrate these points by numerically solving the single-fluid equations of mass, momentum, and energy conservation for a thermally driven wind, including the effects of coronal heating, heat conduction, and radiative losses (*cf.* Hammer 1982; Withbroe 1988). The flow is taken to be along a radially oriented flux tube, with the magnetic field $B(r)$ falling off as $r^{-\nu}$ for $r \lesssim 2.5\ R_\odot$ and as r^{-2} for $r \gg 2.5\ R_\odot$. The model explicitly includes a chromospheric-coronal transition region, where the downward heat flux is balanced by radiation and an outward enthalpy flux (which in turn determines the mass flux). To demonstrate that a coronal heating rate that depends mainly on the local magnetic field strength will lead

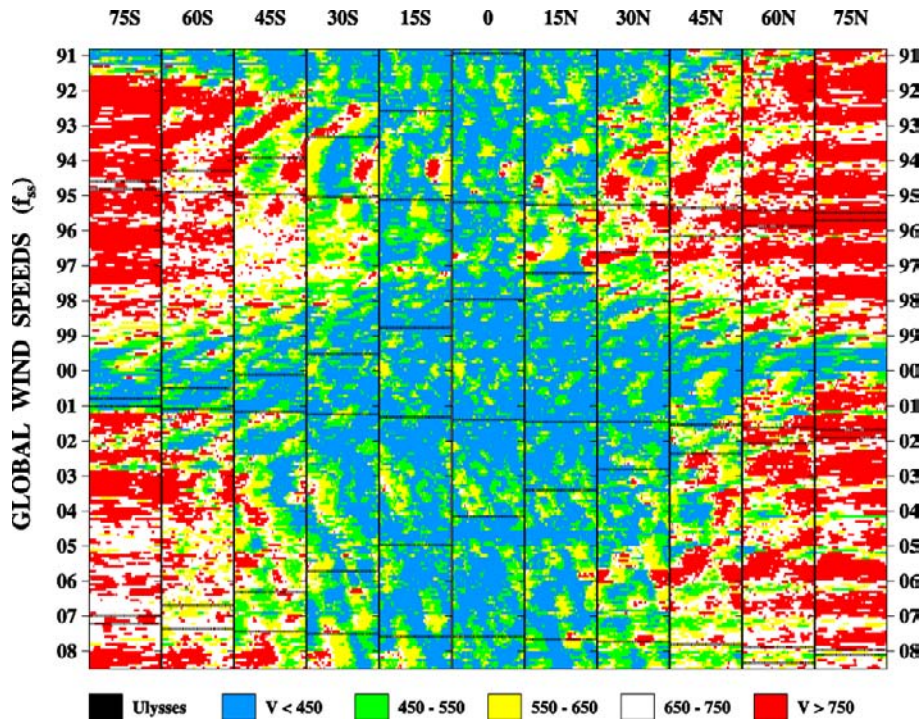


Fig. 6 Multilatitude stackplot array showing the global evolution of the solar wind speed during 1990–2008, as derived from the expansion factor relationship. In each latitude panel, successive rows of pixels represent successive 27.3 day Carrington rotations, with longitude increasing from left to right. Color-coding as in Fig. 5. *Black horizontal lines* indicate the latitude trajectory of *Ulysses*

to an inverse relationship between wind speed and expansion factor, we arbitrarily adopt a heating function of the form

$$F_{\rm h} = F_{\rm h0} \left(\frac{B}{B_0} \right)^{\mu}, \tag{3}$$

where, for definiteness, we take $F_{\rm h0} = 8 \times 10^5$ erg cm^{-2} s^{-1} and $\mu = 3/2$. Figure 7 shows the steady-state profiles of flow speed $u(r)$, temperature $T(r)$, proton (or electron) density $n(r)$, and proton flux $n(r)u(r)$ obtained by setting ν equal to 2, 3, and 4. As the magnetic falloff rate increases, the location of the temperature maximum moves inward, the mass flux density at the coronal base increases, the temperatures fall in the outer corona, and the flow velocity at the outer boundary decreases. Similar results hold for any $\mu > 1$.

A possible source of the heating in coronal holes is reconnection between the unipolar flux concentrations and the ubiquitous "magnetic carpet," consisting of small bipoles that continually emerge at the photosphere; this network and intranetwork activity may in turn give rise to MHD waves that propagate outward into the corona before eventually dissipating (see, e.g., Parker 1991; Schrijver et al. 1998). According to Cranmer et al. (2007), however, incompressible Alfvén waves are generated by granular motions and subsequently damped via a turbulent cascade. Since the volumetric heating rate varies as $L_\perp^{-1} \propto B^{1/2}$, where $L_\perp$ is the transverse correlation length for the turbulence (*cf.* Hollweg 1986), Cranmer et al. likewise find that the wind speed is inversely correlated with the rate of magnetic falloff.

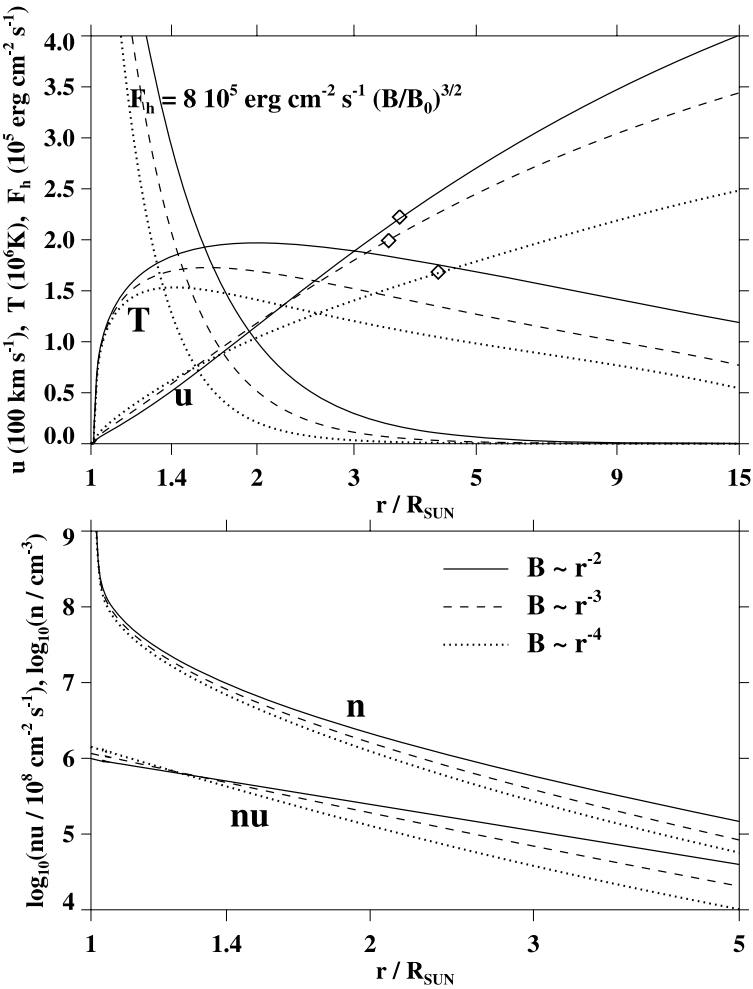


Fig. 7 Three thermally driven wind solutions obtained by varying the magnetic falloff rate in the coronal heating function $F_{\rm h}(r) = 8 \times 10^5$ erg cm^{-2} s^{-1} $(B/B_0)^{3/2}$. *Solid lines*: $\nu = 2$. *Dashed lines*: $\nu = 3$. *Dotted lines*: $\nu = 4$. *Diamonds* mark the location of the sonic point

The relationship between network activity and coronal heating is especially clear in polar plumes. In this case, strong localized heating occurs near the base of the plume, where a small bipole (in the form of an EUV bright point) undergoes reconnection with a unipolar flux concentration inside the polar hole. This extra base heating drives a large downward heat flux and raises the density everywhere along the flux tube, while causing the temperature and flow speed above the dissipation region to *decrease* (Wang 1994), as confirmed by SUMER and UVCS observations (Wilhelm et al. 1998; Giordano et al. 2000; Teriaca et al. 2003). The plume decays after the minority-polarity flux is canceled on the ∼1 day timescale of the supergranular convection.

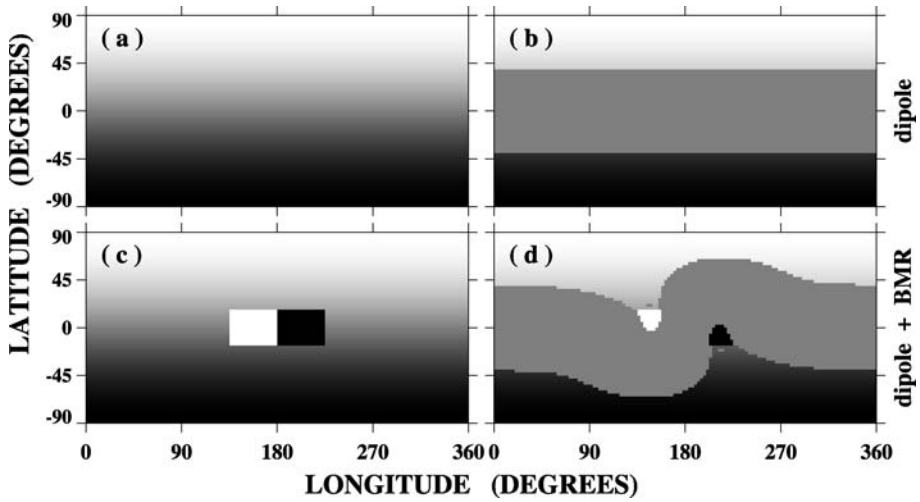


Fig. 8 Interaction between an equatorial BMR and the polar fields. (**a**) Axisymmetric dipole field of strength 1 G; (**b**) corresponding distribution of open flux (*gray* represents closed fields). (**c**) Superposition of a BMR of strength 5 G and an axisymmetric dipole of strength 1 G; (**d**) corresponding distribution of open flux. The flux distributions are all plotted at the solar surface

5 Magnetic Reconnection and the Rotation of Coronal Holes

The tendency for coronal holes to rotate more rigidly than the photosphere is most striking in the equatorward extensions of the polar holes (Timothy et al. 1975). This behavior becomes less puzzling once it is understood how these extensions are formed. Consider an idealized initial configuration consisting of an axisymmetric dipole field with its associated polar holes (Fig. 8, top panels). After depositing an east-west oriented BMR at the equator, we obtain a pair of equatorward extensions that link each polar hole to the like-polarity sector of the bipole (Fig. 8, bottom panels).

We now include the photospheric differential rotation, with angular velocity profile $\omega(L) = 13.38 - 2.30\sin^2 L - 1.62\sin^4 L$ deg day^{-1} (Snodgrass 1983), and allow the system to evolve with time. The leftmost panels in Fig. 9 show the photospheric field after 0, 1, and 2 (27.3 day) rotations, the middle panels show the corresponding open field regions, while the rightmost panels illustrate how the same regions would evolve if they sheared at the local plasma rate. We see that the PFSS-derived polar hole extensions hardly change their shape at all during the simulation. This result follows immediately from the fact that the large-scale photospheric flux distribution, which uniquely determines the coronal field in the current-free approximation, is practically time-independent in a frame that corotates with the BMR.

In reality, in the presence of a plasma, the coronal field must undergo continual reconnection in order to remain close to a potential state (see the MHD simulation of Lionello et al. 2005). Possible evidence for "interchange reconnection" between open and closed field lines is provided by coronagraph observations of density inhomogeneities that propagate outward along the heliospheric current/plasma sheet (Sheeley et al. 1997; Wang et al. 1998a; Crooker et al. 2004). The blobs appear to originate from the closed portions of helmet streamers and to be squeezed out through their cusps (Fig. 10). One interpretation of these white-light observations is that the streamer loops are undergoing reconnection with neighboring open field lines at the Y-point, leading to an exchange of footpoints in such a way

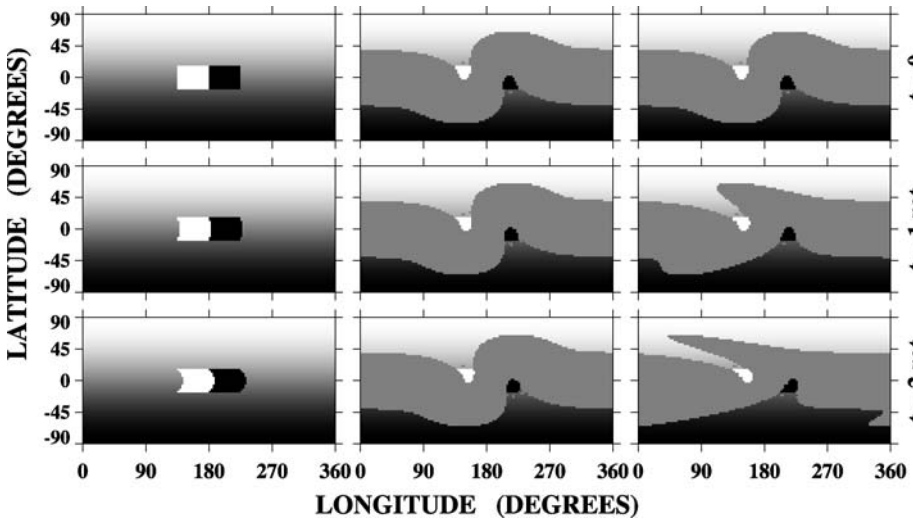


Fig. 9 Rotational evolution of a configuration consisting of a 5 G BMR at the equator and a 1 G axisymmetric dipole field. *Left*: Photospheric field after the lapse of 0, 1, and 2 rotations. *Middle*: Corresponding open field regions. *Right*: Open field regions as they would appear if they rotated at the photospheric plasma rate

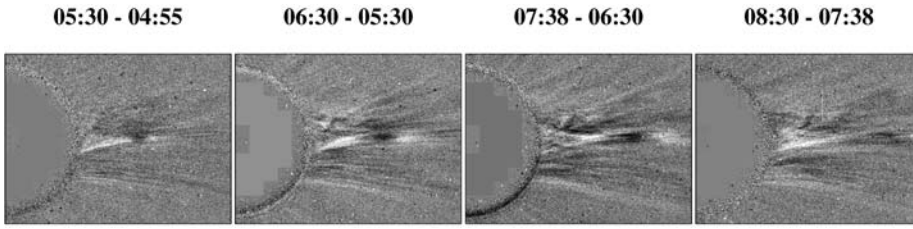


Fig. 10 Sequence of *SOHO* LASCO C2 running-difference images showing the ejection of a plasma blob from the cusp of a helmet streamer, 25 February 1997. The blob first appears as a slight density enhancement (*white* feature) near $r \sim 3.5\ R_{\odot}$. Note the background of fine raylike structures threading the plasma sheet, which may represent open field lines that have undergone interchange reconnection with the closed streamer loops

as to oppose the deformation (by rotational shearing or supergranular convective motions) of the coronal hole boundaries. At the same time, material is injected into the plasma sheet without eroding the helmet streamer or changing the total amount of open flux.

In addition to the outward-moving streamer blobs, the LASCO C2 coronagraph has detected thousands of inflow events at heliocentric distances of 2 to 5 $R_{\odot}$, again concentrated around the heliospheric current/plasma sheet (Sheeley and Wang 2002). A typical example is displayed in Fig. 11: the inward-moving structure leaves a narrow, dark trail in its wake; as it approaches the inner edge of the coronagraph field of view at $r \sim 2\ R_{\odot}$, it decelerates and takes on a cusp-like appearance. Such events, seen most frequently during times of high solar activity, are strongly suggestive of the closing-down of magnetic flux at coronal hole boundaries or (in some cases) in the aftermath of CMEs. The U-loops that result from the disconnection process should be detectable at 1 AU as disruptions in the suprathermal electron strahl streaming away from the Sun. While heat flux dropouts are common near the heliospheric current sheet, Crooker and Pagel (2008) conclude that they may signal either

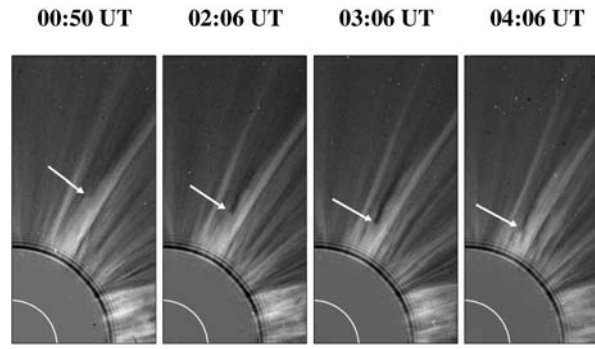


Fig. 11 Inflow observed with the LASCO C2 coronagraph on 25 October 1999. The sinking column of streamer material leaves a dark depletion trail in its wake and takes on a cusp-like appearance below $r \sim 2.5\ R_\odot$

disconnection or interchange reconnection (where one end of the field line remains anchored to the Sun).

6 Coronal Holes, Jets, and ^{3}He-Rich Particle Events

Coronal holes are copious emitters of X-ray and EUV jets (Shibata et al. 1992; Moses et al. 1997; Cirtain et al. 2007). The ejections are triggered by X-point reconnection between the small bipoles that continually emerge inside the holes and the overlying open flux, which acts to collimate the hot plasma and channel it out into the heliosphere (Shimojo and Shibata 2000). A single flaring bright point may emit several jets during its lifetime. The brighter and faster jets are also detected in white light beyond $r \sim 2\ R_\odot$, with their leading edges traveling at speeds of ~400–1000 km s^{-1} (Wang et al. 1998b).

The largest jets originate from small, flaring active regions located inside or near the boundaries of low-latitude coronal holes. In some cases, the source of the jets may be connected to Earth along the Parker spiral, allowing the energetic particles associated with these reconnection events to be observed *in situ*.

Figure 12 shows two EUV/white-light jet events originating from an active region near the west limb on 5 October 2002. The change in direction between the EUV jet and the corresponding white-light ejection above $r \sim 2\ R_\odot$ can be understood from the PFSS extrapolation in Fig. 13, which shows the southward-pointing open field lines adjacent to the flaring region bending sharply northward at greater heights. Over a 5-day interval beginning on October 3, the same small active region produced as many as 16 jet events, some of which consisted of multiple ejections (see Fig. 14). Early on October 5, the ULEIS detector on *ACE* recorded a steep increase in the flux of ^{3}He and Fe ions as connection to the source was established; the flux remained high over the next two days, and then fell on October 7 (Fig. 15). The highest peak in the particle intensity occurred ~8 hr after the event shown in the bottom panels of Fig. 12. Recent studies suggest that such impulsive solar energetic particle (SEP) events invariably originate from small active regions next to or inside coronal holes, whose open field lines channel the fractionated products of the reconnection process into the heliosphere (Reames 2002; Wang et al. 2006; Pick et al. 2006; Nitta et al. 2006).

7 Concluding Remarks

In this overview, we have emphasized the role of active regions, surface flux transport, and magnetic reconnection in the formation, evolution, and rotation of coronal holes. Much of

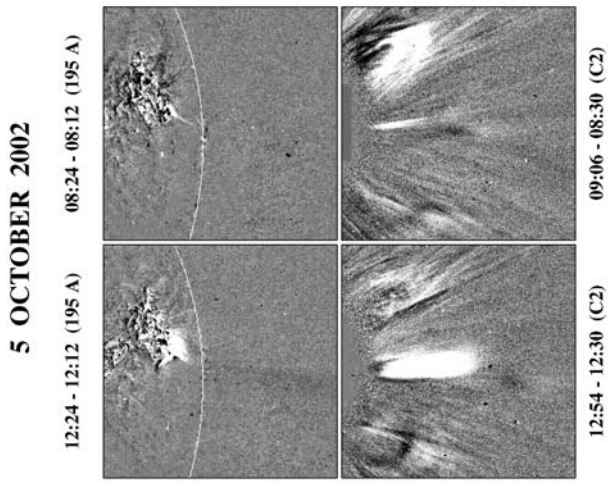


Fig. 12 Difference images showing two "homologous" *SOHO* LASCO/EIT jet events on 5 October 2002. The Fe XII 19.5 nm jets are on the *left*, while their white-light counterparts beyond $r \sim 2\ R_{\odot}$ are on the *right*. Note the change in direction of the jets as they propagate from the solar surface to the outer corona (compare Fig. 13)

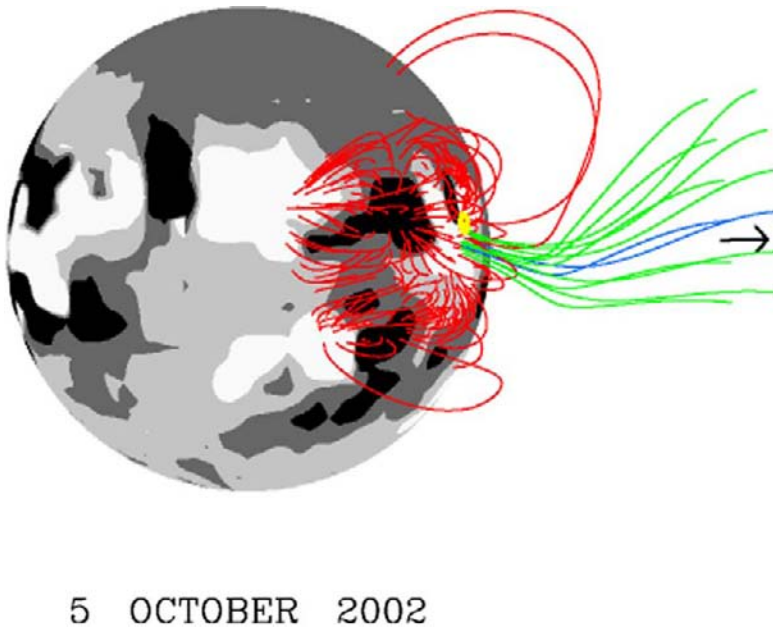


Fig. 13 Field-line configuration of the source region of the 5 October 2002 jets, derived by extrapolating the observed photospheric field. The open field lines (coded *blue* if directed into the ecliptic, *green* otherwise) point southward near the solar surface but bend northward at greater heights, thus accounting for the differing orientations of the EUV and white-light jets in Fig. 12. The *yellow dot* marks the location of the flaring source; the *arrow* indicates the direction of the LASCO C2 jets

this discussion has been based on approximating the coronal field as current-free. A major limitation of this approach is that it does not tell us how coronal loops open up to form long-lived coronal holes (are CME events involved, as suggested by Luhmann et al. 1998?), how the field-line reconnection that maintains their quasi-rigid rotation actually takes place, or how the open flux eventually closes down. These questions can addressed both observa-

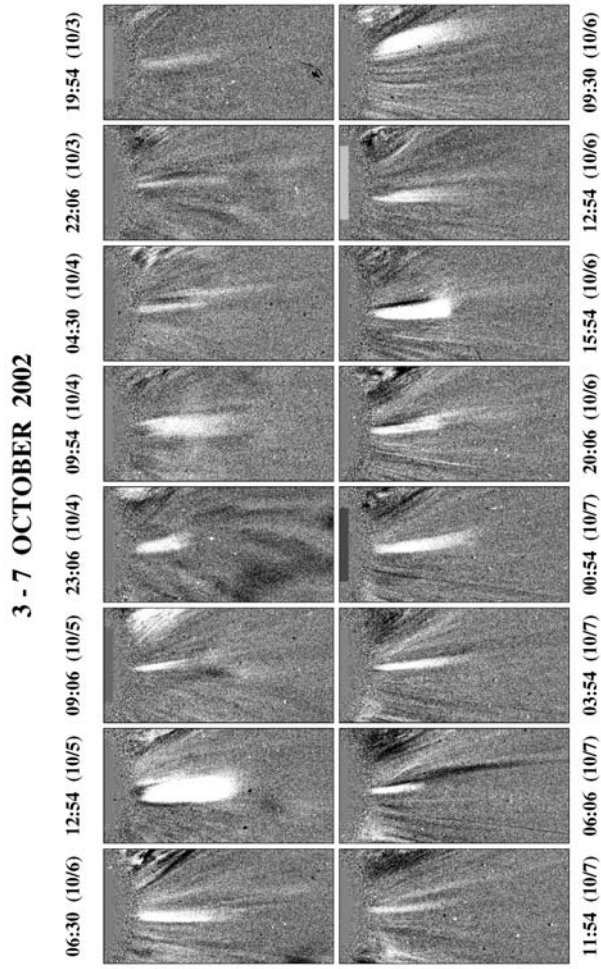


Fig. 14 The white-light jets shown in Fig. 12 were just two of 16 such events recorded by the LASCO C2 coronagraph from the same flaring active-region–coronal-hole system during 3–7 October 2002. The average speed of the jets was close to 700 km s^{-1}

tionally and with the help of 3D MHD simulations. In particular, the SECCHI white-light and EUV instruments on the twin *STEREO* spacecraft will make it easier to relate CMEs, slow streamer expansions/disruptions, streamer blobs, and inflow events (all of which are best observed near the sky plane) to changes in coronal hole boundaries (which are most visible on the disk).

Another unresolved issue concerns the sources of low-speed solar wind, which is characterized by high temporal and spatial variability and distinctive compositional properties (Bame et al. 1977; Geiss et al. 1995; von Steiger et al. 2000; Ko et al. 2006). It is often argued that this plasma must originate outside coronal holes, i.e., from closed field regions (Schwadron et al. 1999; Zurbuchen et al. 2000; Woo et al. 2004; Feldman et al. 2005; McComas et al. 2007; Zurbuchen 2007). We have suggested instead that the bulk of the slow wind comes either from just inside the boundaries of large coronal holes or from the small, rapidly evolving open field regions present around active regions. The high freeze-in temperatures (or $n_{O^{7+}}/n_{O^{6+}}$ ratios) measured in the slow wind and its tendency to be relatively enriched in elements of low first-ionization potential (FIP) are then attributed to the heating being concentrated near the coronal base, which steepens the local temperature gradient

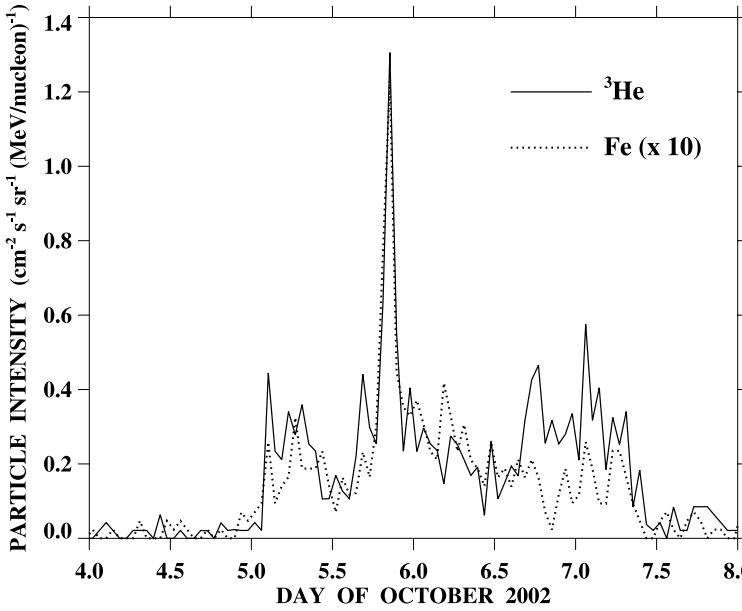


Fig. 15 Ion intensities in the 0.32–0.45 MeV nucleon^{-1} range measured by *ACE*/ULEIS during 4–7 October 2002 (compare Fig. 14). *Solid curve*: ^{3}He. *Dotted curve*: Fe. The highest peak occurs ∼8 hr after the flare/jet event displayed in the bottom panels of Fig. 12

and leads to enhanced chromospheric evaporation (Wang and Sheeley 2003). Self-consistent modeling of coronal heating and solar wind acceleration, along the lines of Cranmer et al. (2007) and Suzuki and Inutsuka (2006), may eventually help to settle this debate.

Acknowledgements I am indebted to N.R. Sheeley Jr. for his long-standing and continuing collaboration, to R. Grappin for developing the solar wind code used to obtain the results in Fig. 7, and to A. Balogh for inviting me to attend the ISSI Workshop "Origin and Dynamics of Solar Magnetism." Financial support was provided by NASA and the Office of Naval Research.

References

M.D. Altschuler, G. Newkirk Jr., Sol. Phys. **9**, 131 (1969)
C.N. Arge, V.J. Pizzo, J. Geophys. Res. **105**, 10465 (2000)
C.N. Arge, E. Hildner, V.J. Pizzo, J.W. Harvey, J. Geophys. Res. **107**, A10 (2002), SSH 16-1
A. Balogh, E.J. Smith, B.T. Tsurutani, D.J. Southwood, R.J. Forsyth, T.S. Horbury, Science **268**, 1007 (1995)
S.J. Bame, J.R. Asbridge, W.C. Feldman, J.T. Gosling, J. Geophys. Res. **82**, 1487 (1977)
J.W. Cirtain et al., Science **318**, 1580 (2007)
S.R. Cranmer, Space Sci. Rev. **101**, 229 (2002)
S.R. Cranmer, A.A. van Ballegooijen, R.J. Edgar, Astrophys. J. Suppl. Ser. **171**, 520 (2007)
N.U. Crooker, C. Pagel, J. Geophys. Res. **113**, A02106 (2008)
N.U. Crooker, C.-L. Huang, S.M. Lamassa, D.E. Larson, S.W. Kahler, H.E. Spence, J. Geophys. Res. **109**, A03107 (2004)
U. Feldman, E. Landi, N.A. Schwadron, J. Geophys. Res. **110**, A07109 (2005)
J. Geiss, G. Gloeckler, R. von Steiger, Space Sci. Rev. **72**, 49 (1995)
S. Giordano, E. Antonucci, G. Noci, M. Romoli, J.L. Kohl, Astrophys. J. **531**, L79 (2000)
R. Hammer, Astrophys. J. **259**, 767 (1982)
J.V. Hollweg, J. Geophys. Res. **91**, 4111 (1986)

Y.-K. Ko, J.C. Raymond, T.H. Zurbuchen, P. Riley, J.M. Raines, L. Strachan, Astrophys. J. **646**, 1275 (2006)
M. Kojima, K. Fujiki, T. Ohmi, M. Tokumaru, A. Yokobe, K. Hakamada, J. Geophys. Res. **104**, 16993 (1999)
R.A. Kopp, T.E. Holzer, Sol. Phys. **49**, 43 (1976)
E. Leer, T.E. Holzer, J. Geophys. Res. **85**, 4681 (1980)
R.H. Levine, Sol. Phys. **79**, 203 (1982)
R.H. Levine, M.D. Altschuler, J.W. Harvey, J. Geophys. Res. **82**, 1061 (1977)
P.C. Liewer, M. Neugebauer, T. Zurbuchen, Sol. Phys. **223**, 209 (2004)
R. Lionello, P. Riley, J.A. Linker, Z. Mikić, Astrophys. J. **625**, 463 (2005)
M. Lockwood, R. Stamper, M.N. Wild, Nature **399**, 437 (1999)
J.G. Luhmann, J.T. Gosling, J.T. Hoeksema, X. Zhao, J. Geophys. Res. **103**, 6585 (1998)
J.G. Luhmann, Y. Li, C.N. Arge, P.R. Gazis, R. Ulrich, J. Geophys. Res. **107**, A8 (2002), SMP 3-1
D.J. McComas et al., Rev. Geophys. **45**, RG1004 (2007)
D. Moses et al., Sol. Phys. **175**, 571 (1997)
R.H. Munro, B.V. Jackson, Astrophys. J. **213**, 874 (1977)
K. Mursula, B. Zieger, J.H. Vilppola, Sol. Phys. **212**, 201 (2003)
M. Neugebauer et al., J. Geophys. Res. **103**, 14587 (1998)
M. Neugebauer, P.C. Liewer, E.J. Smith, R.M. Skoug, T.H. Zurbuchen, J. Geophys. Res. **107**, A12 (2002), SSH 13-1
N.V. Nitta, D.V. Reames, M.L. DeRosa, Y. Liu, S. Yashiro, N. Gopalswamy, Astrophys. J. **650**, 438 (2006)
E.N. Parker, Astrophys. J. **372**, 719 (1991)
M. Pick, G.M. Mason, Y.-M. Wang, C. Tan, L. Wang, Astrophys. J. **648**, 1247 (2006)
D.V. Reames, Astrophys. J. **571**, L63 (2002)
J.D. Richardson, K.I. Paularena, J.W. Belcher, A.J. Lazarus, Geophys. Res. Lett. **21**, 1559 (1994)
K. Saito, Publ. Astron. Soc. Jpn. **17**, 1 (1965)
T. Sakao et al., Science **318**, 1585 (2007)
K.H. Schatten, Cosm. Electrodyn. **2**, 232 (1971)
K.H. Schatten, J.M. Wilcox, N.F. Ness, Sol. Phys. **6**, 442 (1969)
C.J. Schrijver, M.L. DeRosa, Sol. Phys. **212**, 165 (2003)
C.J. Schrijver et al., Nature **394**, 152 (1998)
N.A. Schwadron, L.A. Fisk, T.H. Zurbuchen, Astrophys. J. **521**, 859 (1999)
N.R. Sheeley Jr., Y.-M. Wang, Astrophys. J. **579**, 874 (2002)
N.R. Sheeley Jr. et al., Astrophys. J. **484**, 472 (1997)
K. Shibata et al., Publ. Astron. Soc. Jpn. **44**, L173 (1992)
M. Shimojo, K. Shibata, Astrophys. J. **542**, 1100 (2000)
S.M. Silverman, R. Shapiro, J. Geophys. Res. **88**, 6310 (1983)
E.J. Smith, A. Balogh, R.J. Forsyth, D.J. McComas, Geophys. Res. Lett. **28**, 4159 (2001)
H.B. Snodgrass, Astrophys. J. **270**, 288 (1983)
S.T. Suess, S.F. Nerney, Sol. Phys. **40**, 487 (1975)
T.K. Suzuki, S. Inutsuka, J. Geophys. Res. **111**, A06101 (2006)
L. Svalgaard, T.L. Duvall Jr., P.H. Scherrer, Sol. Phys. **58**, 225 (1978)
L. Teriaca, G. Poletto, M. Romoli, D.A. Biesecker, Astrophys. J. **588**, 566 (2003)
A.F. Timothy, A.S. Krieger, G.S. Vaiana, Sol. Phys. **42**, 135 (1975)
R.K. Ulrich, S. Evans, J.E. Boyden, L. Webster, Astrophys. J. Suppl. Ser. **139**, 259 (2002)
R. von Steiger et al., J. Geophys. Res. **105**, 27217 (2000)
Y.-M. Wang, Astrophys. J. **435**, L153 (1994)
Y.-M. Wang, N.R. Sheeley Jr., Astrophys. J. **355**, 726 (1990a)
Y.-M. Wang, N.R. Sheeley Jr., Astrophys. J. **365**, 372 (1990b)
Y.-M. Wang, N.R. Sheeley Jr., Astrophys. J. **447**, L143 (1995)
Y.-M. Wang, N.R. Sheeley Jr., Astrophys. J. **587**, 818 (2003)
Y.-M. Wang, S.H. Hawley, N.R. Sheeley Jr., Science **271**, 464 (1996)
Y.-M. Wang et al., Astrophys. J. **498**, L165 (1998a)
Y.-M. Wang et al., Astrophys. J. **508**, 899 (1998b)
Y.-M. Wang, M. Pick, G.M. Mason, Astrophys. J. **639**, 495 (2006)
K. Wilhelm, E. Marsch, B.N. Dwivedi, D.M. Hassler, P. Lemaire, A.H. Gabriel, M.C.E. Huber, Astrophys. J. **500**, 1023 (1998)
G.L. Withbroe, Astrophys. J. **325**, 442 (1988)
R. Woo, S.R. Habbal, U. Feldman, Astrophys. J. **612**, 1171 (2004)
J.B. Zirker (ed.), *Coronal Holes and High Speed Wind Streams* (Colorado Assoc. Univ. Press, Boulder, 1977)
T.H. Zurbuchen, Annu. Rev. Astron. Astrophys. **45**, 297 (2007)
T.H. Zurbuchen, S. Hefti, L.A. Fisk, G. Gloeckler, N.A. Schwadron, J. Geophys. Res. **105**, 18327 (2000)

Solar Cycle Forecasting

David H. Hathaway

Originally published in the journal Space Science Reviews, Volume 144, Nos 1–4, 401–412.
DOI: 10.1007/s11214-008-9430-4

Abstract Predicting the behavior of a solar cycle after it is well underway (2–3 years after minimum) can be done with a fair degree of skill using auto-regression and curve fitting techniques that don't require any knowledge of the physics involved. Predicting the amplitude of a solar cycle near, or before, the time of solar cycle minimum can be done using precursors such as geomagnetic activity and polar fields that do have some connection to the physics but the connections are uncertain and the precursors provide less reliable forecasts. Predictions for the amplitude of cycle 24 using these precursor techniques give drastically different values. Recently, dynamo models have been used directly with assimilated data to predict the amplitude of sunspot cycle 24 but have also given significantly different predictions. While others have questioned both the predictability of the solar cycle and the ability of current dynamo models to provide predictions, it is clear that cycle 24 will help to discriminate between some opposing dynamo models.

Keywords Solar activity · Sunspot cycle · Solar cycle forecasting

"Prediction is very difficult, especially about the future"—Niels Bohr

1 Introduction

Predicting the solar cycle is indeed very difficult. A cursory examination of the sunspot record reveals a wide range of cycle amplitudes (Fig. 1). Over the last 24 cycles the average amplitude (in terms of the 13-month-smoothed monthly averages of the daily sunspot number) was about 113. Over the last 400 years the cycle amplitudes have varied widely—from basically zero through the Maunder Minimum to the two small cycles of the Dalton Minimum at the start of the 19th century (amplitudes of 49.2 and 48.7) to the recent string of

D.H. Hathaway (✉)
NASA/MSFC, Huntsville, AL 35812, USA
e-mail: David.Hathaway@nasa.gov
url: http://solarscience.msfc.nasa.gov/

M.J. Thompson et al. (eds.), *The Origin and Dynamics of Solar Magnetism*.
DOI: 10.1007/978-1-4419-0239-9_20

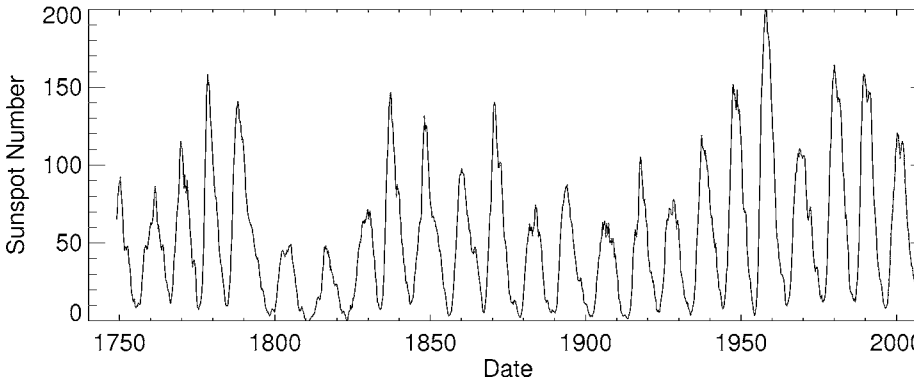


Fig. 1 Smoothed Relative sunspot number since 1749. The solar cycles vary in size, length, and shape

large cycles (amplitudes of 151.8, 201.3, 110.6, 164.5, 158.5, and 120.8). In addition to the changes in the amplitude of the cycle, there are changes in cycle length and cycle shape.

The daily sunspot number is a somewhat contrived indicator of solar activity but its long historical record makes it invaluable. Sunspots range in size from those barely discernable in telescopic views to those large enough to be seen with the naked eye. Counts of individual sunspots will vary depending on the telescope, the atmospheric seeing, and the observer. This led Rudolf Wolf (1852) to devise a Relative sunspot number, R, based primarily on the number of spot groups, G, (which are far more easily and consistently counted) with the number of individual spots in a group, N, adding one-tenth as much such that

$$R = k\,(10G + N) \tag{1}$$

where k is a correction factor for the observer. Although somewhat contrived, this Relative sunspot number is well correlated with other, more physical, measures of the solar cycle. Figure 2 shows examples of four other solar activity indicators plotted against the smoothed Relative sunspot number. All four (sunspot area, 10.7 cm radio flux, total solar irradiance, and M- and X-class flares per month) are well correlated and linearly related to the Relative sunspot number.

The close relationship between solar activity and the Relative sunspot number, together with the long historical record of sunspot observations, has led most efforts at predicting the solar cycle to predict sunspot number. The predicted sunspot numbers can then be used as predictors for other sources of solar activity through the relationships between sunspot number and solar activity source.

Solar cycles overlap by 2–3 years around the time of cycle minima and the magnetic cycle consists of two sunspot cycles. Yet, each sunspot cycle can, and usually is, treated individually. Separating one cycle from the next is not, however, as simple as one might expect. There is a tendency to simply take the month of the minimum of the smoothed sunspot number (or other activity indicator) as the time of cycle minimum and the dividing point between cycles. With this method the month of minimum depends upon the smoothing used (the 13-month smoothing filter has half weights for the months at either end but still passes high frequencies), can be sensitive to small variations in activity near the time of minimum, and can give more than one minimum month. This has led to consideration of other indicators of minimum, including the number of new cycle spots relative to the number of old cycle spots and the number of spotless days

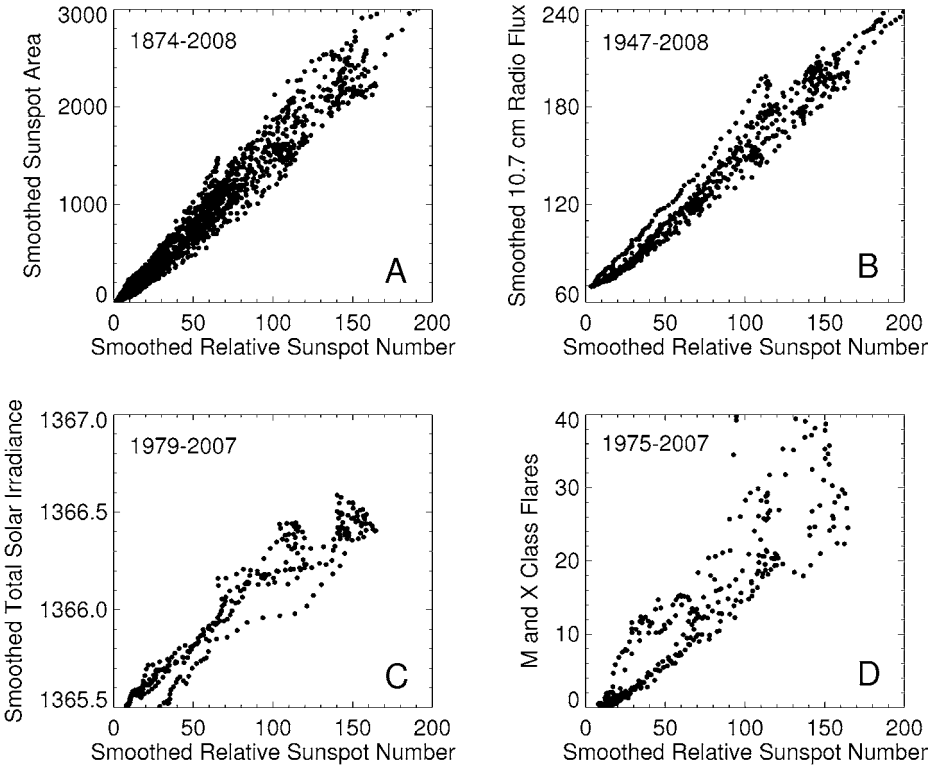


Fig. 2 Four solar activity indicators plotted versus the Relative sunspot number. The smoothed daily sunspot area from 1874 to 2008 is shown in panel **A**. The smoothed daily 10.7 cm radio flux from 1947 to 2008 is shown in panel **B**. The smoothed daily total solar irradiance (PMOD) from 1979 to 2007 is shown in panel **C**. The smoothed monthly number of M- and X-Class flares from 1975 to 2007 is shown in panel **D**. All four solar activity indicators show a nearly linear dependence on the smoothed Relative sunspot number

(e.g. Waldmeier 1961; McKinnon 1987; Harvey and White 1999). An "official" list of the dates and values for the maxima and minima of the sunspot cycles can be found at ftp://ftp.ngdc.noaa.gov/STP/SOLAR_DATA/SUNSPOT_NUMBERS/maxmin.new.

2 Ongoing Cycle Predictions

While sunspot cycles display a significant range of amplitudes as well as variations in length and shape, they also exhibit some fairly consistent behavior that can be utilized in predicting future activity. The sunspot cycles are typically asymmetric in shape with a rapid rise to maximum and a slower decline to minimum. The exceptions to this are some of the early cycles where the observations are sparse and more uncertain. Figure 3 shows an "average" cycle constructed by finding the average length and amplitude, stretching each cycle to match those average values, and then averaging the cycle curves together.

One popular and often used method for predicting solar activity was first described by McNish and Lincoln (1949). As a cycle progresses the smoothed monthly sunspot numbers are compared to the average cycle for the same number of months since minimum. The difference between the two is used to project future differences between predicted and mean

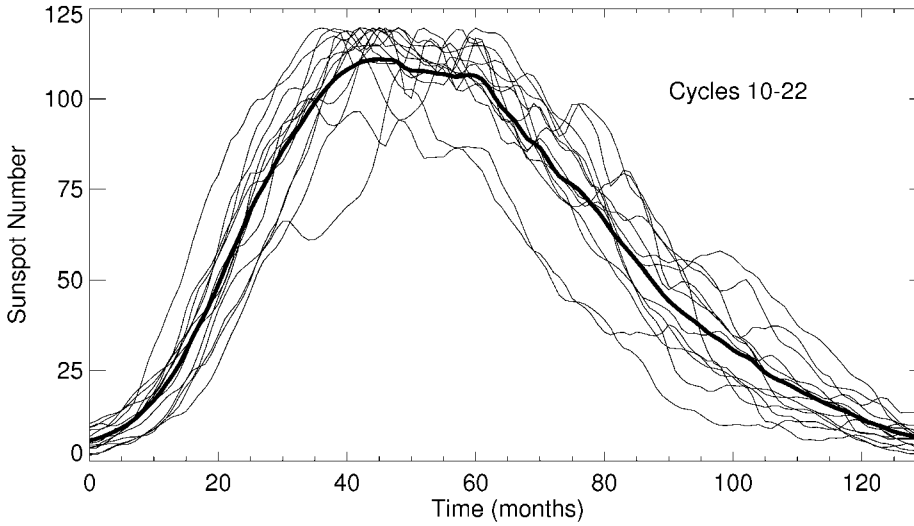


Fig. 3 An average cycle (*thick line*) constructed from the well observed cycles—cycles 10–22. The average amplitude was about 120 and the average cycle length was about 130 months. The average cycle has a rapid rise to maximum and a slower decline to minimum

cycle. The McNish–Lincoln regression technique originally used yearly values and only projected one year into the future. Later improvements to the technique use monthly values and use an auto-regression to predict the remainder of the cycle.

One problem with the modified McNish–Lincoln technique is that it does not account for systematic changes in the shape of the cycle with cycle amplitude. Wolf (1861) noted that small cycles tend to be longer than big cycles while Waldmeier (1935) found that small cycles tend to take longer to reach maximum than do big cycles (the "Waldmeier Effect"). Another problem with the McNish–Lincoln method is its sensitivity to choices for the date of cycle minimum. Both the systematic changes in shape and the sensitivity to cycle minimum choice can be accounted for with techniques that fit the monthly data to parametric curves (e.g. Stewart and Panofsky 1938; Elling and Schwentek 1992; Hathaway et al. 1994). Hathaway et al. (1994) found a two-parameter function which closely mimics the changing shape of the sunspot cycle. Prediction requires fitting the data to the function with a best fit for an initial starting time, t_0, and amplitude, A, where the function is given by

$$R(t) = A(t - t_0)^3 / \{\exp[(t - t_0)^2 / b(A)^2] - 0.71\} \tag{2}$$

where time, t, is measured in months. The change in shape of the curve is given by the width parameter b which is a function of the amplitude and is given by

$$b(A) = 27.12 + 25.15/(A \times 1000)^{1/4}. \tag{3}$$

Both the Modified McNish–Lincoln and the curve-fitting techniques work nicely once a sunspot cycle is well under way. The critical point seems to be 2–3 years after minimum near the time of the inflection point on the rise to maximum. Predictions for cycle 23 using the Modified McNish–Lincoln and the Hathaway, Wilson, & Reichmann curve-fitting techniques in March 1999 (30-months after minimum) are shown in Fig. 4. Since cycle 23 had an amplitude very close to the average of cycles 10–22, both of these predictions are very

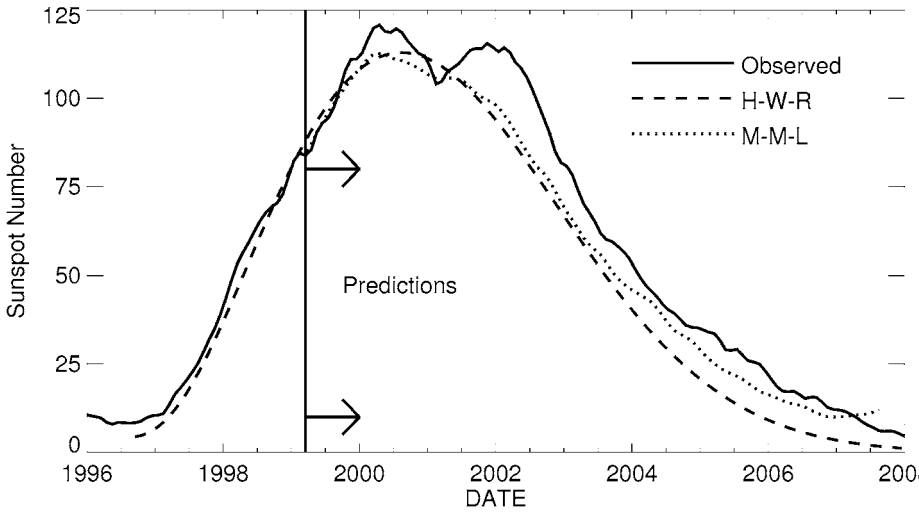


Fig. 4 Predictions for cycle 23 based on data up to March 1999. Both the Modified-McNish–Lincoln, M-M-L (*dotted line*), and the Hathaway–Wilson–Reichmann curve-fitting, H-W-R (*dashed line*) provide good, and similar, predictions for this average sized cycle. These types of predictions don't become reliable until 2–3 years after minimum (cycle 23 minimum was taken as September 1996)

similar. Distinct differences are seen for larger or smaller cycles and when different dates are taken for minimum with the McNish–Lincoln method.

3 Upcoming Cycle Predictions with Precursors

Predicting the size and timing of a cycle prior to its start (or even during the first year or two of the cycle) requires methods other than auto-regression or curve-fitting. There is a long, and growing, list of measured quantities that can and have been used to predict future cycle amplitudes. Prediction methods range from simple climatological means to physics-based dynamos with assimilated data.

The mean amplitude of the last n cycles gives the benchmark for other prediction techniques. The mean of the last 23 cycle amplitudes is 114.1 ± 40.4 where the error is the standard deviation of the mean. This represents a prediction without any skill. If other methods cannot predict with significantly better accuracy they have little use.

One class of prediction techniques is based on trends and periodicities in the cycle amplitudes. In general there has been an upward trend in cycle amplitudes since the Maunder Minimum. Projecting this trend to the next cycle gives a prediction slightly better than the mean. A number of periodicities have been noted in the cycle amplitude record. Gleissberg (1939) noted a long-period variation in cycle amplitudes with a period of seven or eight cycles. Gnevyshev and Ohl (1948) noted a two-cycle periodicity with the odd numbered cycle having larger amplitude than the preceding even numbered cycle. Ahluwalia (1998) noted a three-cycle sawtooth shaped periodicity in the six-cycle record of the geomagnetic *Ap* index.

Another class of prediction techniques uses the characteristics of the preceding cycle as indicators of the size of the next cycle. Wilson et al. (1998) and Solanki et al. (2002) found that the length (period) of the preceding cycle is inversely correlated to the amplitude of the following cycle. Another indicator of the size of the next cycle is the level of activity

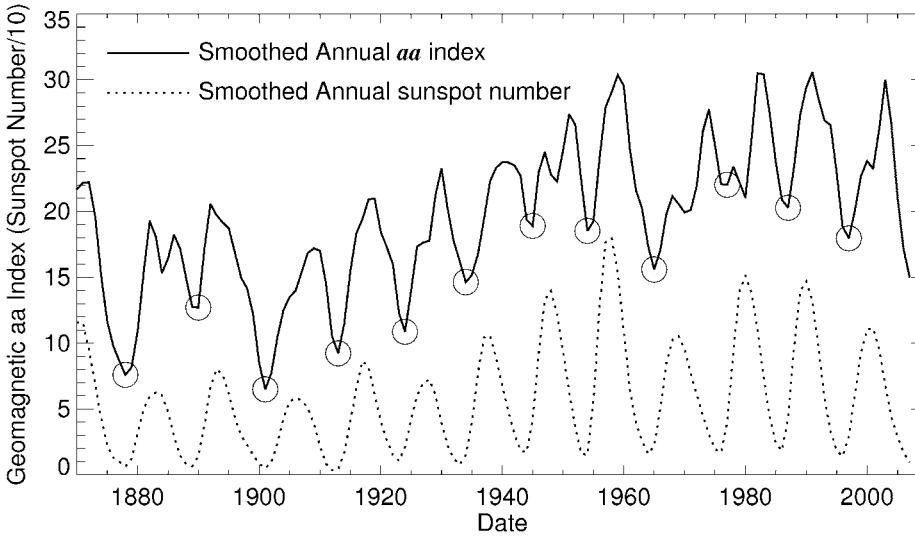


Fig. 5 Smoothed (3-year average with half weights on the ends) annual *aa* index (*solid line*) and sunspot numbers (*dotted line*). The levels of the *aa* index at its minima (*circled*) are good indicators for the maxima of the following sunspot cycles

at minimum—the amplitude of the following cycle is correlated with the smoothed sunspot number at the preceding minimum (Brown 1976). This type of technique has led to searches for activity indicators that are correlated with future cycle amplitude. Javaraiah (2007), for example, has found sunspot areas from intervals of time and latitude that correlate very well with future cycle activity.

One class of precursors for future cycle amplitudes that has worked well in the past uses geomagnetic activity during the preceding cycle at or near the time of minimum as an indicator of the amplitude for the next cycle. These "Geomagnetic Precursors" use indices for geomagnetic activity (rapid changes in the Earth's magnetic field strength and/or direction at ground stations due to solar wind interaction with the Earth's magnetosphere) that extend back to 1844. Ohl (1966) found that the minimum level of geomagnetic activity seen in the *aa* index near the time of sunspot cycle minimum was a good predictor for the amplitude of the next cycle. This is illustrated in Fig. 5. One problem with this method concerns the timing of the *aa* index minima—they often occur well after sunspot cycle minimum and therefore do not give a much advanced prediction.

Two significant variations on this method circumvent the timing problem. Feynman (1982) noted that geomagnetic activity has two different sources—one due to solar activity (flares, CMEs, and filament eruptions) that follows the sunspot cycle and another due to recurrent high speed solar wind streams that peaks during the decline of each cycle. She separated the two by finding the sunspot number dependence of the base level of geomagnetic activity and removing it to reveal the "interplanetary" component of geomagnetic activity. This is illustrated in Fig. 6 using the modifications described by Hathaway and Wilson (2006). The peaks in the interplanetary component occur prior to sunspot cycle minimum and are very good indicators for the amplitude of the following sunspot cycle. Thompson (1993) also noted that some geomagnetic activity during the previous cycle served as a predictor for the amplitude of the following cycle but, instead of trying to separate the two, he simply related the geomagnetic activity (as represented by the number of days with the geo-

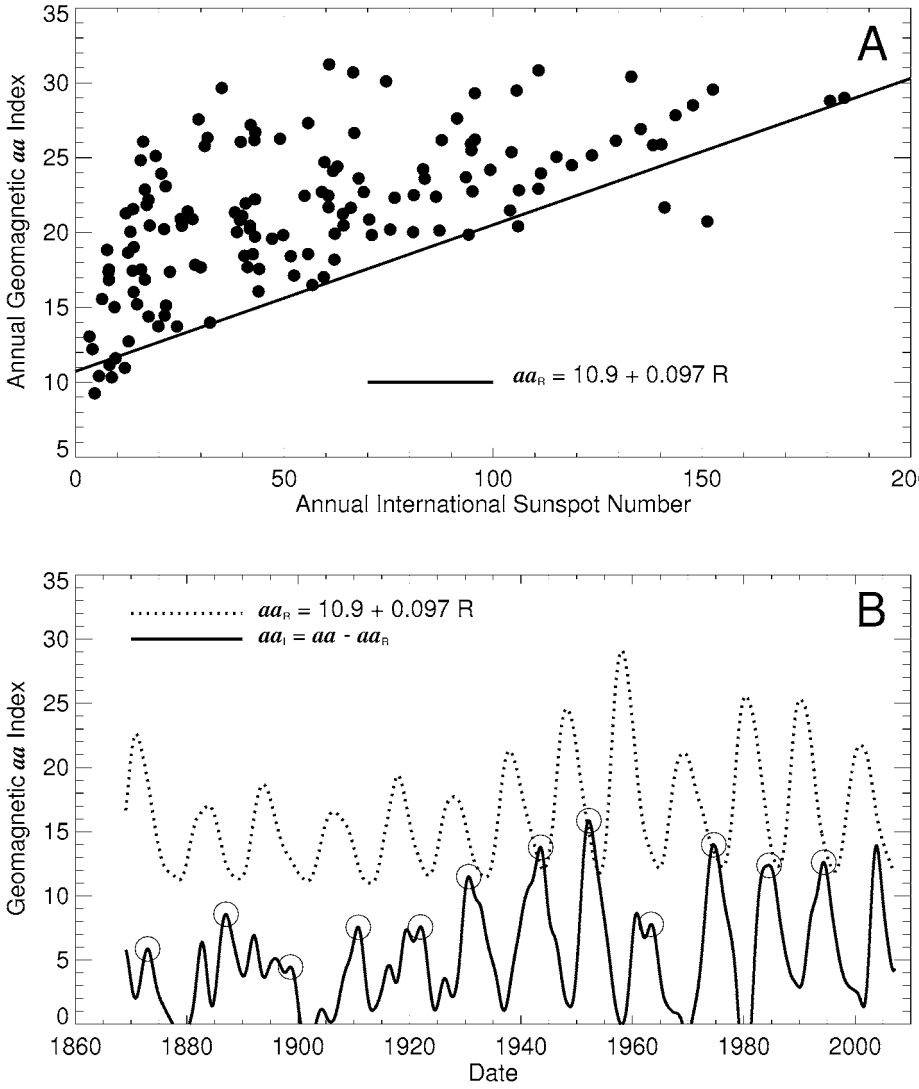


Fig. 6 Feynman's method for separating geomagnetic activity into a component, aa_R, proportional to sunspot number and the remaining "interplanetary" component, aa_I. Panel **A** shows annual values of the geomagnetic index *aa* plotted against annual values of the sunspot number. A straight line fit through the lower values as described by Hathaway and Wilson (2006) gives the solar activity component, aa_R. Panel **B** shows each component as a function of time. The (*circled*) peaks in the interplanetary component are predictors for the amplitude of the sunspot cycle (represented here by the solar activity component)

magnetic *Ap* index ≥25) during one cycle to the sum of the amplitudes of that cycle and the following cycle. Predictions for the amplitude of a sunspot cycle are available well before minimum with these two Geomagnetic Precursor methods.

Hathaway et al. (1999) tested these precursor methods by backing-up in time to 1950, calibrating each precursor method using only data prior to the time, and then using each method to predict cycles 19–22, updating the data and recalibrating each method for each

Table 1 Precursor prediction method errors (Predicted—Observed) for cycles 19–23

Prediction method	Cycle 19	Cycle 20	Cycle 21	Cycle 22	Cycle 23	RMS
Mean cycle	−97.4	−1.6	−55.4	−46.7	−6.9	54.4
Even–odd	−60.1	—	−26.7	—	61.4	52.0
Maximum–minimum	−109.7	24.9	−18.6	−8.1	5.2	51.2
Amplitude–period	−75.3	18.4	−73.5	−25.6	15.0	49.6
Secular trend	−96.4	14.6	−40.6	−25.4	18.9	49.3
Three cycle sawtooth	−96.5	14.6	−38.5	−25.4	18.8	49.0
Gleissberg cycle	−64.8	48.0	−36.9	−31.8	−0.9	42.1
Ohl's method	−55.4	−5.9	2.3	−9.1	10.5	28.7
Feynman's method	−43.3	−22.4	−1.0	−14.8	25.9	28.6
Thompson's method	−17.8	8.7	−26.5	−13.6	40.5	27.0
Combined method	−30.6	−6.9	−13.8	−14.2	33.2	22.3

remaining cycle. The results of this test were examined for both accuracy and stability (i.e. did the relationships used in the method vary significantly from one cycle to the next). An updated (including cycle 23 and corrections to the data) version of their Table 3 is given here as Table 1. The RMS errors in the predictions show that the geomagnetic precursor methods (Ohl's method, Feynman's method, and Thompson's method) consistently outperform the other tested methods. Furthermore, these geomagnetic precursor methods are also more stable. For example, as time progressed from cycle 19 to cycle 23 the Gleissberg cycle period changed from 7.5-cycles to 8.5-cycles and the mean cycle amplitude changed from 103.9 to 114.1 while the relationships between geomagnetic indicators and sunspot cycle amplitude were relatively unchanged. Hathaway et al. (1999) also noted that the prediction errors from Feynman's method were uncorrelated with those from Thompson's method so that a Combined Precursor from the average of the two would provide an improved prediction. This is shown in the final row of Table 1.

It does not appear that the minimum for cycle 24 has been reached as of this writing (June 2008). This isn't a problem for Feynman's method. Her method gives $R_{\max}(24) = 150.2 \pm 28.6$ based on a peak in the smoothed *aa* index in the fall of 2003. Currently Ohl's method gives $R_{\max}(24) = 109.2 \pm 28.7$ but this estimate will continue to fall until the minimum in the smoothed *aa* index is reached. Thompson's method gives $R_{\max}(24) = 114.4 \pm 27.0$ but this estimate will continue to rise until sunspot minimum is reached to end the counting of geomagnetically disturbed days during cycle 23. The Combined Precursor Method of Hathaway et al. (1999) gives $R_{\max}(24) = 132.3 \pm 22.3$. This estimate will also rise slightly until minimum is reached.

The physics behind the geomagnetic precursors is uncertain. The geomagnetic disturbances that produce the precursor signal are primarily due to high speed solar wind streams from low latitude coronal holes late in a cycle. Schatten and Sofia (1987) suggested that this geomagnetic activity near the time of sunspot cycle minimum is related to the strength of the Sun's polar magnetic field which is, in turn, related to the strength of the following maximum (see next section on dynamo based predictions). Cameron and Schüssler (2007) suggest that it is simply the overlap of the sunspot cycles and the Waldmeier Effect that leads to these precursor relationships with the next cycle's amplitude.

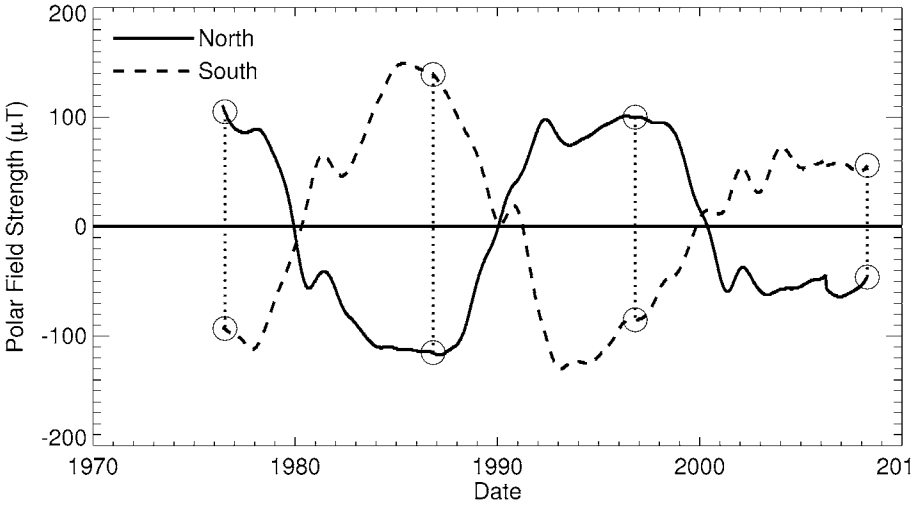


Fig. 7 Polar field strength measurements from the Wilcox Solar Observatory. The field strength poleward of 55° in the north (*solid line*) and south (*dashed line*) near the time of sunspot cycle minima (1976.5, 1986.8, 1996.8, and 2008.3—*circled values* connected by *dotted lines*) is used as a predictor for the amplitude of the following cycle

4 Upcoming Cycle Predictions with Dynamo Models

Dynamo models for the Sun's magnetic field and its evolution have led to predictions based on aspects of those models. Schatten et al. (1978) suggested using the strength of the Sun's polar field as a predictor for the amplitude of the following cycle based on the Babcock (1961) dynamo model. Good measurements of the Sun's polar field are difficult to obtain. The field is weak and predominantly radially directed and thus nearly transverse to our line-of-sight. This makes the Zeeman signature weak and prone to the detrimental effects of scattered light. Nevertheless, systematic measurements of the polar fields have been made at the Wilcox Solar Observatory since 1976 and have been used by Schatten and his colleagues to predict cycles 21–24. These polar field measurements are shown in Fig. 7. While the physical basis for these predictions is appealing, the fact that the necessary measurements are only available for the last three cycles is a distinct problem. It is unclear when the measurements should be taken. Predictions by this group for previous cycles have given different values at different times. The RMS differences between the published predictions and the observed cycle amplitudes suggest that these predictions are about as good as the geomagnetic precursor predictions. The polar fields are obviously much weaker during the current minimum. This has led to a prediction of $R_{\max}(24) = 75 \pm 8$ by Svalgaard et al. (2005)—about half the size of the previous three cycles based on the polar fields being about half as strong. Note that the stated error in this estimate is the error in the measurement of the polar field. A more reasonable error for the prediction itself, based on previous predictions, is ± 30.

In the Babcock (1961) dynamo model the polar field at minimum is representative of the poloidal field that is sheared out by differential rotation to produce the toroidal field that erupts as active regions during the following cycle. Diffusion of the erupting active region magnetic field (along with the Joy's Law tilt of these active regions) then leads to the accumulation of opposite polarity fields at the poles and the ultimate reversal of the polar fields.

Over the last decade dynamo models have started to include the effect of the Sun's meridional circulation and found that it can play a significant role in the magnetic dynamo (c.f. Dikpati and Charbonneau 1999). In these models the speed of the meridional circulation sets the cycle period and influences both the strength of the polar fields and the amplitudes of following cycles. Two predictions have recently been made based on flux transport dynamos with assimilated data—with very different results.

Dikpati et al. (2006) predicted an amplitude for cycle 24 of 150–180 using a flux transport dynamo that included a rotation profile and a near surface meridional flow based on helioseismic observations. They modeled the axisymmetric poloidal and toroidal magnetic field using a meridional flow that returns to the equator at the base of the convection zone and used two source terms for the poloidal field—one at the surface due to the Joy's Law tilt of the emerging active regions and one in the tachocline due to hydrodynamic and MHD instabilities. The diffusivity in the model is a function of depth with a surface diffusivity of 5×10^{12} cm^2 s^{-1} falling to 5×10^{10} cm^2 s^{-1} at $r = 0.9\ R_\odot$. They drive the model with a surface source of poloidal field that depends upon the sunspot areas observed since 1874. Measurements of the meridional flow speed prior to 1996 are highly uncertain (c.f. Hathaway 1996) so they maintained a constant flow speed prior to 1996 and forced each of those earlier cycles to have a constant period as a consequence. The surface poloidal source term drifted linearly from 30° to 5° over each cycle with an amplitude that depended on the observed sunspot areas. They based their prediction on the strength of the toroidal field produced in the tachocline. They found excellent agreement between this toroidal field strength and the amplitude of each of the last eight cycles (the four earlier cycles—during the initialization phase—were also well fit but not with the degree of agreement of the later cycles). The correlation they find between the predicted toroidal field and the cycle amplitudes is significantly better than that found with the geomagnetic precursors. When they kept the meridional flow speed at the same constant level during cycle 23 they found $R_{\max}(24) \sim 180$. When they allowed the meridional flow speed to drop by 40% as was seen from 1996–2002 they found $R_{\max}(24) \sim 150$ and further predicted that cycle 24 would start late.

Choudhuri et al. (2007) predicted an amplitude for cycle 24 of 80 using a similar flux-transport dynamo but with the surface poloidal field at minimum as the assimilated data. They used a similar axisymmetric model for the poloidal and toroidal fields but with a meridional flow that extends below the base of the convection zone and a diffusivity that remains high throughout the convection zone. In their model the toroidal field in the tachocline produces flux eruptions when its strength exceeds a given limit. They compare the number of eruptions to the observed sunspot numbers and use this as the predictor for cycle 24. They assimilate data by instantaneously changing the poloidal field at minimum throughout most of the convection zone to make it match the dipole moment obtained from the Wilcox Solar Observatory observations (Fig. 7). They found an excellent fit to the last three cycles (the full extent of the data) and found $R_{\max}(24) \sim 80$, in agreement with the polar field prediction of Svalgaard et al. (2005).

Criticism has been leveled against all of these dynamo-based predictions. Dikpati et al. (2006) criticized the use of polar field strengths to predict the sunspot cycle peak that follows by four years by questioning how those fields could be carried down to the low latitude tachocline in such a short time. Cameron and Schüssler (2007) produced a simplified 1D flux transport model and showed that with similar parameters to those used by Dikpati et al. (2006) the flux transport across the equator was an excellent predictor for the amplitude of the next cycle but the predictive skill was lost when more realistic parameterizations of the active region emergence were used. Yeates et al. (2008) compared an advection-dominated model like that of Dikpati et al. (2006) to a diffusion-dominated models like that

of Choudhuri et al. (2007) and concluded that the diffusion-dominated model was better because it gave a better fit to the relationship between meridional flow speed and cycle amplitude. Dikpati et al. (2008) returned with a study of the use of polar fields and cross equatorial flux as predictors of cycle amplitudes and concluded that their tachocline toroidal flux was the best indicator. Furthermore, they found that the polar fields followed the current cycle so that the weak polar fields at this minimum are due to the weakened meridional flow. The strongest criticism of these dynamo-based predictions was give by Tobias et al. (2006) and Bushby and Tobias (2007). They conclude that the solar dynamo is deterministically chaotic and thus inherently unpredictable.

5 Conclusions

Solar cycle forecasting has made great gains in the last few years. It has progressed from using statistical correlations to using numerical models for the physical processes with assimilated data from observations. Statistics are still useful and they indicate that cycle 24 will be particularly telling. It is apparent that cycle 23 will be a long cycle. As of this writing the Sun is still dominated by cycle 23 spots. This makes the period for cycle 23 at least 142 months (September 2006 to June 2008). Statistically this suggests a small cycle for cycle 24—in accordance with the predictions based on the weak polar fields (Svalgaard et al. 2005; Choudhuri et al. 2007). This statistic also makes the Dikpati et al. (2006) prediction for a large but late starting cycle extraordinary. Cycle 24 will likely show that one (or possibly both) of these models is incorrect. A large, late starting cycle would rule out polar fields as good predictors and provide strong support (given its unlikeliness) for the Dikpati et al. (2006) model. A small, late starting cycle would rule out the long (2–3 cycle) memory and small diffusivity of that same model. An average sized cycle 24 is apt to prolong the debate but, given the extremes in the predictions, could indicate problems with both—and possible agreement with the unpredictability suggested by Tobias et al. (2006).

References

H.S. Ahluwalia, The predicted size of cycle 23 based on the inferred three-cycle quasi-periodicity of the planetary index Ap. J. Geophys. Res. **103**(A6), 12,103–12,109 (1998)
H.W. Babcock, The topology of the Sun's magnetic field and the 22-year cycle. Astrophys. J. **133**, 572–587 (1961)
G.M. Brown, What determines sunspot maximum? Mon. Not. Roy. Astron. Soc. **174**, 185–190 (1976)
P.J. Bushby, S.T. Tobias, On predicting the solar cycle using mean-field models. Astrophys. J. **661**, 1289–1296 (2007)
R. Cameron, M. Schüssler, Solar cycle prediction using precursors and flux transport models. Astrophys. J. **659**, 801–811 (2007)
A.R. Choudhuri, P. Chatterjee, J. Jiang, Predicting solar cycle 24 with a solar dynamo model. Phys. Rev. Lett. **98**, 131103 (2007)
M. Dikpati, P. Charbonneau, A Babcock-Leighton flux transport dynamo with solar-like differential rotation. Astrophys. J. **518**, 508–520 (1999)
M. Dikpati, G. de Toma, P.A. Gilman, Predicting the strength of solar cycle 24 using a flux-transport dynamo-based tool. Geophys. Res. Lett. **33**, L05102 (2006)
M. Dikpati, G. de Toma, P.A. Gilman, Polar flux, cross-equatorial flux, and dynamo generated tachocline toroidal flux as predictors of solar cycles. Astrophys. J. **675**, 920–930 (2008)
W. Elling, H. Schwentek, Fitting the sunspot cycles 10-21 by a modified F-distribution density function. Sol. Phys. **137**, 155–165 (1992)
J. Feynman, Geomagnetic and solar wind cycles, 1900–1975. J. Geophys. Res. **87**(A8), 6153–6162 (1982)
W. Gleissberg, A long-periodic fluctuation of the Sun-spot numbers. The Observatory **62**, 158–159 (1939)

M.N. Gnevyshev, A.I. Ohl, On the 22-year solar activity cycle. Astron. Z. **25**, 18–20 (1948)
K.L. Harvey, O.R. White, What is solar cycle minimum? J. Geophys. Res. **104**(A9), 19,759–19,764 (1999)
D.H. Hathaway, Doppler measurements of the Sun's meridional flow. Astrophys. J. **460**, 1027–1033 (1996)
D.H. Hathaway, R.M. Wilson, E.J. Reichmann, The shape of the solar cycle. Sol. Phys. **151**, 177–190 (1994)
D.H. Hathaway, R.M. Wilson, Geomagnetic activity indicates large amplitude for sunspot cycle 24. Geophys. Res. Lett. **33**, L18101 (2006)
D.H. Hathaway, R.M. Wilson, E.J. Reichmann, A synthesis of solar cycle prediction techniques. J. Geophys. Res. **104**(A10), 22,375—22,388 (1999)
J. Javaraiah, North-south asymmetry in solar activity: predicting the amplitude of the next solar cycle. Mon. Not. R. Astron. Soc. **377**, L34–L38 (2007)
J.A. McKinnon, 'Sunspot numbers 1610–1985 based on sunspot-activity in the years 1610–1960, Rep. UAG-95 World Data Center A for Sol.-Terr. Phys., Boulder, Colo. (1987), 105 pp
A.G. McNish, J.V. Lincoln, Prediction of sunspot numbers. Trans. AGU **30**, 673–685 (1949)
A.I. Ohl, Forecast of sunspot maximum number of cycle 20. Solice Danie **9**, 84 (1966)
K.H. Schatten, S. Sofia, Forecast of an exceptionally large even-numbered solar cycle. Geophys. Res. Lett. **14**(6), 632–635 (1987)
K.H. Schatten, P.H. Scherrer, L. Svalgaard, J.M. Wilcox, Using dynamo theory to predict the sunspot number during solar cycle 21. Geophys. Res. Lett. **5**(5), 411–414 (1978)
S.K. Solanki, N.A. Krivova, M. Schüssler, M. Fligge, Search for a relationship between solar cycle amplitude and length. Astron. Astrophys. **396**, 1029–1035 (2002)
J.Q. Stewart, H.A.A. Panofsky, The mathematical characteristics of sunspot variations. Astrophys. J. **88**, 385–407 (1938)
L. Svalgaard, E.W. Cliver, Y. Kamide, Sunspot cycle 24: Smallest cycle in 100 years? Geophys. Res. Lett. **32**, L01104 (2005)
R.J. Thompson, A technique for predicting the amplitude of the solar cycle. Sol. Phys. **148**, 383–388 (1993)
S. Tobias, D. Hughes, N. Weiss, Unpredictable Sun leaves researchers in the dark. Nature **442**, 26 (2006)
M. Waldmeier, Neue eigenschaften der sonnenfleckenkurve. Astron. Mitt. Zurich **14**(133), 105–130 (1935)
M. Waldmeier, The sunspot-activity in the years 1610–1960, Zürich Schulthess, Zürich (1961), 171 pp
R.M. Wilson, D.H. Hathaway, E.J. Reichmann, An estimate for the size of cycle 23 based on near minimum conditions. J. Geophys. Res. **103**(A4), 6595–6603 (1998)
R. Wolf, Nachrichten von der Sternwarte in Bern. Sonnenflecken Beobachtungen in der zweiten Hafte des Jahres 1851. Naturf. Gesell. Bern Mitt. **229/230**, 41–48 (1852)
R. Wolf, Abstract of his latest results. Mon. Not. R. Astron. Soc. **21**, 77–78 (1861)
A.R. Yeates, D. Nandy, D.H. Mackay, Exploring the physical basis of solar cycle predictions: Flux transport dynamics and persistence of memory in advection- vs. diffusion-dominated solar convection zones. Astrophys. J. **673**, 544–556 (2008)

Coronal Magnetism: Difficulties and Prospects

Peter J. Cargill

Originally published in the journal Space Science Reviews, Volume 144, Nos 1–4, 413–421.
DOI: 10.1007/s11214-008-9446-9

Abstract Prospects for advances in understanding the properties of the coronal magnetic field are discussed. A new generation of ground-based instrumentation presents possibilities of improved direct measurements of the field (the Advanced Technology Solar Telescope: ATST) and its inference from radio observations (the Frequency Agile Solar Radiotelescope: FASR). The latter in particular promises major advances in determining the structure of the strong magnetic fields present in active regions. Interpreting observations of coronal oscillations using MHD wave models to infer a magnetic field strength has become popular. While limb observations yield field strengths compatible with those obtained from infrared spectroscopy, disc observations yield values that seem on the low side, suggesting the need for a programme of forward modelling with realistic global magnetic fields. Global magnetic field models can now provide information on the field in the corona, and towards the Earth through the solar wind. Major challenges for such modelling are the incorporation of small-scale plasma effects.

Keywords Sun corona magnetic fields

1 Introduction

Ever since the realisation almost 70 years ago that the corona was a fully ionised plasma with a temperature in excess of 10^6 K, it has been generally accepted that it was heated by the dissipation of currents associated with the Sun's magnetic field. However, the process whereby the heating occurs remains elusive. Billings (1966: p. 260) summarised the situation as follows:

P.J. Cargill (✉)
Space and Atmospheric Physics, The Blackett Laboratory, Imperial College, London SW7 2BW, UK
e-mail: p.cargill@imperial.ac.uk

P.J. Cargill
School of Mathematics and Statistics, University of St Andrews, St Andrews, Scotland KY16 9SS, UK

M.J. Thompson et al. (eds.), *The Origin and Dynamics of Solar Magnetism*.
DOI: 10.1007/978-1-4419-0239-9_21

> "The most striking aspect of the subject of magnetic fields in the corona is the frequency and variety of situations for which they are postulated, compared to the scarcity of any definite information concerning them."

and 42 years later one cannot disagree. By definite information, we mean measurements of the field strength and direction. In contrast, the magnetic field in the solar photosphere (especially sunspots) has been measured for the past 100 years (this ISSI workshop falls on the centenary of Hale's original work). Increasingly precise measurements have been made, latterly from space with the Michelson Doppler Imager (MDI: Scherrer et al. 1995) instrument on the Solar and Heliospheric Observatory (SOHO) and the more recent Solar Optical Telescope (SOT: Tsuneta et al. 2008) on Hinode. These are discussed in other papers in this volume.

What sort of magnitude can we expect the coronal field to have? Using multi-wavelength observations of the quiet, active and flaring corona, one can obtain rough estimates of what at least the minimum magnitude must be. (i) Consider an active region loop structure to have a characteristic temperature of a few MK and a density of order 10^{10} cm^{-3}. To confine such a pressure in the observed loop structures requires a field strength well in excess of 20 G. For loops in the quiet Sun, one needs $\gg$ a few G. (ii) A solar flare releases 10^{32} ergs over a cubic volume of scale perhaps 5×10^9 cm. To power the flare, one needs to dissipate a field of 140 G throughout this volume. This value is a significant under-estimate of the true field strength, since only the energy associated with the non-potential component is available for the flare. (iii) If one takes a typical radial interplanetary magnetic field at 1 AU of 5 nT, and extrapolates back to the Sun, one comes up with a field magnitude in the outer corona of a few G. So the coronal field strength must cover a very broad range, which in turn will depend on the photospheric field, and whatever complexity photospheric motions can introduce into the corona.

Direct measurements of coronal fields on the limb using the Zeeman effect are to a large extent stymied by the difficulty in actually detecting it in competition with thermal and non-thermal line broadening. Concepts that involve measurements on the limb using the Hanle effect are being developed, but need more work. Radio observations have long been seen as a promising avenue for the determination of strong coronal field magnitudes in active regions on the disk, but have significant interpretational difficulties. Future observational facilities being proposed that would use some of these techniques are discussed in Sect. 2.

Alternative approaches for determining field properties rely on theoretical knowledge of how coronal plasmas and magnetic fields are expected to behave. One carries out large-scale extrapolations based on the measured photospheric fields (and increasingly the photospheric *vector* field). Maxwell's equations and Newton's laws are solved, and give a prediction of the 3-D magnetic field at all points within a bounded space, subject to appropriate boundary conditions. In addition, the freezing of a highly conducting plasma and the magnetic field means that observed coronal plasma structures are often assumed to outline magnetic lines of force, hence images at EUV and X-ray wavelength are argued to show a global field topology. (There are spectroscopic subtleties and caveats in such an interpretation.) The newer field of coronal seismology combines theoretical properties of coronal waves with observations of oscillations to infer a magnetic field magnitude. These approaches are discussed in Sect. 3. One also needs to step back and assess how such measurements and models can actually advance the solution of major problems in the corona: why is it there, why does it flare and erupt etc. This is done in Sect. 4.

2 Approaches Based on "Direct" Measurement

Several methods have been proposed to measure the magnetic field in the corona. Some of these techniques are best used in prominences, and are beyond our scope. The report of Judge et al. (2001) is highly recommended as a summary, and is available at the time of writing at: http://www.cosmo.ucar.edu/tech-notes.jsp.

2.1 Using the Zeeman Effect

The main problem with measuring the coronal field directly lies in the difficulty in detecting the Zeeman effect at coronal temperatures. The shift of wavelength due to Zeeman splitting is: $\frac{\Delta\lambda}{\lambda} \approx (\frac{e}{4\pi m_e c^2})\bar{g}\lambda B = 4.7 \times 10^{-13}\bar{g}\lambda B$ (Landau and Lifshitz 1988; Zirin 1988) where λ is in Angstrom, B in Gauss and $\bar{g}$ the Lande-g factor. In the corona, the detectability of Zeeman splitting competes with thermal and non-thermal broadening of emission lines. We consider only thermal broadening: non-thermal broadening typically gives an additional factor of two or so (e.g. Doschek et al. 2007). Thermal broadening is described by $\frac{\Delta\lambda}{\lambda} \approx \frac{V_{ti}}{c}$, where $V_{ti}^2 = k_B T_i / m_i$ is defined by the relevant ion species: we will consider iron: $m_i = 56 m_p$. Then, the ratio of Zeeman splitting to thermal broadening is:

$$\frac{(\Delta\lambda/\lambda)_Z}{(\Delta\lambda/\lambda)_T} \approx \left(\frac{e}{4\pi m_e c}\right)\frac{\bar{g}\lambda B}{V_{ti}} = 1.4 \times 10^{-2}\left(\frac{\bar{g}\lambda B}{V_{ti}}\right)$$

Consider now a typical EUV line with wavelength 200 Å with a coronal field of 200 G at a temperature of 2 MK. This gives a ratio of 10^{-4} or so. The problem is made worse for shorter wavelengths (X-ray) and only improves for coronal emission lines in the visible and infrared.

This latter point has been long appreciated, and has lead to some noble efforts in the last decade to make direct IR coronal measurements. Lin et al. (2000) used the longitudinal Zeeman effect in an Fe XIII line (10 747 Å) to deduce a field strength of 10 and 33 G in two active regions at 0.12 and 0.15 R_s above the solar surface. Spatial averaging over a large domain was necessary. In a later paper, Lin et al. (2004) used the SOLARC coronagraph to demonstrate the imaging of relatively weak fields (a few G) approximately 0.15 R_s above an active region. It should be stressed that this is a challenging analysis, involving the extraction of very weak signals from the Stokes parameters, and long integration times (70 minutes for the case shown). This integration time is far longer than the main coronal timescale, namely the Alfven transit time across a region of the corona (1–2 min for a field of 10 G, a scale of 10^{10} cm and a density of a few 10^8 cm^{-3}). Thus, the temporal resolution needs to be better: a few seconds.

Kramer et al. (2006) have provided an overview of the limitations of this method in trying to reconstruct the coronal magnetic field. Of particular concern is the integration of the emission along an extensive line of sight, the radiation being optically thin at coronal temperatures. At the heights used, one might argue that large-scale coronal structures dominate, suggesting some uniformity along the line of sight. However, this is by no means clear, leading to the measured "field" being a convolution of many structures. Multi-point observations provide another option, though Kramer et al. (2006) note difficulties in the tomographic reconstruction of the vector field, as well as the need for the overall topology to be quite stable over a fraction of a solar rotation. Their forward modelling approach is invaluable in assessing many effects, such as noise, magnetic complexity in coronal reconstruction.

While the lack of further SOLARC results is disappointing, leaving the generality of the sole published result unclear, these observations are important in establishing an element of

truth in the coronal field strength. The importance of making coronal magnetic field measurements off the limb is such that this technique needs to be pursued. Here one looks to the perhaps sub-arc second resolution proposed for the Advanced Technology Solar Telescope (ATST: Rimmele et al. 2008).

There has been interest for some time in using the Hanle effect as a way to measure the coronal field (e.g. Trujillo-Bueno and Asensio-Ramos 2007; Raouafi et al. 2008). The technique makes use of UV emission lines (900–1300 Å), so any observations must be made from space. At the present time, work needs to be done on both the details of the atomic physics, and how any observations would be interpreted, but the payoff for success would seem to be such that a small trial mission would be appropriate soon.

2.2 Using Gyroresonance Emission

The above IR measurements are limited to being made off the limb, and a considerable distance above the solar surface, and so address problems associated with large-scale coronal structures and perhaps the inner solar wind. The strong magnetic fields of active regions require a different approach. Electrons gyrating in a magnetic field radiate at the electron cyclotron frequency: $f_{ce} = (eB/m_e c)/2\pi = 2.8 \times 10^6 |B|$ Hz, and its harmonics, where B is in Gauss. This is gyroresonant emission and for active region fields of a few hundred Gauss, gives GHz frequencies. For lower frequencies, the atmosphere blocks radio waves, so imposing a lower limit on the field strengths that can be detected. It has been argued for many years that such emission, and associated absorption, can be used to make measurements of the active region coronal magnetic field strength. While the relevant plasma physics of the emission/absorption processes is quite complex, it is also well understood (e.g. Melrose 1985; White 2004), and experienced practitioners have evolved robust techniques for getting round many of the ambiguities (e.g. which harmonic's emission is being seen).

The great advantage of such radio measurements is that they can be made from the ground, so requiring inexpensive hardware (at least compared to that flown in space). Over the years, radio arrays have been able to image the Sun at a few discrete frequencies (and hence, subject to some caveats, can measure fixed field strengths). The most striking result is that strong fields are inferred: well in excess of a kG near sunspots (which is not too surprising), but of order 500 G–1 kG at higher levels. If one refers back to the field magnitudes quoted in the Introduction that are needed to account for flares and active regions, these numbers are reassuring. Figure 1 (Lee et al. 1997) gives an overview of present day techniques where a few frequencies are used.

Clearly the use of a continuous range of frequencies corresponding to anticipated active region field strengths is desirable. The Frequency Agile Solar Radiotelescope (FASR: Bastian 2004) aims to provide frequency coverage in the range 0.1–30 GHz with a spatial resolution of one arc second. Thus, instead of having three radio maps as shown in Fig. 1, there will be many. The difficulty with FASR (and indeed with the technique in general), is that there is no absolute height information about the magnetic field. Thus, full interpretation will require a programme of surface magnetic field and velocity measurements (such as ATST can undertake), coupled with EUV and X-ray imaging, as well as modelling.

3 Estimates Based on Theory

3.1 Coronal Seismology

The now large field of coronal seismology is based on the simple idea that in a structured magnetised plasma, the MHD wave modes obey dispersion relations that relate frequency

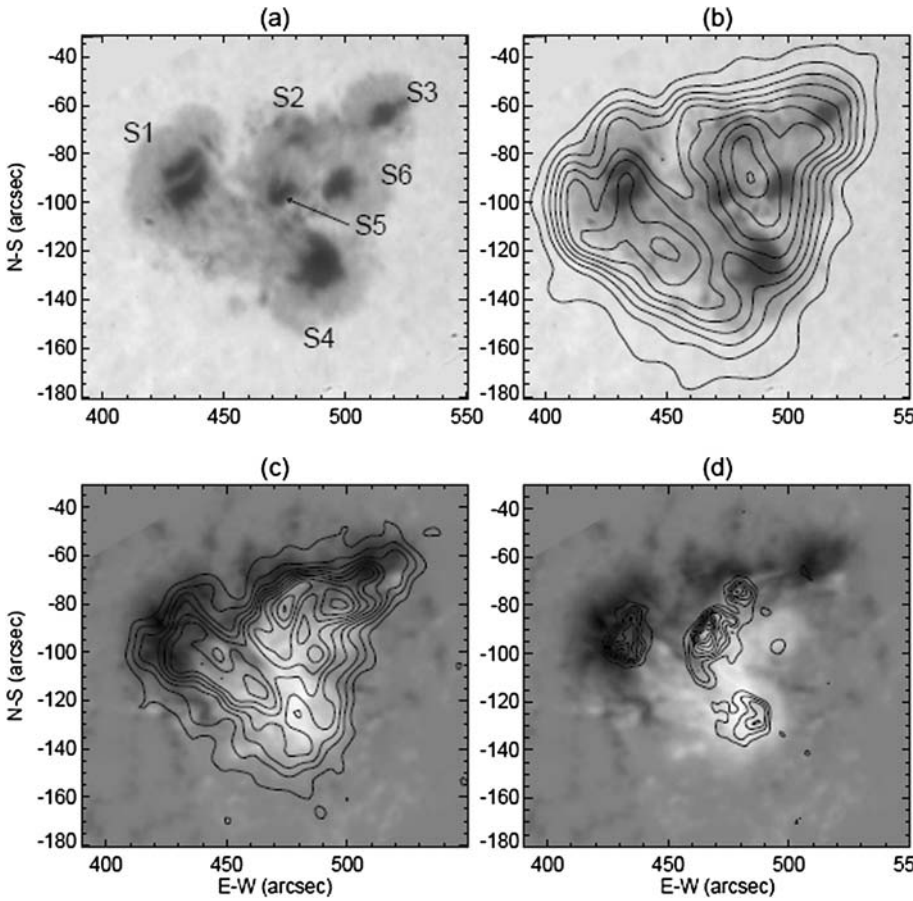


Fig. 1 VLA observations of a solar active region showing the magnetic field and temperature distribution in the corona. The white-light image **a** shows a number of spots. Panel **b** overlays contours of the VLA 5 GHz emission on the white-light image: the radio image corresponds to the electron temperature distribution on the surface in the corona where the magnetic field B equals 450 G. Panels **c** and **d** show contours of the 8.4 GHz ($B = 750$ G) and 15 GHz emission ($B = 1350$ G), respectively, overlaid on a longitudinal photospheric magnetogram. From Lee et al. (1997)

and wavelength to the plasma properties (magnetic field, density, structure). Thus if one can measure or infer some of these, the possibility exists of "backing out" a field magnitude, using a specific model for the corona. The basis for this approach was laid some time ago. Uchida (1968) argued that Moreton waves could be used to infer the magnetosonic speed in the corona (and, with a density model, the magnetic field strength). Subsequently, Edwin and Roberts (1983) published a fundamental paper that laid out the theory of wave propagation in a structured corona, with an isolated flux tube modelling a coronal loop. Structuring of the magnetic field makes the MHD wave modes dispersive, and, given suitable observations and assumptions, it was recognised that there was a possibility of solving the dispersion relations for the magnetic field intensity.

Modern EUV imaging from the TRACE and SOHO spacecraft has made this approach feasible. Nakariakov et al. (1999) analysed a loop undergoing damped transverse oscillations, apparently set into motion by a flare. By measuring the oscillation frequency and

wavelength from TRACE images, and making an educated estimate of the density, they were able to infer a magnetic field intensity of 10–20 G, though with significant error bars. Further analysis of similar events by Verwichte et al. (2004) gave field strengths in the range 9–46 G. These results provided the "proof of concept" of coronal seismology.

This is not the place to document the vast literature on this topic (see Nakariakov and Verwichte 2005), but some remarks are pertinent. Most importantly, this methodology is worth pursuing in that it has the potential of providing an independent determination of the coronal magnetic field strength. The difficulties are more subtle. The main one is the uncertainty of the technique in more complex coronal field geometries. The corona is permeated everywhere by magnetic field, with only selected regions being "illuminated" in EUV/X-ray emission. The Edwin and Roberts model has the benefit of simplicity, and, within the framework of an isolated flux tube within a uniform external magnetic field does permit further complexity (Roberts 2008). However, it seems as if it is time to move beyond it, and try more realistic and global magnetic geometries. This provides an excellent opportunity for a "forward modelling" approach wherein one introduces an oscillation locally into a global field, constructs "observables", and then sees if the inferred magnetic field strength is the real one. It would also permit a direct comparison with the simple models mentioned above. Such a calculation need only require a "linear" numerical model, and so is simpler to run and interpret than a full MHD simulation. There are numerous force-free field models in the literature (see Low 1996 for an overview) and calculations of this sort would provide confidence to a wider community that coronal seismology results are relevant.

A second worry, at least for this author, is the relative weakness of the inferred magnetic fields. The oscillating loops are often associated with large flares, and the deduced field strengths of 20–40 G seem to be rather weak, even for the post-flare phase (recall the values for flare magnetic fields noted earlier). While one might expect that the field strength will decline with altitude, observations suggest this may not be the case (e.g Lopez Fuentes et al. 2006) with the complexity associated with meso-scale currents maintaining relatively uniform loop structures. Once again forward modelling of waves in realistic fields is badly needed.

Another example of coronal seismology is the recent work of Tomczyk et al. (2007). Using the HAO Coronal Multi-channel Polarimeter (CoMP) with emission at Fe XIII 10 747 Å, they obtained a sequence of images showing oscillations in the outer solar corona at about 0.1 R_s above the surface. These were interpreted as Alfven waves propagating outward along loops. Using standard time-series analysis, they were able to infer the phase speed of the waves, and their power. The latter is too small to be interesting for coronal heating (10 erg/cm^2/s), but, with an estimate of the density, they claimed a coronal field strength there of between 8 and 26 G. This is reassuringly similar to that found by the direct IR measurements of Lin et al. at similar heights. Recent comments on this work by Van Doorsselaere et al. (2008) correctly make the point that the pure Alfven mode does not exist in a structured atmosphere, but impose a model (Edwin and Roberts 1983) that may not best describe the magnetic geometry in which these waves are present. ATST provides future opportunities for this sort of analysis.

3.2 Large-Scale Modelling and Coronal Geometry

We now pass from the realms of coronal measurements to ways of inferring coronal fields using only surface measurements. The subject of extrapolating the coronal field from photospheric measurements has a long history, beginning with the use of purely longitudinal fields and, as measurements improved (e.g. the Advanced Stokes Polarimeter: Lites et al.

1995, and the Hinode SOT), to make use of the full vector magnetic field, at least for local modelling. There have been two distinct types of model. One attempts to reconstruct an equilibrium coronal magnetic field, subject to a prescribed volume and appropriate boundary conditions, for a given photospheric magnetic field. Amari et al. (1999) give a particularly clear discussion of the mathematical complexities that arise from the requirement to solve the non-linear force-free equations as a boundary value problem. Schrijver et al. (2006) and Metcalf et al. (2008) provide a comparison of different methods, and emphasise the extremely delicate nature of the calculations. The future of this approach is very promising. The quantification of the usefulness of force-free reconstruction techniques means that workers can proceed with more confidence. In addition, the new data from Hinode/SOT of the vector magnetic field will provide the high-resolution input at the coronal base that will unquestionably shed light on fine structure, provided of course that the reconstruction techniques remain robust.

A second approach treats the evolution of the coronal magnetic field as an initial value problem. The common approach is to specify a simple initial coronal magnetic field, and then allow it to evolve in response to observed photospheric flows and/or injection of magnetic flux. While this avoids many of the concerns about mathematical ill-posedness that surround the construction of force-free equilibrium, other problems arise. Perhaps the most comprehensive series of models are those developed by Mikic, Linker and collaborators (e.g. Riley et al. 2006) which not only model the global corona, but can "couple" to models of the interplanetary medium (e.g. Odstrcil et al. 2002). Progress in this field is, to a large degree, limited by computational capacity and ingenuity in developing faster algorithms, though the continued robustness of Moore's law means that, in principle, larger simulations will always be run.

The real challenges would appear to lie elsewhere. The physics in the dynamic corona is widely believed to be determined by very small scales, well below any feasible resolution, and involving physics not dealt with by the ideal MHD equations. Magnetic dissipation and reconnection at localised current sheets has a long history (e.g. Priest and Forbes 2000), and it seems clear that the dissipation process itself involves both collisional and collisionless turbulent processes over a range of scales. The problem one is interested in addressing in the corona is that of "forced reconnection": namely how current sheets respond to strong external driving. Some have argued (see a discussion in Cargill et al. 1996) that an "ideal" MHD code will model this adequately since any forced current sheet will steepen until numerical diffusion steps in. In a time-averaged sense, one probably gets a reasonable picture of how much flux reconnects. But, in addition to being aesthetically unpleasing, this approach does not provide information of key observables: relative heating of electrons and ions, line broadening due to small-scale turbulence, particle acceleration. But, given the spatial and temporal scales required, how can one reconcile this physics with global MHD models? This is the real challenge faced by those modelling coronal fields. Full particle codes are probably not the answer since they are expensive to run, and cannot model global scales. A more promising approach would appear to be a multi-scale one, where local models of the dissipative processes are developed, and are coupled into a global code as and when required. While this must involve a great deal of basic plasma physics research, there needs to be an effort to develop simple test cases to see how such an approach might work.

4 Summary

Over the next decade we can expect the following to be achieved:

(i) Off-limb observations using IR spectroscopy with ATST will provide a large data base of the coronal magnetic field properties, subject to mitigation of line-of-sight effects.
(ii) Radio observations from FASR will produce high-resolution maps of the magnetic field at strengths expected in active regions, and, with supporting observations, can provide a 3-D magnetic field map.
(iii) More sophisticated models of coronal oscillations will validate (or otherwise) the current results of field strength from coronal seismology.
(iv) 3-D computer models of the evolution of coronal magnetic fields will be able to accommodate increasingly accurate photospheric vector input, and to better resolve the small scales expected.

It is clear that (i) and (ii) above represent the best potential observations that can be achieved at the time of the instrument design, given any financial constraints imposed on the respective projects. But it is also incumbent on us to discuss how these can solve the questions of coronal physics. The "corona-centric" questions are of course those old chestnuts: "what heats the corona?", "what causes a flare?", "why are flares such efficient particle accelerators?", "what causes a CME to lift off?", etc. And this is where things get a bit more difficult. For at least the first three of these, what is needed are very high resolution (sub-arc sec) measurements of what the magnetic field is doing in the corona (in response to the photosphere) on sub-second timescales, as well as the resultant plasma response on commensurate spatial and temporal scales over a wide range of energies. It would be presumptuous to ask IR spectroscopy as proposed for ATST, or gyroresonance imaging as proposed for FASR, to achieve this when they both already represent a massive advance over the current situation. For ATST, limb magnetic field measurements represent just one of many scientific goals of the telescope, and we know so little at present that any new results will be invaluable. With FASR, the magnetic field measurements represent the high priority science goals and, with supporting data and modelling, will provide a new view of the magnetic field of an active region.

But what is lacking are measurements of what we believe to be the fundamental scales in coronal physics, and one must question whether they can ever be made. So one should perhaps look to help from theory in answering our basic questions. It is also essential that a programme of forward modelling be carried out in order to understand what "observables" are needed to verify and refute theories. There is a pressing need for progress in combining large-scale MHD models with those of small-scale dissipation, especially to understand the feedback of the dissipation on the global dynamics. The feedback challenge is even greater in solar flares. Despite many years of work in plasma physics, we do not know how flares are such efficient accelerators (e.g. Miller et al. 1997), nor do we understand the feedback processes on magnetic fields that occur when 50% or so of the flare energy goes into particles with energies $>$ a few keV.

With a combination of high resolution vector magnetic field measurements in the photosphere, radio measurements in active regions, and a strong effort at global and local modelling, the prospects for major advances in coronal magnetic fields and their activity in the next decade are excellent.

Acknowledgements I thank Andre Balogh and Mike Thompson for the invitation to attend this workshop, and the many colleagues who have answered my questions about the topics covered in this paper. Thanks also to Jim Klimchuk for helpful comments on a draft manuscript.

References

T. Amari, T.Z. Boulmezaoud, Z. Mikic, Astron. Astrophys. **350**, 1051 (1999)

T.S. Bastian, in *Solar and Space Weather Radiophysics*, ed. by D.E. Gary, C.U. Keller (Kluwer, Dordrecht, 2004), p. 47
D. Billings, *A Guide to the Solar Corona* (Academic Press, New York, 1966)
P.J. Cargill, J. Chen, D.S. Spicer, S.T. Zalesak, J. Geophys. Res. **101**, 4855 (1996)
G.A. Doschek et al., Astrophys. J. **667**, L109 (2007)
P.M. Edwin, B. Roberts, Sol. Phys. **88**, 179 (1983)
P. Judge, R. Casini, S. Tomczyk, D.P. Edwards, E. Francis, NCAR/TN-466-STR, 2001
M. Kramer, B. Inhester, S. Solanki, Astron. Astrophys. **456**, 665 (2006)
L. Landau, E.M. Lifshitz, *Quantum Mechanics (Non-Relativistic Theory)* (Pergamon, Elmsford, 1988)
J. Lee, S.M. White, N. Gopalswamy, M.R. Kundu, Sol. Phys. **174**, 175 (1997)
H. Lin, M. Penn, S. Tomczyk, Astrophys. J. **541**, L83 (2000)
H. Lin, J.R. Kuhn, R. Coulter, Astrophys. J. **613**, L177 (2004)
B.W. Lites, B.C. Low, V. Martinez Pillet, P. Seagraves, A. Skumanich, Z.A. Frank, R.A. Shine, S. Tsuneta, Astrophys. J. **446**, 877 (1995)
M.C. Lopez Fuentes, J.A. Klimchuk, P. Demoulin, Astrophys. J. **639**, 459 (2006)
B.C. Low, Sol. Phys. **167**, 217 (1996)
D.B. Melrose, in *Solar Radiophysics*, ed. by D.J. McLean, N.R. Labrum (CUP, 1985). Chap. 9
T.R. Metcalf et al., Sol. Phys. **247**, 269 (2008)
J.A. Miller, P.J. Cargill et al., J. Geophys. Res. **102**, 14632 (1997)
V.M. Nakariakov, E. Verwichte, Living Rev. Sol. Phys. **2005**(3) (2005)
V.M. Nakariakov, L. Ofman, E.E. DeLuca, B. Roberts, J.M. Davila, Science **285**, 862 (1999)
D. Odstrcil et al., J. Geophys. Res. **107** (2002). doi:10.1029/2002JA009334
E.R. Priest, T.G. Forbes, *Magnetic Reconnection* (CUP, 2000)
N.-E. Raouafi, S.K. Solanki, T. Wiegelmann, arXiv:0801.2202v1 (2008)
P. Riley, J.A. Linker, Z. Mikić, R. Lionello, S.A. Ledvina, J.G. Luhmann, Astrophys. J. **653**, 1510 (2006)
T. Rimmele et al., Adv. Space Res. **42**(1), 78 (2008)
B. Roberts, in *IAU Symp*. **247** (2008)
P. Scherrer et al., Sol. Phys. **162**, 129 (1995)
C.J. Schrijver et al., Sol. Phys. **235**, 161 (2006)
S. Tomczyk, S.W. McIntosh, S.L. Keil, P.G. Judge, T. Schad, D.H. Seeley, J. Edmondson, Science **317**, 1192 (2007)
J. Trujillo-Bueno, A. Asensio-Ramos, Astrophys. J. **655**, 642 (2007)
S. Tsuneta et al., Sol. Phys. **249**, 195 (2008)
Y. Uchida, Sol. Phys. **4**, 30 (1968)
T. Van Doorsselaere, V.M. Nakariakov, E. Verwichte, Astrophys. J. **676**, L73 (2008)
E. Verwichte, V.M. Nakariakov, L. Ofman, E.E. Deluca, Sol. Phys. **223**, 77 (2004)
S.M. White, in *Solar and Space Weather Radiophysics*, ed. by D.E. Gary, C.U. Keller (Kluwer, Dordrecht, 2004), p. 89
H. Zirin, *Astrophysics of the Sun* (CUP, 1988)

ISSI Workshop on Solar Magnetism: Concluding Remarks

Jean-Paul Zahn

Originally published in the journal Space Science Reviews, Volume 144, Nos 1–4, 423–428.
DOI: 10.1007/s11214-009-9492-y

Keywords Sun · Magnetism · Dynamo

1 Introduction

Spectacular progress has been recently achieved in observing the Sun, thanks in particular to space-borne instruments that allow to capture the far UV and X-ray spectrum. Yokkoh, EIT on SoHO, TRACE, RHESSI and HINODE have brought us splendid pictures and movies of the active Sun, which have again been displayed during this workshop. It is well established that this activity is due to the magnetic field, ever since such a field has been detected in sunspots one hundred years ago (Hale 1908); without that field the Sun would be a rather dull object! But we still don't have a satisfactory explanation for the origin of this magnetic field, and that is why ISSI organized this workshop entirely devoted to the solar magnetism.

It is impossible to summarize such an intense week in just a few pages, especially for someone like me who is not an expert in this field (I suspect that this was the reason for asking me to deliver these concluding remarks!). Therefore I choose to focus here on what was clearly the central issue of our meeting, namely the origin of solar magnetism. I am aware that this is not fair for many other fascinating subjects that have been discussed, such as for instance the structure of sunspots, the fibril nature of the magnetic field, the mechanism of coronal mass ejections, and I hope that the contributing authors will forgive me.

The current paradigm for the generation of the solar field is still that proposed by Gene Parker in his seminal paper of 1955 (Parker 1955). It was so already at the first symposium I ever attended, in 1963, that of Rottach-Egern on stellar and solar magnetic fields. Turbulent convection and differential rotation play a key role in Parker's scheme. A large-scale poloidal field is sheared by the differential rotation into a large-scale toroidal field, which emerges here and there at the surface in the form of bipolar active regions. That toroidal

J.-P. Zahn (✉)
LUTH, Observatoire de Paris, 92195 Meudon, France
e-mail: jean-paul.zahn@obspm.fr

M.J. Thompson et al. (eds.), *The Origin and Dynamics of Solar Magnetism*.
DOI: 10.1007/978-1-4419-0239-9_22

field is twisted back by cyclonic motions into a poloidal component of opposite polarity. These Ω and α mechanisms, as they have been called respectively, operate somewhere in the convection zone or in its vicinity, but not necessarily at the same place. It is commonly believed that the Ω mechanism should be seated in the tachocline, since that layer possesses both a stable stratification and a strong shear, whereas the α-mechanism could operate either in the bulk of the convection zone, according to Parker's original view, or at the surface, as in the Babcock-Leighton scenario.

Although this picture is certainly oversimplified, it served as the reference throughout our workshop. The reason is that it offers a straightforward interpretation for many observed properties of the solar cycle, such as Hale's polarity rules: active regions are dipolar and aligned roughly in longitude, and they have opposite polarities in each hemisphere, which alternate between successive sunspot cycles. Moreover, the α–Ω dynamo can be formulated easily as a two-dimensional model involving the differential rotation and the large-scale magnetic field, with some adequate parametrization of the action of the small scales.

2 Mean Field Dynamo

During our workshop, Nigel Weiss reviewed the elements of the mean field dynamo (Weiss and Thompson 2009). In mean field models, the full set of equations that govern the dynamics and the evolution of the large-scale magnetic field is trimmed to just two equations that describe the conversion of poloidal field into toroidal field, and vice-versa. The large-scale flows are imposed: in its original form, the mean field model involved only the differential rotation, but meridional circulation was later added to reproduce the observed migration of the activity belt to the equator, along the cycle (yielding what is called the butterfly diagram). The results depend heavily on how these large-scale flows are imposed; the duration of the cycle, for instance, is determined by the speed of the meridional circulation. Fortunately, as Mike Thompson remarked, some guidance is provided by seismic sounding and by surface flow measurements. But it is not clear, for instance, how deep the meridional circulation should dig into the radiation zone.

The small-scale motions are not implemented as such, but through their postulated effect on the large-scale fields; it is assumed that, due to their cyclonicity, they create a mean electromotive force, which may be described by a tensor that is related to the helicity of these small-scale flows (Steenbeck et al. 1966). The Ohmic diffusion of the field is expected to be enhanced by the turbulence, which calls for the introduction of a turbulent diffusivity. More sophisticated models also attempt to account for a non-Boussinesq effect that is clearly present in 3D simulations, namely the mean fields being pumped down by the turbulent plumes.

In its kinematic version, the problem is linear in magnetic field, and the parameters are adjusted such as to avoid exponential growth or decay; no prediction can then be made on the strength of the magnetic field. One way to introduce some non-linearity is to implement the so-called α-quenching, which is supposed to represent the feedback of the magnetic field on the small-scale motions that are responsible for the α-effect, but it is not clear how that quenching should be implemented (see Brandenburg 2009).

Mean field models of the flux transport type are now being used also to try to forecast the solar cycles, as was described by Mausumi Dikpati and David Hathaway (Dikpati and Gilman 2009; Hathaway 2009). Peter Gilman went as far as to announce that 'the age of solar cycle prediction using dynamo models has begun'. However his optimism is not shared by everybody, partly because very simple model equations can easily yield chaotic solutions, as

was illustrated by Ed Spiegel (2009). And the Sun may well exhibit such chaotic behavior, as hinted by the Maunder minimum. More work is needed to estimate how far these predictions based on mean field models actually reach, and to check whether the trajectories in phase space can be corrected by data assimilation, much like what is done in weather forecasting, a procedure that was presented by one of the junior participants, Laurène Jouve.

Weiss and Thompson (2009) concluded their critical review by stating that such mean field models can 'yield plausible results when the arbitrary parameters are carefully tuned', but they insisted that these models, although certainly *instructive*, are only *illustrative* (the italics are theirs). And Parker (2009) warned that this is not sufficient for a scientific understanding of the solar dynamo, whereas Peter Gilman stated the contrary. Parker even added that 'given four free parameters you can provide a gratifying fit to the New York skyline'. Thus no consensus could be reached on that issue.

3 Dynamo Action at Multiple Scales?

The Sun is a multiscale object, a fact that has been stressed again by several participants. For Åke Nordlund (cf. Wedemeyer-Böhm et al. 2009), the scales of the velocity field are all linked together since they show self-similarity: he even made the provocative statement that one should forget about granulation, supergranulation, etc. But this opinion was disputed by Nadège Meunier (cf. Meunier and Zhao 2009): through granule tracking over a large field of view, she and her collaborators found that supergranulation produces clearly a peak in the kinetic energy spectrum, whose height increases with the duration of the observation run (Rieutord et al. 2008). On the other hand, although the frequency distribution of magnetic flux is rather smooth over the whole range of scales (cf. van Driel and Culhane 2009), magnetic field is organized in distinct features, such as active regions, ephemeral regions and the inter-network field. Jan Stenflo recalled that a substantial fraction of the magnetic energy could be hidden in unresolved fields, that cannot be detected by the Zeeman effect—some estimates of that hidden component have been obtained through the Hanle effect (see de Wijn et al. 2009).

Is this small-scale field due to genuine local dynamo action, which according to Nordlund could occur at every depth? Steve Tobias (2009) reminded us of the increasing evidence that almost any turbulent flow generates small-scale magnetic fields, provided the magnetic Reynolds number is large enough—a condition that is amply fulfilled in the solar convection zone. And such small-scale dynamo action seems indeed to be present in the Sun—it probably manifests itself in the magnetograms by the so-called 'pepper and salt' pattern that covers the quiet regions. But will it actually contribute to the global cyclic dynamo?

The answer may be given by the careful analysis of flux emergence. The beautiful TRACE and HINODE movies presented by Allan Title demonstrate that flux emerges everywhere and that reconnection occurs everywhere too, but one is left asking whether that field is produced locally, or much deeper down. Concerning the active regions, with their bipolar structure, the origin is clearly a large-scale toroidal field generated in the deep interior, where it is probably fragmented in tubes. An important clue was given by Bruce Lites (2009): according to him, the active region filaments result from emerging twisted flux systems, much as had been described by Leka et al. (1996), as we were reminded by van Driel. Sasha Kosovichev (2009) too confirmed from local helioseismology that the magnetic flux is already concentrated in strong field structures in the near-surface layers. Shall we conclude that these twisted structures have been formed well below, where the convective motions feel the Coriolis force, or are they the result of some local MHD instability? Another hint

was provided by Saku Tsunata, who showed that the 'patchy' polar field, which can reach kilogauss strength, is of the same polarity as the global field.

To summarize, we have no clear idea yet about the role of the small-scale magnetic fields. Do they play an active part in the global dynamo, or are they just a by-product, an epiphenomenon? Are the small-scale cyclonic motions, that are expected in the rotating convection zone, sufficient to twist the poloidal field into a toroidal one, without genuine local dynamo action? The question was left open.

4 Global Simulations

There was wide agreement, among the participants, that one had to go beyond mean field models: it is the three-dimensional global simulations, which are now built with the massive parallel supercomputers, that represent the future, as Nigel Weiss put it (see Thompson and Weiss 2009). Our colleagues geophysicists have laid the path, and they are quite a bit ahead of us, as was demonstrated by Ulrich Christensen (2009). The latest results obtained for the Sun were discussed by Sacha Brun and Matthias Rempel (Brun and Rempel 2009). Provided the resolution is high enough to allow for sufficiently 'turbulent' flows, these simulations now succeed in rendering the differential rotation as observed in the Sun, with a fast equator and slow poles, and with the rotation rate varying little with depth in the convection zone. The solutions display also meridional circulation, in the poleward direction at low latitude, as observed near the surface.

Of course, these models have their limitations. The numerical domain cannot include yet the near surface layers, because the local scale-height becomes there too short and the convective eddies too small to be resolved. Hence present calculations do not encompass the upper shear layer revealed by helioseismology, which might also play a role in the dynamo loop. More importantly, most simulations do not extend into the radiation zone below, because thermal relaxation proceeds very slowly there; the tachocline is then mimicked by appropriate boundary conditions, which are known to have a non negligible impact on the resulting flows.

What about the magnetic field? Above a magnetic Reynolds number of about 400, the convection zone exhibits strong dynamo action, on all convective scales, with a magnetic energy that reaches $1/10$ of the total kinetic energy (including that of the differential rotation). But in the earliest of such simulations there was no sign of a global field: the mean field (averaged in longitude) was much weaker than the fluctuating component (Brun et al. 2004). This has changed recently with the implementation of a tachocline layer, as explained by Brun: the turbulent poloidal field is then pumped by convective downdrafts into that stable layer, where it is sheared and organized into a large-scale toroidal field (Browning et al. 2006). Thus the outcome is much as predicted by the α–Ω dynamo, although there the poloidal field is smoothed in the convection zone and it is that large-scale component which is then sheared by the differential rotation. What is still missing in the dynamo loop is the process that would regenerate the poloidal field, although it seems that all ingredients are gathered that are deemed to play a role in the α mechanism: convection, rotation and meridional circulation. Does it mean that this process occurs in the Sun on scales that are still unresolved by the numerical simulations?

What is lacking too is the transport of the toroidal field to the surface, where it would emerge to form the observed bipolar active regions. Since this process does not occur spontaneously in the simulations, for reasons that are not elucidated yet, it is triggered by putting a flux tube at the base of the convection zone. The same 3D global codes can be used to

model the rise of that tube, as was described by Jouve. Previously, the action of the convective environment was ascribed in 2D to a drag force opposite to the velocity of the tube, and the conclusions were somewhat dubious. Now these flux tubes are experiencing the actual impact of descending plumes, which is directed vertically, and they emerge at lower latitudes. However, there are still some discrepancies with the observations, such as the too large tilt of the emerging dipoles with respect to the latitude circle; this is probably the consequence of having to impose a strong twist to the flux tube, to prevent it from splitting in two counter vortices while it rises to the surface.

These 3D simulations also demonstrate how the convective motions pump down the magnetic field, a non-Boussinesq and non-diffusive effect that one attempts to represent in the mean field models by the so-called Λ terms (cf. Rüdiger 1989). Furthermore, they raise the question of what saturates the growth of the magnetic field: how can it be something like the α quenching one invokes in the mean field dynamo, when the α effect seems to be absent here? There is still much to understand in these calculations!

5 The Key: Comparing the Sun with Other Stars

If I may express a slight regret, it is that we didn't pay enough attention during our workshop to the behavior of other stars; this could certainly help us to comprehend the solar dynamo, by disentangling the roles of rotation, differential rotation and mass (hence the depth of the convection zone).

It has been known for a while already that chromospheric activity, which is tightly correlated with the magnetic field, is present in all stars possessing an outer convection zone; the activity level increases with the rotation rate, and therefore it declines with age. This activity tends to fluctuate erratically in the younger, more active stars, whereas the older, less active stars display regular cycles.

More detailed information is now gathered by the new generation stellar spectropolarimeters, specifically Espadons at the Canada–France–Hawaii telescope and Narval at Pic du Midi (Donati 2003). These powerful instruments allow to map the large-scale surface field, together with the differential rotation, through Doppler–Zeeman imaging. A large survey is under way, to check whether the cycles displayed by the chromospheric activity are associated with field reversals, as in the Sun. (Such a reversal has just been observed, but in a fast rotating 1.3 $M_\odot$ star.) The first results of the survey confirm that the strength of the magnetic field increases indeed with rotation; furthermore, the slow rotators show a larger proportion of poloidal field than the fast rotators, in which the field is mainly toroidal (Petit et al. 2008). The monitoring is pursued to detect cycles and—who knows?—catch a star in its Maunder minimum.

Another survey established that these trends are observed also among the less massive M stars. Moreover it revealed that fully convective stars have strong large-scale poloidal fields, and little (surface) differential rotation. Thus there is no need of a tachocline to generate large-scale fields, as one could be tempted to conclude from the solar case. But is it required to produce cycles?

To conclude, the picture that emerges is that of a great variety of stellar dynamos, some in which the poloidal field dominates and others where it is the toroidal field, some with more differential rotation and others with less, some exhibiting cycles and others not, etc. And it is quite possible that a star may switch from one regime to another, as was perhaps the case during the Maunder minimum.

Therefore I believe that, to solve the problem of the solar dynamo, we will benefit more in the future from the comparison between stars of different mass and rotation rate, than

from pursuing the detailed study of all the manifestations of the solar activity. But how can one possibly resist the temptation of adding new terms to the vocabulary, such as 'gappy', 'Sven's mess', 'straw', 'clapotisphere', and 'fluctuosphere'?

Acknowledgements In the name of all participants, I wish to express again our warm thanks to André Balogh, to Roger Bonnet, to ISSI and its entire staff, for organizing this most interesting workshop!

References

A. Brandenburg, Advances in theory and simulations of large-scale dynamos. Space Sci. Rev. (2009, this issue)
M.K. Browning, M.S. Miesch, A.S. Brun, J. Toomre, Dynamo action in the solar convection zone and tachocline: pumping and organization of the toroidal fields. Astrophys. J. **648**, L157 (2006)
A.S. Brun, M. Rempel, Large scale flows in the solar convection zone. Space Sci. Rev. (2009, this issue)
A.S. Brun, M.S. Miesch, J. Toomre, Global-scale turbulent convection and magnetic dynamo action in the solar envelope. Astrophys. J. **614**, 1073 (2004)
U. Christensen, Planetary dynamos from a solar perspective. Space Sci. Rev. (2009, this issue)
G. de Wijn, J.O. Stenflo, S.K. Solanki, S. Tsuneta, Small-scale solar magnetic fields. Space Sci. Rev. (2009, this issue)
M. Dikpati, P.A. Gilman, Flux-transport solar dynamos. Space Sci. Rev. (2009, this issue)
J.-F. Donati, ESPaDOnS: An Echelle SpectroPolarimetric device for the observation of stars at CFHT. ASP Conf. Ser. **307**, 41 (2003)
G.E. Hale, On the probable existence of a magnetic field in sun-spots. Astrophys. J. **28**, 315 (1908)
D. Hathaway, Solar cycle forecasting. Space Sci. Rev. (2009, this issue)
A.G. Kosovichev, Photospheric and subphotospheric dynamics of emerging magnetic flux. Space Sci. Rev. (2009, this issue)
K.D. Leka, R.C. Canfield, A.N. McClymont, L. van Driel-Gesztelyi, Evidence for current-carrying emerging flux. Astrophys. J. **462**, 547 (1996)
B.W. Lites, The topology and behavior of magnetic fields emerging at the solar photosphere. Space Sci. Rev. (2009, this issue)
N. Meunier, J. Zhao, Observations of photospheric and interior dynamics and magnetic fields: from large-scale to small-scale flows. Space Sci. Rev. (2009, this issue)
E.G. Parker, Hydromagnetic dynamo models. Astrophys. J. **122**, 293 (1955)
E.G. Parker, Solar magnetism: the state of our knowledge and ignorance. Space Sci. Rev. (2009, this issue)
P. Petit, B. Dintrans, S.K. Solanki, J.-F. Donati, M. Aurière, F. Lignières, J. Morin, F. Paletou, J. Ramirez Velez, C. Catala, R. Fares, Toroidal versus poloidal magnetic fields in Sun-like stars: a rotation threshold. Mon. Not. R. Astron. Soc. **388**, 80 (2008)
M. Rieutord, N. Meunier, T. Roudier, S. Rondi, F. Beigbeder, L. Parès, Solar supergranulation revealed by granules tracking. Astron. Astrophys. **479**, L17 (2008)
G. Rüdiger, *Differential Rotation and Stellar Convection—Sun and the Solar Stars* (Akademie, Berlin, 1989)
E.A. Spiegel, Chaos and intermittency in the solar cycle. Space Sci. Rev. (2009, this issue)
M. Steenbeck, F. Krause, K.H. Rädler, A calculation of the mean electromotive force in an electrically conducting fluid in turbulent motion under the influence of Coriolis forces. Zeitschrift Naturforschung A **21**, 369 (1966)
M.J. Thompson, N.O. Weiss, The solar dynamo. Space Sci. Rev. (2009, this issue)
S. Tobias, The role of penetration, rotation and shear on convective dynamos. Space Sci. Rev. (2009, this issue)
L. van Driel, J.L. Culhane, Magnetic flux emergence, activity, eruptions and magnetic clouds: following magnetic field from the Sun to the heliosphere. Space Sci. Rev. (2009, this issue)
S. Wedemeyer-Böhm, A. Lagg, Å. Nordlund, Coupling from the photosphere to the chromosphere and the corona. Space Sci. Rev. (2009, this issue)
N.O. Weiss, M.J. Thompson, Space Sci. Rev. (2009, this issue)

Space Science Series of ISSI

1. R. von Steiger, R. Lallement and M.A. Lee (eds.): *The Heliosphere in the Local Interstellar Medium*. 1996 ISBN 0-7923-4320-4
2. B. Hultqvist and M. Øieroset (eds.): *Transport Across the Boundaries of the Magnetosphere*. 1997 ISBN 0-7923-4788-9
3. L.A. Fisk, J.R. Jokipii, G.M. Simnett, R. von Steiger and K.-P. Wenzel (eds.): *Cosmic Rays in the Heliosphere*. 1998 ISBN 0-7923-5069-3
4. N. Prantzos, M. Tosi and R. von Steiger (eds.): *Primordial Nuclei and Their Galactic Evolution*. 1998 ISBN 0-7923-5114-2
5. C. Fröhlich, M.C.E. Huber, S.K. Solanki and R. von Steiger (eds.): *Solar Composition and its Evolution – From Core to Corona*. 1998 ISBN 0-7923-5496-6
6. B. Hultqvist, M. Øieroset, Goetz Paschmann and R. Treumann (eds.): *Magnetospheric Plasma Sources and Losses*. 1999 ISBN 0-7923-5846-5
7. A. Balogh, J.T. Gosling, J.R. Jokipii, R. Kallenbach and H. Kunow (eds.): *Co-rotating Interaction Regions*. 1999 ISBN 0-7923-6080-X
8. K. Altwegg, P. Ehrenfreund, J. Geiss and W. Huebner (eds.): *Composition and Origin of Cometary Materials*. 1999 ISBN 0-7923-6154-7
9. W. Benz, R. Kallenbach and G.W. Lugmair (eds.): *From Dust to Terrestrial Planets*. 2000 ISBN 0-7923-6467-8
10. J.W. Bieber, E. Eroshenko, P. Evenson, E.O. Flückiger and R. Kallenbach (eds.): *Cosmic Rays and Earth*. 2000 ISBN 0-7923-6712-X
11. E. Friis-Christensen, C. Fröhlich, J.D. Haigh, M. Schüssler and R. von Steiger (eds.): *Solar Variability and Climate*. 2000 ISBN 0-7923-6741-3
12. R. Kallenbach, J. Geiss and W.K. Hartmann (eds.): *Chronology and Evolution of Mars*. 2001 ISBN 0-7923-7051-1
13. R. Diehl, E. Parizot, R. Kallenbach and R. von Steiger (eds.): *The Astrophysics of Galactic Cosmic Rays*. 2001 ISBN 0-7923-7051-1
14. Ph. Jetzer, K. Pretzl and R. von Steiger (eds.): *Matter in the Universe*. 2001 ISBN 1-4020-0666-7
15. G. Paschmann, S. Haaland and R. Treumann (eds.): *Auroral Plasma Physics*. 2002 ISBN 1-4020-0963-1
16. R. Kallenbach, T. Encrenaz, J. Geiss, K. Mauersberger, T.C. Owen and F. Robert (eds.): *Solar System History from Isotopic Signatures of Volatile Elements*. 2003 ISBN 1-4020-1177-6
17. G. Beutler, M.R. Drinkwater, R. Rummel and R. von Steiger (eds.): *Earth Gravity Field from Space – from Sensors to Earth Sciences*. 2003 ISBN 1-4020-1408-2
18. D. Winterhalter, M. Acuña and A. Zakharov (eds.): *"Mars" Magnetism and its Interaction with the Solar Wind*. 2004 ISBN 1-4020-2048-1
19. T. Encrenaz, R. Kallenbach, T.C. Owen and C. Sotin: *The Outer Planets and their Moons* ISBN 1-4020-3362-1
20. G. Paschmann, S.J. Schwartz, C.P. Escoubet and S. Haaland (eds.): *Outer Magnetospheric Boundaries: Cluster Results* ISBN 1-4020-3488-1
21. H. Kunow, N.U. Crooker, J.A. Linker, R. Schwenn and R. von Steiger (eds.): *Coronal Mass Ejections* ISBN 978-0-387-45086-5

22. D.N. Baker, B. Klecker, S.J. Schwartz, R. Schwenn and R. von Steiger (eds.): *Solar Dynamics and its Effects on the Heliosphere and Earth* ISBN 978-0-387-69531-0
23. Y. Calisesi, R.-M. Bonnet, L. Gray, J. Langen and M. Lockwood (eds.): *Solar Variability and Planetary Climates* ISBN 978-0-387-48339-9
24. K.E. Fishbaugh, P. Lognonné, F. Raulin, D.J. Des Marais, O. Korablev (eds.): *Geology and Habitability of Terrestrial Planets* ISBN 978-0-387-74287-8
25. O. Botta, J.L. Bada, J. Gomez-Elvira, E. Javaux, F. Selsis, R. Summons (eds.): *Strategies of Life Detection* ISBN 978-0-387-77515-9
26. A. Balogh, L. Ksanfomality, R. von Steiger (eds.): *Mercury* ISBN 978-0-387-77538-8
27. R. von Steiger, G. Gloeckler, G.M. Mason (eds.): *The Composition of Matter* ISBN 978-0-387-74183-3
28. H. Balsiger, K. Altwegg, W. Huebner, T.C. Owen, R. Schulz (eds.): *Origin and Early Evolution of Comet Nuclei, Workshop honouring Johannes Geiss on the occasion of his 80th birthday* ISBN 978-0-387-85454-0
29. A.F. Nagy, A. Balogh, T.E. Cravens, M. Mendillo, I. Mueller-Wodarg (eds.): *Comparative Aeronomy* ISBN 978-0-387-87824-9
30. F. Leblanc, K.L. Aplin, Y. Yair, R.G. Harrison, J.P. Lebreton and M. Blanc (eds.): *Planetary Atmospheric Electricity* ISBN 987-0-387-87663-4
31. J.L. Linsky, V. Izmodenov, E. Möbius, R. von Steiger (eds.): *From the Outer Heliosphere to the Local Bubble: Comparison of New Observations with Theory* ISBN 978-1-4419-0246-7
32. M.J. Thompson, A. Balogh, J.L. Culhane, Å. Nordlund, S.K. Solanki, J.-P. Zahn (eds.): *The Origin and Dynamics of Solar Magnetism* ISBN 978-1-4419-0238-2

Springer – Dordrecht / Boston / London