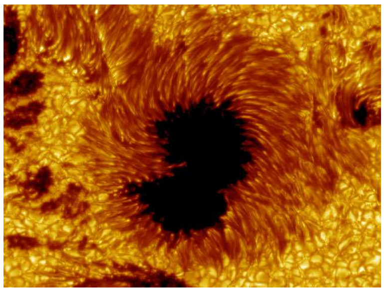
2

Solar magnetic fields.

History, tragedy or comedy?

Dr. Philip Judge (High Altitude Observatory)
National Center for Atmospheric Research, PO Box 3000, Boulder CO 80307-3000, USA

"Double, double, toil and trouble; Fire burn, and cauldron bubble!" – Three Witches, MacBeth



Abstract

Many children know that the Sun sometimes has spots. Sunspots were certainly known to ancient civilizations. Even so, it is sobering that we still do not have a unique answer to the simple question:

"Why, according to the basic laws of physics, is the Sun obliged to form a sunspot?"
– E. N. Parker, ca. 1990 (paraphrased)

This chapter summarizes how magnetic fields make their presence known in the spectra of solar plasmas, but the main purpose is to point to properties of solar magnetic fields that are surprising from the point of view of elementary physical ideas. These ideas include the physics behind the regeneration of magnetic fields that appears to require "dynamo" action and magnetic field dissipation, the role of magnetic field topology in highly conducting plasmas,

and the energetic and dynamic effects of magnetic field rearrangement and dissipation that lead to nearly all phenomena of interest to modern human life. As a leitmotif, and in contrast to those who draw attention to obvious complexities in the Sun's magnetic field, I ask another question: Why does the Sun's magnetism show so much order in the presence of such chaos?

2.1 Introductory Remarks

When asked to review this subject for this Winter School, I confess to being a little surprised, for there are many solar physicists more qualified than I, owing to their particular areas of specialization, who can discuss both observations and theory relating to solar magnetism. In recent years I have tried to understand the physical problems presented by modern solar observations, and have offered some short courses and lectures which attempt to identify some elementary problems. By elementary, I do not mean "simple", but rather those building blocks of a complex nonlinear system that a graduate student in physics might understand and retain. To me, it seems that modern solar physics boils down to one thing: the evolution of solar magnetic fields. So, this chapter is going to be concerned with explaining elementary aspects of the complex interaction of magnetic fields and plasma in the Sun, from the point of view of Newton, and Maxwell.

Right off the bat, I draw attention to a curious but central issue, captured in one of the best and exquisite images yet obtained of a sunspot (Figure 2.1). Sunspots exist in spite of the fact that 100% of the solar luminosity is carried by turbulent convection that extends across the outer 30% of the solar interior. The topmost convective layer is seen clearly in the figure as cellular structures outside the sunspot itself. Why is the Sun obliged to produce such a thing, such order from chaos?

Sunspots exhibit order in many other ways. In the 1800s Schwabe found that sunspots come and go cyclically with periods of 11 years. The magnetic nature of sunspots was discovered by Hale (1908), a decade or so after Zeeman discovered the effect of magnetic fields on spectral line intensities. Hale *et al.* (1919) later reported that sunspot magnetic fields change polarity from one cycle to the next, so that the Sun exhibits a 22 year cycling magnetic variation. Following the development of magnetohydrodynamics ("MHD", e.g. Alfven, 1950), sunspots are the most obvious manifestation of a large scale MHD "dynamo". Various datasets show that the Sun's cycling magnetism is not atypical when compared to stars. Armed with powerful telescopes on the ground and in space and with recordings of solar magnetism encoded in paleo-climate records, we have many more observations

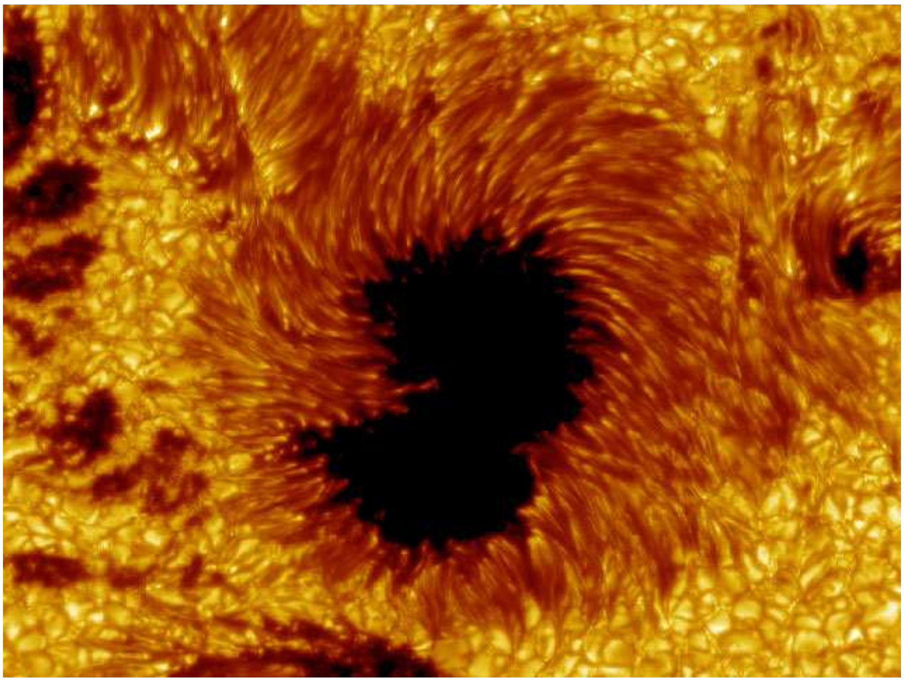


Fig. 2.1. The mystery of sunspots is represented in this figure that appeared in *Sky and Telescope* magazine on July 23 2003, an image from the Swedish Solar Telescope, a project led by G. Scharmer. Neither the nature of the umbra (darkest regions) nor penumbra (filamentary structures) is really understood, and the origin of why the magnetic field should emerge in such a concentrated form remains a mystery. The contrast between the large-scale order in the penumbral filament patterns and the turbulent convection outside the sunspot is dramatic.

of solar magnetic fields, from time scales exceeding 10^4 years to less than 1 second.

In the 1930s and 1940s, the Sun continued to surprise us, when spectroscopy showed that the temperature of plasma above the visible surface *increases* with distance from the Sun's center. This was in apparent violation of the second law of thermodynamics, and so not to be taken lightly. The obvious observational manifestation of this increased temperature is the solar corona. There is of course no such violation- instead we must invoke non-thermal processes for which thermodynamics does not apply, in contrast to the thermal modes of energy transport of radiative transport and convection present in the Sun's interior. In the latter half of 20^{th} century, scientists have trodden a rather tortuous path towards the conclusion that the dissipation of energy associated with solar magnetism is intrically related to the formation of a corona.

Below we will look at the behavior of solar magnetism on large scales, meaning global scales on the order of the solar radius $R_\odot$, and those $\sim 0.01-0.1R_\odot$ that are characteristic of sunspots and active regions. Accordingly we will use the framework of MHD. This does not mean that we can ignore small scales- quite the contrary, one characteristic of MHD (and hydrodynamics) is the development of small scales and a dynamical coupling across all physical scales. We speak of "turbulence".

Figure 2.2 shows more "order from chaos"– namely the relatively well--organized thermal structure of the solar corona on global scales. In part this results from the non-linear dependence of the energy loss due to electron heat conduction on electron temperature (Rosner *et al.*, 1978; Judge, 2002), but nevertheless, the degree of order in such images is remarkable given that the coronal heating mechanism is believed to occur on unobservably small scales (section 2.7.3). Therefore, we have a puzzle: What is the origin of the large scale order in the corona given that we know that dissipation occurs on very small scales?

For velocity field $\mathbf{u}$, in MHD the evolution of the solar magnetic field is described by an "induction equation", the simplest form of which is

$$\frac{\partial \mathbf{B}}{\partial t} = \mathbf{curl}\,(\mathbf{u} \times \mathbf{B}) + \eta \nabla^2 \mathbf{B}, \quad \eta = 1/\mu_0 \sigma, \tag{2.1}$$

where σ is a scalar electrical conductivity. This equation is readily derived from Faraday's Law and Ohm's law – itself a combination of electron and ion equations of motion undergoing collisions evaluated in the frame moving with the plasma's center of mass. The velocity field $\mathbf{u}$ of the Sun is observed to contain large (rotation, circulation, solar wind) and small (convection, turbulence, waves) scales.

First consider the second term in equation (2.1). Kinetic theory (Ohm's law) for collision dominated plasmas gives us values for the conductivity σ (e.g. Braginskii, 1965), from which we find that $\eta \sim 10^4$ cm^2 s^{-1} in the Sun's interior. Choosing $\ell \sim R_\odot/3 = 2 \times 10^{10}$ cm we find that the diffusive term in equation (2.1) has a time scale of $\ell^2/\eta \sim 10^9$ years. Yet the Sun exhibits magnetic variations many orders of magnitude faster. Next, note that the equation appears linear in $\mathbf{B}$, but this is misleading unless $\mathbf{u}$ is truly independent of $\mathbf{B}$. In real plasmas this is rarely the case since the plasma experiences a Lorentz force (effect of $\mathbf{j} \times \mathbf{B}$ on $\mathbf{u}$). MHD deals the coupling of the induction equation (or more complex versions of it) with equations of motion for the plasma in which $\mathbf{u}$ and $\mathbf{B}$ and other fluid variables (density, pressure, temperature, ...) evolve together.

The Sun is the archetypal object for study under the "high magnetic

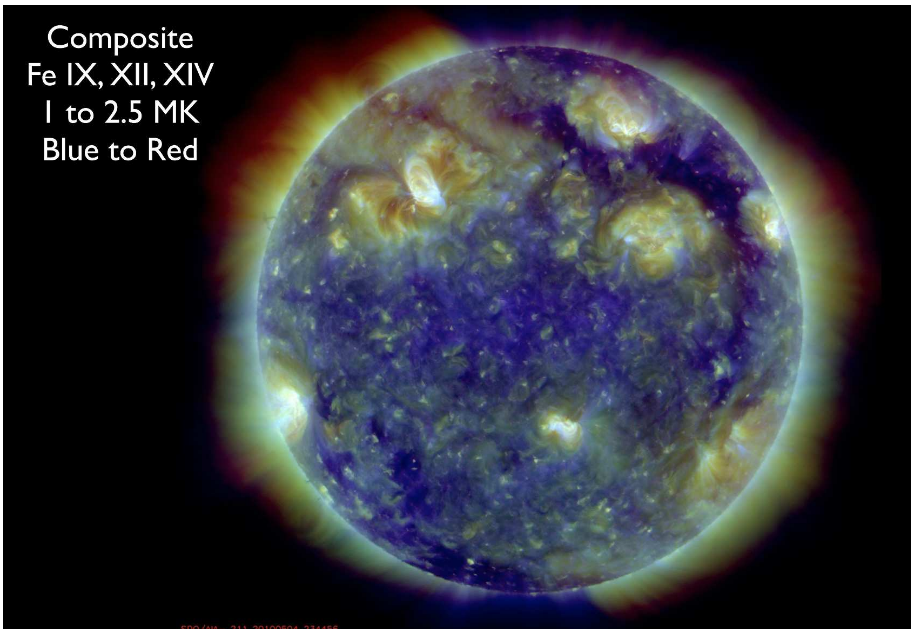


Fig. 2.2. A three color image taken by the AIA instrument on the SDO spacecraft on May 4 2010. The corona is not only well organized magnetically (Figure 2.15) but also thermally, since this image shows blue regions (electron temperatures near 1 million K, 1MK), green (2 MK) and red regions (3 MK). Entire regions of the Sun somehow know how to be at given temperatures that are very different from other regions on the Sun. The dark features are coronal holes- regions of fast plasma outflow.

Reynolds number" regime, $R_M = u\ell/\eta \gg 1$, where u and ℓ are characteristic speeds and lengths of the motion of the fluid. In this regime the first term of equation (2.1) dominates on large scales. In the limit of zero diffusion η (the "ideal MHD" limit), tubes of magnetic flux are tied forever to the the plasma that they entrain– "Alfvén's theorem". In this case the behavior of the plasma and fields must obey not only equations of motion but also an infinite number of topological constraints. It is a curious fact, almost a poetic tragedy, that ideal MHD leads to its own demise (2.7.3). In "natural" systems like the Sun, as opposed to idealized models with high degrees of order, Parker (1972); Parker (1994) has argued that topological constraints plus the equations of motion over-determine the solutions. The system evolves by trying to form mathematical "tangential discontinuities". But even in the presence of very small but finite plasma resistivities, in trying to become singular, the Maxwell stresses make steeper gradients until ideal MHD causes its own demise- small scales ℓ develop that the second term in equation (2.1) eventually leads to non-ideal behavior. Shakespeare might have written:

"Fair is foul, and foul is fair." – Witches, Act I, scene I, MacBeth

Here then are some central issues discussed below: In terms of the generation of magnetic fields, *why is the Sun obliged to vary cyclically, flipping the large scale fields every 11 years, and why must it appear most clearly as a sunspot?* In terms of the dissipation of magnetic field, *why does every solar-like star possess a corona, or chromosphere which is a partially ionized region between the visible surface and corona? Why must flaring occur?* I will use some simple physical arguments to identify some cutting edge problems. I am guided by some recent arguments by Parker (2009) and Spruit (2011). Pedagogical articles by Casini and Landi Degl'Innocenti (2008) and Rempel (2009) are also recommended.

2.2 Solar magnetism: so what?

Sunspots are small enough, and the Sun bright enough, that the occasional reports of them by ancient civilizations are of limited use. The first quantitative measurements of sunspots really required the invention of the telescope and the projection of solar images. Spots began to be counted in around 1610. It was not until 1908 that Hale proved that the spots were concentrations of strong magnetic fields, but his discovery meant that spots could then be used to probe solar magnetism back to the early 1600s, the time of Galileo.

Sunspots generally appear benign to the naked eye- so why should humanity care about these most obvious manifestations of a varying magnetic field on the Sun? The first rapid changes (minutes) seen in sunspots were reported by Carrington (1859) in a complex group of sunspots, confirmed independently by Hodgson. In the book by Young (1892), we find that this phenomenon

"... was immediately followed by a magnetic storm of unusual intensity, the auroral displays being most magnificent on both sides of the Atlantic, and even in Australia".

Now, aurorae are accompanied by magnetic disturbances on the ground, affecting compasses. These disturbances occur because the changing magnetism on the Sun, which leads to flares including the one observed by Carrington, produces high-energy radiative and particle disturbances at the earth. In turn, these disturbances induce electrical currents in the earth's magnetosphere and ionosphere. In 1859, telegraph wires hundreds of miles long were affected by currents along them induced through the changing magnetic fields in the ionosphere, such that signals could be transmitted with no applied EMF!

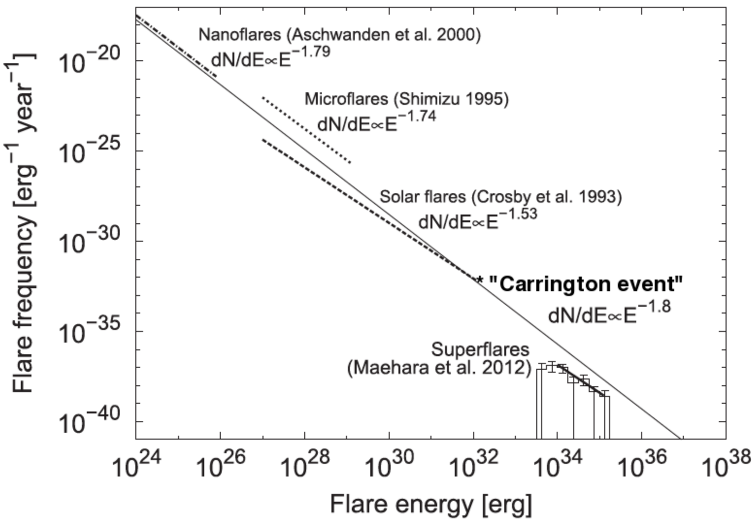


Fig. 2.3. Distributions of flares for the Sun in comparison with stars, modified from Shibata *et al.* (2013) to include the "Carrington Event". The solid histogram shows the frequency distribution of superflares on G V -type stars with rotational period > 10 d and effective temperatures of 5600–6000 K.

Towards the end of the 20$^{\text{th}}$ century it became clear that flares are also associated with the large scale ejection of material in the magnetized corona, "coronal mass ejections" (CMEs). In many flaring events, solar energetic particles (vastly supra-thermal) are emitted by the evolving fields in the corona, and via shocks in interplanetary space induced by CMEs. All of these phenomena, driven by evolving solar magnetic fields, put at risk modern society, as we become more and more dependent on spacecraft flying in the ionosphere, magnetosphere and in interplanetary space, and as our need for electrical power ever increases. Power grids and satellites, are just the most obvious "Achilles Heels" of our technological society. The huge geomagnetic effects of the "Carrington event" of 1859, (with energy estimated at $\approx 10^{32}$ erg by Shibata *et al.*, 2013, probably a lower limit), suggest that it is the largest flare recorded in history. The question arises as to how our infrastructure might be affected by such a strong flare, given that plasma ejection associated with smaller flares (X class between 11 and 13 March 1989) disrupted power transmission in Quebec. Given the spectacular aurorae that accompany such magnetic storms, a 20^{th} century descendent of MacBeth might conceivably have said

"So foul and fair a day I have not seen." – MacBeth's opening line

Recent work combining stellar and solar data (Shibata *et al.*, 2013, , see Figure 2.3) allows us to estimate the rate of occurrence of flares of Carrington amplitude and higher. Although solar data are based upon EUV and X-ray data and stellar data are based upon visible wavelength photometry from the *Kepler* satellite, Shibata et al. find that a flare of 100× the Carrington energy is expected roughly once in 800 years.

We must therefore attempt to understand why the Sun, in particular its magnetism, behaves as it does.

2.3 Measuring solar magnetic fields

2.3.1 Remote sensing

This section draws on the pedagogical article by Casini and Landi Degl'Innocenti (2008), another very nice article is that of Lites (2000) with several examples of solar measurements. Determining physical properties in solar plasmas is an exercise in "remote sensing". In astronomy, the origin of remote sensing is traced to Kirchhoff and Bunsen (1860) who identified spectral lines in the laboratory which coincided with Fraunhofer's dark lines in the solar spectrum recorded some 4 decades earlier. "Plasma spectroscopy" in fact began the modern era of astrophysics, it remains a primary tool – without it we would know little about the universe. Spectra of atoms, molecules and ions embedded in plasmas emit and absorb photons in a manner that encodes conditions in the plasma and any electric and magnetic fields threading the plasma. Simple examples are bulk flows of plasma reflected by Doppler shifts of the spectrum (expansion of the universe), temperatures of stellar atmospheres are reflected in the relative strengths of lines belonging to molecules, neutral atoms or ions. Curiously, after Fraunhofer's work but before that of Kirchhoff & Bunsen, philosopher Auguste Comte (1835) wrote

"On the subject of stars... we shall never be able by any means to study their chemical composition... or even their density... I regard any notion concerning the true mean temperature of the various stars as forever denied to us."

I cannot resist the following quote

"...But swords I smile at, weapons laugh to scorn, Brandish'd by man that's of a woman born." – MacBeth

Compte's "weapons", a tragic, perhaps comedic prediction of the future, serve as a warning to those of "us of a woman born" who try to predict the future of science. (MacBeth, however, is killed quickly after these words by MacDuff, not "born" of woman but delivered by cesarian section).

The effects of magnetic fields in spectra of astronomical plasmas are generally more subtle, because bright objects are usually hot and so spectral lines are broadened by thermal motions. Only when magnetic fields are strong can one use spectroscopy alone. More generally, one must also use "plasma polarization spectroscopy" because the magnetic fields are weaker. Magnetic fields break symmetries in current-carrying systems such as atoms, so it is not surprising that magnetically-induced atomic polarization leads to polarized spectra.

Spectral line polarization originates in plasmas in essentially two ways (from Casini and Landi Degl'Innocenti, 2008): Firstly, magnetic substates may be unevenly populated, so the transition components† ($JM \to J'M'$) no longer combine with the particular weights (natural populations) needed for the total polarization to vanish, such as in local thermodynamic equilibrium (LTE). Unequal populations occur when *atomic excitation/de-excitation processes are anisotropic*. Secondly, even with atomic substates populated naturally, the substates may be *separated in energy*, so that a spectral analysis of atomic transitions reveals varying polarization properties with wavelength. This second case is most familiar to solar physicists, leading to the Zeeman and Stark effects.

The two sources of polarized spectral lines are not mutually exclusive. Hanle (1924) experimented with anisotropically excited atoms in the pres-

† Here, J is the total angular momentum quantum number, M the projection of this quantum number on to a particular axis, such as along a magnetic vector $\mathbf{B}$.

ence of magnetic fields, finding that magnetic fields can modify the zero-field polarization with field strengths far below those necessary to produce visible energy separations via the Zeeman effect. Such modifications are referred to as the Hanle effect.

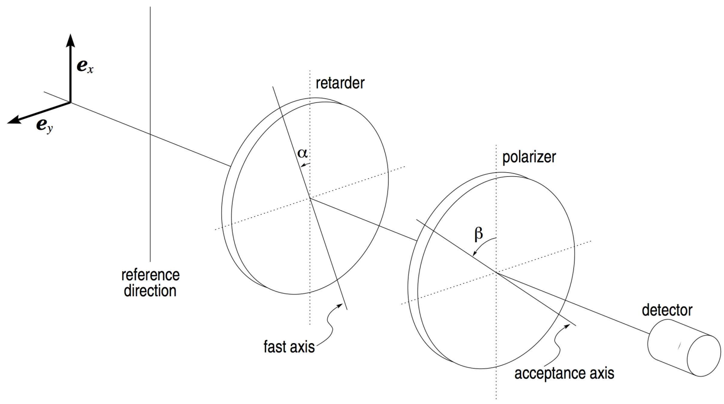


Fig. 2.4. Typical configuration of a polarimeter. Different polarization states of the radiation passing through the instrument are measured by suitable combinations of the retarder and polarizer position angles, α and β. From Casini and Landi Degl'Innocenti (2008).

2.3.2 Measuring polarization of light

Polarization of light is mathematically specified by the two complex Cartesian components, E_x and E_y of the radiation electric field with wave-vector along the z axis. In a physical description a coherency matrix (or polarization tensor) of the radiation field (averaged over the acquisition time and elemental surface of the light detector) is specified,

$$C = \begin{pmatrix} \langle E_x^* E_x \rangle & \langle E_x^* E_y \rangle \\ \langle E_y^* E_x \rangle & \langle E_y^* E_y \rangle \end{pmatrix} \equiv \frac{1}{2} \begin{pmatrix} I+Q & U-iV \\ U+iV & I-Q \end{pmatrix} \tag{2.2}$$

so that four independent parameters are needed in order to describe polarized radiation, either in terms of the electric fields or equivalently the Stokes parameters, I, Q, U, V (Casini and Landi Degl'Innocenti, 2008). Observationally, it is customary use the four (real) Stokes parameters: I is the intensity, Q and U are the two independent parameters needed to describe linear polarization on the $x-y$ plane, and V is the circular-polarization parameter. Casini and Landi Degl'Innocenti (2008) adopt an operational definition for measuring polarization of light in terms of an ideal instrument, consisting of a linear polarizer and retarder (Figure 2.4). A retarder introduces a phase difference between orthogonal directions of polarization, examples being calcite crystals with different refractive indices for light with electric vectors along and perpendicular to certain directions (birefringence). For a $\lambda/4$ retarder, this "polarimeter" registers counts on a detector as functions of angles α and β between the retarder's fast axis, the polarizer's acceptance axis, and a reference direction as:

$$S(\alpha,\beta) = k\,[I + (Q\cos 2\alpha + U\sin 2\alpha)\cos 2(\beta-\alpha) + V\sin 2(\beta-\alpha)] \quad (2.3)$$

where k is a factor that includes detector gain, flat field etc. corrections. A sequence of measurements of $S(\alpha,\beta)$ with α and β varying with time can be used to determine I, Q, U, V with suitable choices of α and β. This involves a 4×4 matrix inversion coupling all four Stokes parameters. The needed sequence of observations is called "polarization modulation".

Several factors are critical in making accurate polarimetric measurements.

- Frequently, Q, U, V are significantly smaller than I, sometimes by orders of magnitude.
- Since complete measurements of all four Stokes parameters requires at least four measurements, sequential in time in most polarimeters, *any (spurious) time variations can lead to measurement errors*. Such errors include *crosstalk*, in which, say, variations in Q *entering the polarimeter from the telescope feed*, due to non-solar sources induce spurious signals in, say U.
- A large source of error and crosstalk includes *atmospheric seeing* (e.g. Lites, 1987). Space polarimeters, such as the SP instrument (Lites *et al.*, 2001) on the Hinode spacecraft, avoid seeing problems (but do have some spacecraft jitter).
- In observations affected by seeing, $I \rightarrow Q, U, V$ crosstalk often dominates. The solution is then to split the beam, with suitable choices of $\beta - \alpha$, to measure effectively $I + aQ + bU + cV$ and $I - aQ - bU - cV$ simultaneously. This yields an 8×4 matrix inversion (Seagraves and Elmore, 1994) or it's algebraic equivalent. The Hinode SP instrument in space has a dual beam capability, but results obtained with one beam are of comparable quality, as one might expect (Lites private communication, ca. 2009).
- The polarimeter measures only the state of polarized light entering the instrument. If the telescope is "polarizing", in the sense that it changes the incoming Stokes array I, Q, U, V via off-axis asymmetries for example, via a matrix $\mathbf{T}$, then this matrix must be determined (a "polarization calibration" is required) and then inverted.

Sometimes we need to measure very small polarizations, say $10^{-4}I$, to determine magnetic fields in the Sun. Such measurements are extremely difficult to perform, but there are ways around them, such as by modulation/demodulation at frequencies far above any frequency induced by seeing or other factors. The ZIMPOL instrument does this with a combination of piezo-electric modulation by crystals and by charge caching detectors (e.g. Gandorfer and Povel, 1997).

2.3.3 The Sun is not always bright enough for polarimetry

It should be recognized that spectropolarimetry at high angular resolution actually is a *photon starved* exercise. This surprising conclusion has been re-emphasized by Landi Degl'Innocenti (2013). At the diffraction limit of any telescope, with $\theta_{DL} = 1.22\lambda/D$ the flux density f_{DL} of photons is independent of telescope aperture D because

$$f_{DL} = I\pi\theta_{DL}^2 = I\pi(1.22\lambda/D)^2 \tag{2.4}$$

per unit area, where I is the solar disk intensity and we have assumed the flux is contained in one Airy disk. The total photon flux per detector pixel is $\pi D^2/2$ times this and independent of D. Using reasonable values for system efficiency, spectral resolution, wavelength, Landi Degl'Innocenti finds that one can accumulate only 10^6 photons per second per pixel, meaning that 100s integrations are needed to achieve a statistical sensitivity of 10^{-4}. While not a problem for stars, the diffraction limit of large solar telescopes corresponds to tens of km at the solar surface, in 100s the Sun's atmosphere can change dramatically on such scales. Trade-offs must be made by a judicious selection of angular resolution, wavelength, instrumental throughput, even with 4m class telescopes.

Lastly, even with great care, it should be pointed out that systematic sources of error (fringes, crosstalk, calibration errors) generally dominate polarization measurements at high sensitivities. Spectropolarimetry presents interesting challenges.

"... that is a step on which I must fall down, or else o'erleap, for in my way it lies." – MacBeth

2.3.4 Zeeman effect

The Zeeman splitting of atomic sub-levels is $h\nu_L \sim \mu_B B$, where ν_L is the Larmor frequency, μ_B is the Bohr magneton. The radiation emitted or absorbed by the associated "σ" ($\Delta M = \pm 1$ and "π" ($\Delta M = 0$) radiative components then depends on the magnitude and direction of the magnetic field vector. Requiring polarizations in excess of a few percent in most instruments, then, as we will see below, this means $\nu_L/\Delta\nu \gtrsim 2 \times 10^{-2}$, where $\Delta\nu$ is the characteristic line width. In the Sun's photosphere, Doppler broadening is usually dominant, with a value of a few km s^{-1}, $\sim 10^{-5}c$. Translated into frequency units, we find that the Zeeman effect is useful for magnetic fields of order

$$\nu_L \sim \frac{\mu_B B}{h} \gtrsim 2 \times 10^{-2} 10^{-5} \frac{c}{\lambda_0} \tag{2.5}$$

$$B \gtrsim 90 \, \frac{5000\text{Å}}{\lambda} \text{ G.} \tag{2.6}$$

Figure 2.5 shows the "classical" analog for the Zeeman effect which applies to $J = 0 \rightarrow J = 1$ transitions, from Lites (2000).

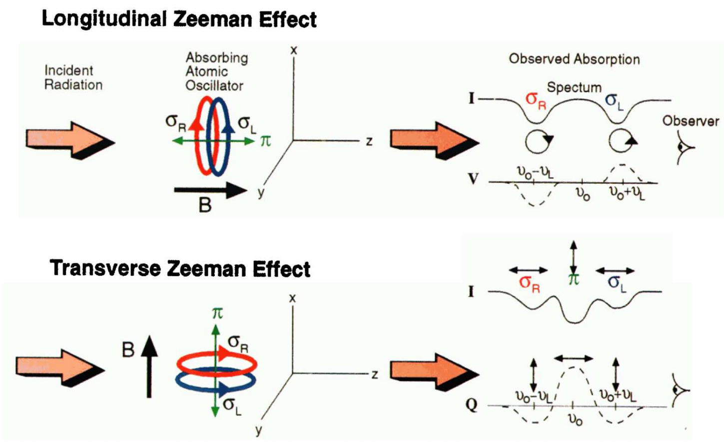


Fig. 2.5. The classical analog of the Zeeman effect an an atomic system represented by classical oscillators. An atom absorbs a little of the incoming continuum light (from the left) with the magnetic field aligned along the z and x axes respectively, showing the longitudinal (top) and transverse (bottom) Zeeman effect. From Lites (2000).

In the upper panel of the figure (longitudinal Zeeman effect), the circularly orbiting oscillators are seen "face on" so that they emit circularly polarized light in the observer's direction. Conversely, in the lower panel, the oscillators are seen "edge on" so that they emit linearly polarized light. Due to symmetry, the "red" and "blue" shifted components are of the same amplitude, and they are shifted in frequency by the Larmor frequency of the atomic level ν_L in both cases.

In the longitudinal-field case, a difference in measurement of left versus right circularly polarized light (this is the definition of Stokes V) will reveal the V polarized profile as shown. The I profile shown is "fully split", i.e. Zeeman splitting exceeds line broadening from the plasma (thermal, bulk motions, collisions), and is symmetric. In the limit that the Zeeman splitting is small, the Stokes I profile is merely broadened a little - the two absorption features $\sigma_{\rm R}$, $\sigma_{\rm L}$ in I in Figure 2.5 being blended into one feature. However, the Stokes V profile then survives with an amplitude which is proportional to the longitudinal field strength, as shown below.

Figure 2.6 shows profiles measured by the Advanced Stokes Polarimeter. The profiles shown are most readily understood using the excellent pedagogical article of Jeffries *et al.* (1989) which focuses on the transport of polarized radiation through an atmosphere. These authors first derive emission and absorption coefficients for light passing through electrons bound to atoms and ions using equations of motion and dielectric theory, in the presence of arbitrarily oriented magnetic fields. The absorption and emission processes are assumed uncorrelated, such as occurs in thermal equilibrium and LTE, and the processes of damping and Larmor precession are assumed far smaller than the "resonance frequency" (i.e. frequency of the line radiation, $\sim 10^{15}$ Hz). They then derive the equations of radiation transport and solutions for simple cases. Jeffries et al. discuss in particular the "weak field limit" (Larmor frequency ν_L much smaller than the combined Doppler and natural widths $\Delta\nu$ of the lines). Their equations (45) and (47), consistency relations related to their transfer equations, readily explain the Q, U, V profiles seen in figure 2.6: the emergent Stokes V profile is a term that is first order in $\nu_L/\Delta\nu$, proportional to the first derivative of the intensity profile and so asymmetric around line center. The QU profiles are second order terms and are symmetric about line center.

This different behavior originates from the dependence of the emitted dipolar radiation on the quantum numbers $\alpha JM \rightarrow \alpha' J' M'$ of the atomic transitions (refer to Figure 2.5). When $\Delta M = \pm 1$, labeled "σ" transitions, the atom emits or absorbs left- and right- circularly polarized light, the radiative transitions being shifted in frequency by $\pm\nu_L$. When $\nu_L/\Delta\nu \ll 1$, we can use a first order expansion of the (left - right) polarization states in frequency or wavelength, yielding $V \propto \nu_L \phi'$ where $\phi' = d\phi/d\nu$. For the $\Delta M = 0$ "π" transitions, the atom emits or absorbs linearly polarized light unshifted in frequency. Unlike Stokes V, the linear polarization (Stokes Q, U) measurements "see" both the σ and π components. When the substate populations are equal (LTE), since they occur in the combination $\pi - (\sigma_{+1} + \sigma_{-1})/2$, the leading order term is second order in $\nu_L/\Delta\nu$, therefore yielding a frequency dependence $\propto \phi''$.

In the work of Jeffries *et al.* (1989), the Zeeman effect is imprinted in the outward directed solar $IQUV$ parameters through a matrix transport equation for the array (I, Q, U, V) in which a 4×4 *matrix* of absorption and emission coefficients contains the appropriate Zeeman wavelength shifts and polarization states. Taking account of the geometry, Jeffries and colleagues present solutions to the simplest radiative transfer equation applicable to photospheric lines, the assumption that the source function varies with (continuum) optical depth τ as $S(\tau) = B_0 + B_1\tau$. Ignoring, for tutorial purposes, complications due to "magneto-optical effects†", we can write their equation (39) for the emergent Stokes parameters at any particular frequency as

$$I \approx B_0 + \frac{\mu B_1}{\Delta}(1 + \eta_I)^3 \tag{2.7}$$

$$Q, U, V \approx -\frac{\mu B_1}{\Delta}(1 + \eta_I)^2 \eta_{Q,U,V} \tag{2.8}$$

where the $\eta_i = \kappa_i/\kappa_C$ are ratios of the line's Stokes component i absorption coefficient to that of the continuum κ_C (their equations 28 ignoring the primed MO terms):

$$\eta_I = \kappa_C \frac{1}{2}\left[\left(\frac{\kappa_r + \kappa_l}{2}\right)(1 + \cos^2\gamma) + \kappa_p \sin^2\gamma\right] \tag{2.9}$$

$$\eta_Q = \kappa_C \frac{1}{2}\left(\kappa_p - \frac{\kappa_r + \kappa_l}{2}\right)\sin^2\gamma \cos 2\chi \tag{2.10}$$

$$\eta_U = \kappa_C \frac{1}{2}\left(\kappa_p - \frac{\kappa_r + \kappa_l}{2}\right)\sin^2\gamma \sin 2\chi \tag{2.11}$$

$$\eta_V = \kappa_C \left(\frac{\kappa_r - \kappa_l}{2}\right)\cos\gamma \tag{2.12}$$

Here, γ and χ are angles that magnetic fields make with the line of sight and on the plane of the sky defined according to the reference direction (Figure 2.4). Note that the azimuthal angle χ occurs only through a sine and cosine function of 2χ, thus the *azimuth* is determined by observations of Q, U only to within 180°. This is the "azimuthal ambiguity". These

† MO effects arise from phase changes introduced by non-unit real parts of the refractive index in the dielectric theory. In the notation of Jeffries et al. the MO effects are contained in the quantities having have a "prime" superscript in their standard transfer equation (35) and subsequently in their variable ϱ. MO effects are negligible in the weak field limit.

equations show algebraically the combinations of energetically shifted left- (κ_l), right- (κ_r) σ components, and the unshifted π component κ_p which lead immediately to the profiles seen in Figure 2.5. The Q, U terms are of the form $2 \times \pi$ component minus the sum of the σ components: this is a difference equation for the second derivative; the V term is simply the first difference between the σ components.

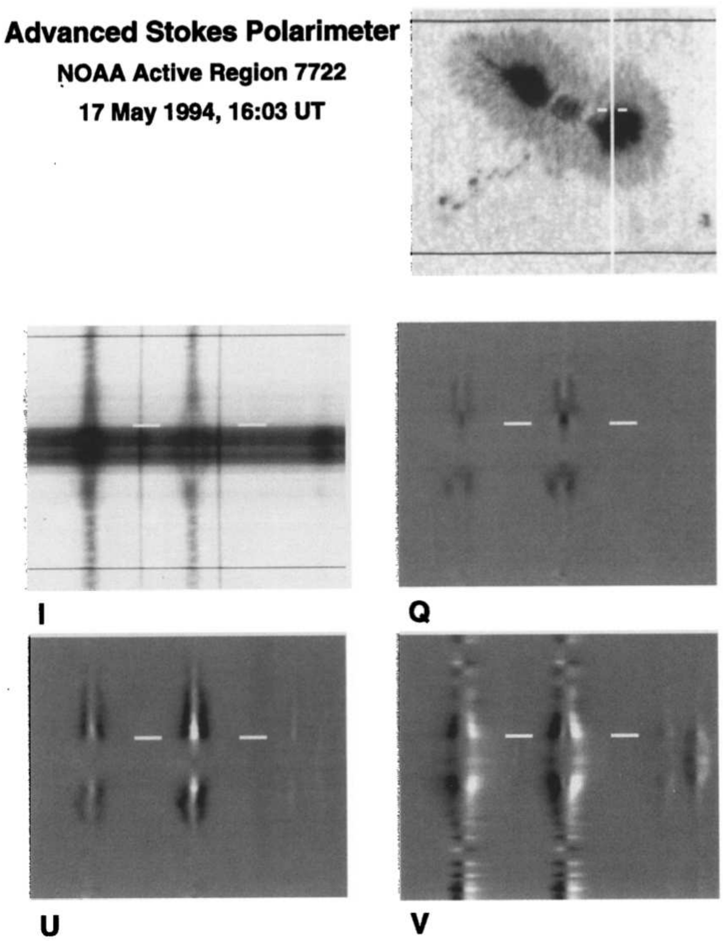


Fig. 2.6. Spectropolarimetric line profiles of a sunspot with polarization generated by the Zeeman effect. The upper panel shows the position of the slit on a continuum image, other panels show Stokes profiles as a function of wavelength (abscissa) and position along the slit. The two broad dark vertical lines are a pair of 630 nm lines of Fe I, they show Zeeman splitting in I and also V over the sunspot umbra and penumbra. Away from the sunspot the Stokes QUV profiles are similar, being formed in the "weak field limit" (see text). From Lites (2000).

In the "weak field limit", magnetic information is contained only in the *amplitudes* of the Stokes profiles, and angular factors, the Q, U, V profile *shapes* being set by derivatives of I. It is not possible to discriminate between two configurations containing the same net magnetic flux per unit area unless additional measurements from, e.g., a line formed outside the weak field regime, are available. In the weak field limit the Zeeman effect merely measures the *magnetic flux density* (Maxwells per square cm) not the *magnetic field strength* (Gauss).

Lastly, it is worth emphasizing that because Zeeman-induced linear polarization is usually small, the bulk of the literature, indeed almost the entire literature that deals with *global* fields, is based upon circularly polarized data from the longitudinal Zeeman effect. Such measurements, called "longitudinal magnetograms", measure only the *net line of sight field component* of **B**. Regular observations began in the 1950s after the invention of the scanning magnetograph (Babcock and Babcock, 1952; Babcock, 1953). With the advent of the SOLIS ground-based instrument and HMI instrument SDO, "vector magnetograms" (using linear and circular polarization measurements) are becoming more commonplace.

2.3.5 Hanle effect

Above, we found that the Zeeman effect requires field strengths in excess of a few tens of G, the polarized light depending on the ratio of Larmor frequency to Doppler width frequency. In the Hanle effect the relevant parameter is the product

$$2\pi\nu_L A^{-1} \sim 1 \tag{2.13}$$

with A the spontaneous decay rate of an atomic level. An atom in a given state with a net magnetic moment μ gyrates around magnetic field lines with frequency $\nu_L = \mu B/h$. If $2\pi\nu_L A^{-1} \sim 1$, line photons are emitted while the atom's azimuthal angle around the magnetic field changes by one radian. In the vector model of the atom an initially polarized state will lose some "memory" of its initial direction, the emission will be *rotated* and the magnitude of polarization *reduced*, since radiative decay is a stochastic process obeying a distribution of lifetimes $\propto \exp(-tA)$.

The Hanle effect is sensitive to weaker magnetic fields than the Zeeman effect. For electric dipole transitions in neutral species, $A \sim 10^8$ s^{-1} for a spectral line at $\lambda_0 = 5000$ Å. Then, with $\nu_0 = 6 \times 10^{14}$ Hz, we have $A/\nu_0 \sim 1.6 \times 10^{-7}$. If we allow for rotation of ~ 0.1 radians to be detectable, then the Hanle effect is sensitive to magnetic fields when

$$\nu_L \sim \frac{0.1 \rightarrow 1}{2\pi A}, \text{ or} \tag{2.14}$$

$$B \sim 1 \rightarrow 10 \text{ G}, \quad \text{(permitted lines of neutrals).} \tag{2.15}$$

The Hanle effect merely *alters existing polarization*, its use in solar magnetic field measurements is related to spectral lines which are polarized by other processes, most commonly by excitation by radiation fields that are anisotropic. This is explicitly then a non-LTE phenomenon since LTE implies detailed balance, isotropic excitation in particular. Unlike the Zeeman effect, Hanle rotation and depolarization is additive, no matter the sign of magnetic field. Therefore it is suited to detection of randomly oriented fields. In this context Kleint *et al.* (2010) used the Hanle effect to investigate small-scale, disordered magnetic fields in the quiet Sun using lines sensitive to magnetic fields formed near the solar limb. Their results, shown in Figure 2.7, show no significant variations around 5 G between 2000 and 2010. In contrast, Zeeman measurements of solar magnetic fields associated with sunspots amount to average field strengths between 4 Mx cm^{-2} and 20 Mx cm^{-2} (Schrijver and Harvey, 1989) over the sunspot cycle.

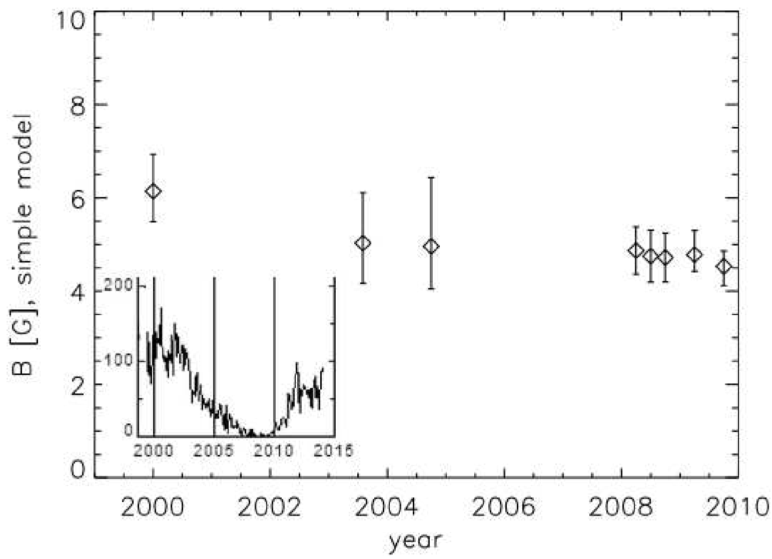


Fig. 2.7. Field strengths measured from the Hanle effect over a decade, near the limb of the quiet Sun from Kleint *et al.* (2010). In contrast, the inset shows sunspot numbers from 1999 to the present, varying by a factor of 10-100 over the same period!

The reader should refer to Casini and Landi Degl'Innocenti (2008) for an accessible pedagogical text on the (far more complex) quantal treatment of the Hanle effect.

2.3.6 Natural systems as recorders of solar magnetism

The large scale (several AU) solar magnetic fields influence the earth by modulating the incoming flux of cosmic rays, see McCracken *et al.* (2013) for a recent review. Radionuclides ^{10}Be and ^{14}C are produced in the atmosphere and sequestered in polar ice and tree rings. Neutrons are also created by cosmic ray interactions with Earth's atmosphere. Neutron fluxes show a clear 22 year modulation of a fundamental 11 year cycle as measured over the past 60+ years (e.g. Beer *et al.*, 2012). Taken together, this has allowed scientists to infer a "solar modulation" function for the past 10,000 years. The function, an indirect measure of heliospheric magnetic fields, is based upon theoretical models of cosmic ray transport (Parker, 1965).

As shown by McCracken *et al.* (2013), the independent ^{10}Be and ^{14}C records are mutually consistent over the past 10,000 years (the "holocene"). Figure 2.8 shows Fourier amplitudes of the ^{10}Be time series reported by McCracken and colleagues. The large-scale (heliospheric) solar magnetic field appears to show a quasi-periodic behavior on time scales between 50 and 10,000 years.

That earth is a detector of solar influences should be no surprise, the radionuclide traces of galactic cosmic ray modulation are one interesting example. The earth responds to solar radiation and particles: aurorae caused by electromagnetic perturbations from solar charged particles (Birkeland and Muir, 1908) influence the magnetosphere, and are correlated with sunspot activity. Indeed Eddy (1976) cited the lack of aurorae in support of the reality of the Maunder Minimum in sunspots. The mere existence of an ionosphere and upper atmospheric ozone prompted Saha (1937) to conclude that the Sun must emit excess UV radiation before the conclusive evidence of a hot corona (Grotrian, 1939; Edlén, 1943).

2.4 The observational record

B.C. Low, a theoretician who studies physical and mathematical properties of the MHD equations, often reminds us that "solar physics is an observationally-driven science". I believe he means that, even in the case of the simplest theoretical model for coupled magnetic fields and plasmas – the MHD picture – the non-linear coupling between the governing equations admit such a broad variety of solutions, that we must be guided by

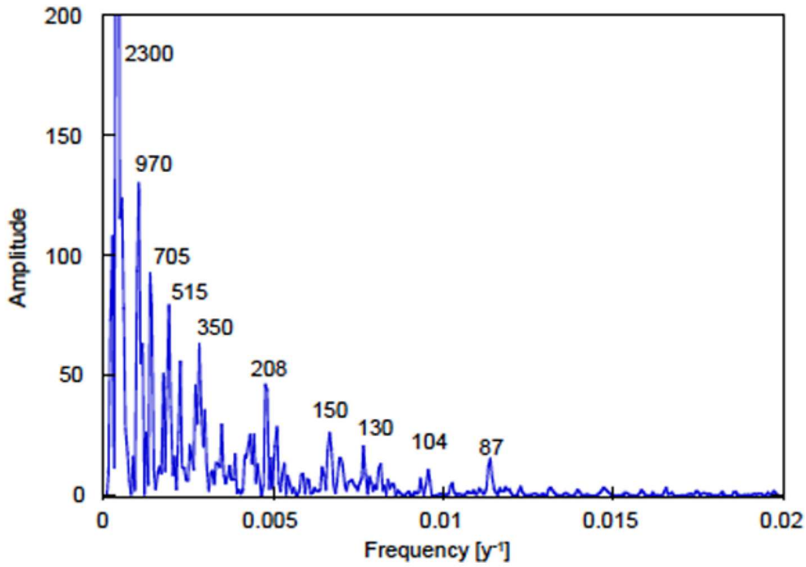


Fig. 2.8. Fourier amplitude spectrum of the ^{10}Be data ice core data versus frequency, for periods between 50 and a few thousand years, taken from McCracken *et al.* (2013). Periods corresponding to the major peaks are annotated in years.

observations. Indeed, the justification for our new facilities, such as the Advanced Technology Solar Telescope, the first major facility for ground--based solar physics for the United States in almost 50 years, is to be able to study the non-linear interactions between plasma and (electro-) magnetic fields threading them, on scales inaccessible to laboratories. So here I review landmark observations in a historical context.

2.4.1 Early hints of variable solar magnetism

Fig 36. of the book by Young (1892) shows spectra from 1870 of Fraunhofer's D lines. The figure is clearly recognizable today as a typical case of spectral lines split by magnetic fields- the Zeeman effect - but because the nature of sunspots was not known until much later these data were interpreted in terms of earlier work showing reversals in chromospheric lines- i.e. the origin of the reversal was interpreted as thermodynamic, not magnetic. During the 1878 total eclipse, during a deep minimum in sunspot number, it was noted that the corona appeared to trace magnetic lines of force.

From 1610 sunspots have been recorded essentially daily. While the data are from heterogeneous sources, several properties of sunspot numbers are

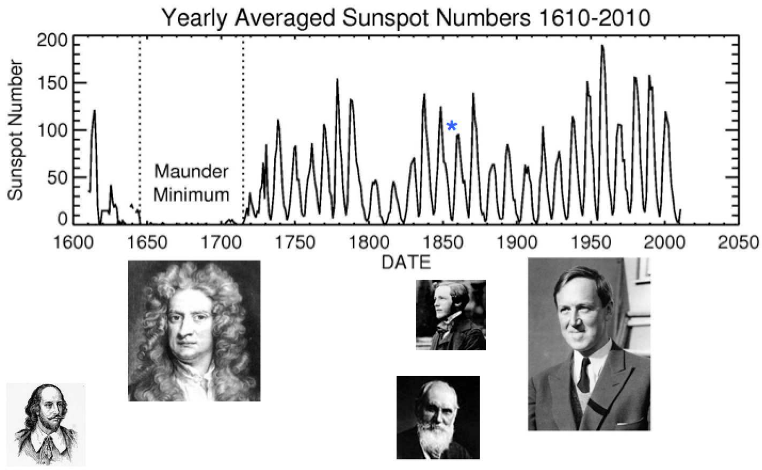


Fig. 2.9. The historical record of sunspot numbers is shown along with some characters discussed in the text. The asterisk at 1859 marks the "Carrington Event". From left, there is Shakespeare, Newton, Carrington (lower) Maxwell, Alfvén. Widths of the images corresponds to life spans.

robust. Figure 2.9 shows sunspot numbers as compiled by David Hathaway†. After 17 years of observations, Schwabe (1844) discovered the cyclic increase and decrease of sunspot numbers. In 1852, Sabine reported that the sunspot cycle period was "absolutely identical" to that of geomagnetic activity, Young (1892) shows data from 1772 to 1880.

2.4.2 Large-scale properties of solar magnetic fields

Here I offer selected observations of relevance of the origin and dissipation of solar magnetic fields, some of these are already evident in Figure 2.10, showing surface magnetic fields as measured using longitudinal magnetographs. I follow lines of argument from Parker (2009); Spruit (2011):

- spots - compact regions of high field strength - are the most obvious signature (Hale, 1908)
- spots emerge only below latitudes of 30°, their distribution of latitude with time follows the "butterfly wing" pattern as spots emerge closer to the equator with time Maunder (1904, reporting on work with wife Annie)
- leading polarities of spots emerge closer to the equator than following polarity, this "tilt" is larger the further spots emerge from the equator ("Joy's law", Hale *et al.*, 1919)

† This is a version of `http://solarscience.msfc.nasa.gov/images/ssn_yearly.jpg`.

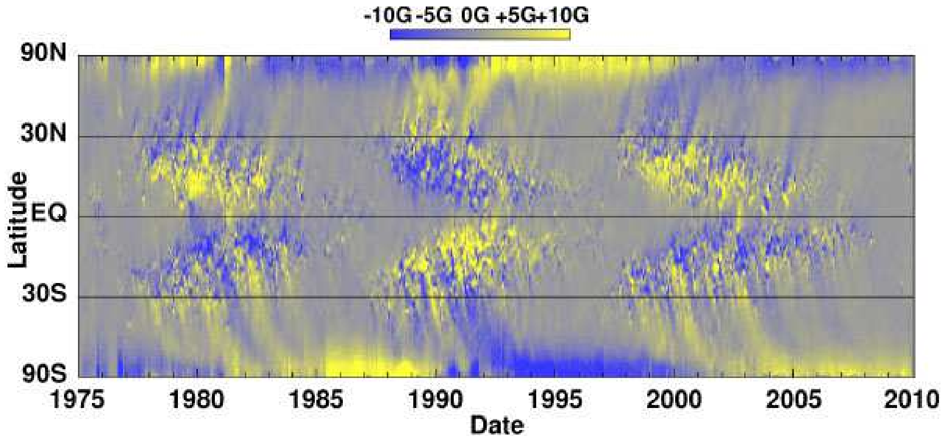


Fig. 2.10. A magnetic butterfly diagram constructed from the radial magnetic field obtained from instruments on Kitt Peak and SOHO. This illustrates Hale's polarity laws, Joy's tilt law, polar field reversals in relation to sunspots, and the apparent transport of magnetic field toward the poles. From Hathaway (2010).

- a 22 year magnetic cycle dominates (Hale *et al.*, 1919) but includes some stochasticity
- the large-scale (dipole, quadrupole) components vary on a 22 year cycle, somewhat out of phase with the spots (Babcock, 1961)
- brightness variations alone vary with a dominant 11 year period and a ~ 26 day rotational period
- spot emergence in N and S hemispheres are not necessarily in phase
- at least one period occurred where sunspots almost disappeared over 70 years (the Maunder Minimum, Eddy, 1976)
- large-scale (heliospheric) fields show stochastic and periodic magnetic variability over the last few thousand years (2.8) as recorded in ^{10}Be and ^{10}C radionuclide data and correlated with neutron monitor data (McCracken *et al.*, 2013)
- spots often re-emerge at specific longitudes
- the Sun's 11 year brightness variations resemble a "cycling" class of similar, slowly rotating G stars, (Baliunas *et al.*, 1995; Judge and Thompson, 2012), although roughly as many stars do not show cycling activity as do on decadal time scales (Figure 2.11)

McCracken *et al.* (2013, and references in the paper) draw further conclusions from sunspot and radionuclide studies over the past 600 yr:

- "The Hale 22 year cycle of solar magnetism, and the heliospheric counterpart, continued throughout the Spoerer and Maunder Minima, and

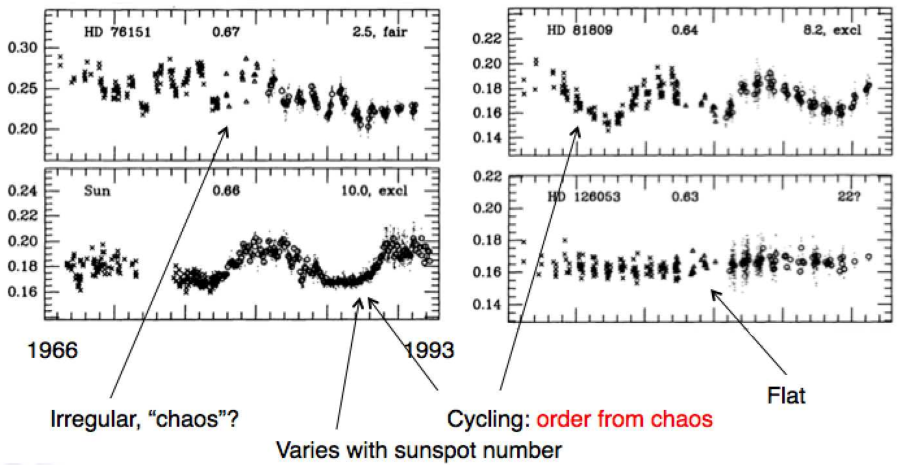


Fig. 2.11. A selection of sun-like stars from the study of Baliunas *et al.* (1995), showing the "Ca II H index" versus time. The solar index varies as shown in the lower left hand panel, in phase with sunspots, the index is therefore a "proxy" for sunspot-like activity on sun-like stars. Roughly as many stars show "cycling" behavior as do not.

- The amplitude of the 24/12-yr modulation of the paleo-cosmic ray record for the Spoerer Grand Minimum implies a ... [factor] 0.5–2.5 variation in the heliospheric magnetic field near Earth."

The continued cycling behavior of the Sun through the Maunder Minimum is hinted at already by the sunspot study of Ribes and Nesme-Ribes (1993). Further characteristics of the evolving solar magnetic field on the surface and in the interior have been reviewed by Hathaway (2010), from which Figure 2.10 is taken. While an obvious point, it must always be remembered that such figures reveal magnetic field patterns evolving both through and across a 2D surface of a 3D system. Care must be taken in interpreting such data only in terms of dynamics across a 2D surface.

2.4.3 Smaller scales, sunspots, flares and CMEs

Spruit (2011) has emphasized more local aspects of sunspots. Spots emerge first by exhibiting fragmented surface magnetic fields with mixed polarities. From this, clumps of opposite polarity then form and diverge without influence by convection. Spruit continues

"This striking behavior is the opposite of diffusion. To force it into a diffusion picture, one would have to reverse the arrow of time. Instead of opposite polarities decaying by diffusing into each other, they segregate out from a mix. The MHD equations are completely symmetric with respect to the sign of the magnetic field, however. There are no flows (no matter how complex) that can separate fields of different signs out of a mixture. This rules out a priori all models attempting to explain the formation of sunspots and active regions by turbulent diffusion. . . The observations. . . demonstrate that the orientation and location of the polarities seen in an active region must already be have been present in the initial conditions: in the layers below the surface from which the magnetic field traveled to the surface."

These and other observations of Spruit are discussed in Section 2.6.1

Spruit then argues that models of dynamo action based upon cyclonic turbulence (originating with the ideas of Parker, 1955) cannot lie at the heart of the solar dynamo. Simply put, how can small-scale turbulence cause the intense, clumped tubes of sunspot flux?

Solar flares mostly (but not always!) originate in the neighborhood of sunspots. Following the development of MHD in the 1940s and 1950s (section 2.5), it was realized, through a process of elimination, that the sudden release of 10^{32} ergs of energy in a few minutes implies the storage and release of magnetic free energy in tenuous regions of the Sun's atmosphere- the corona. The best introduction to this problem I have found is "The Physics of Solar Flares", proceedings of AAS-NASA Symposium, editors Wilmot N. Hess, ed., NASA SP-50 1964.

2.4.4 Summary of observed properties

Somehow, the solar plasma/magnetic field interactions:

- induce global solar variability to time scales of 11 yr and (much) less, compared with a 10^5 yr Kelvin-Helmholz time scale,
- introduce highly variable high energy tails to the distributions of photons and particles,
- produce *order out of chaos*

As noted by Parker (1955), the order must arise because the Sun rotates. The way this happens remains an active research area (Parker, 2009; Spruit, 2011). Again, appealing to stellar behavior proves fruitful- Figure 2.12, from Noyes *et al.* (1984), shows that the Sun lies squarely in the activity levels expected in an ensemble of solar-like stars. This plot has on the abscissa the ratio of rotation period to convective turnover time, the inverse of the Rossby number, to be discussed further below.

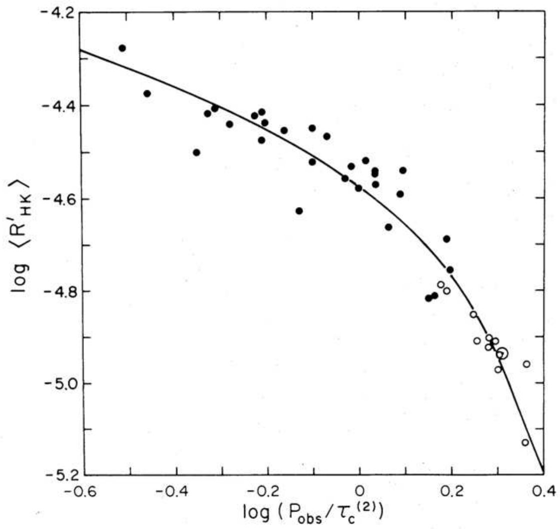


Fig. 2.12. A scatter plot of an activity index (in this case not the S index but the more physical R'_{HK}) versus rotation period/ convective turnover time, the inverse of the Rossby number, figure 8 from Noyes *et al.* (1984).

2.5 Magnetohydrodynamics

The essential ideas behind MHD as a description of plasmas and the electromagnetic fields are firstly, the usual continuum approach is valid for the plasma (many particles in each phase element), imposing a lower limit to the scales at which MHD applies. Even at the low densities of the corona, $\sim 10^8$ particles per cm^3, this lower limit is less one cm. Second, the plasma is quasi-neutral (electrons move quickly to eliminate any electric field on macroscopic scales). Third, we neglect Maxwell's displacement current, which implies that plasma evolution occurs slower than the light crossing time $t \sim \ell/c$ for the plasma ($R_\odot/c \sim 2$ s). Fourth, in the case of "single fluid" MHD, collisions relax different particles (atoms, ions, electrons) on timescales short compared with dynamics, we deal with densities ρ, pressures p which are sums over all constituents, which are all described by single fluid velocity $\mathbf{u}$ and temperature T†.

Another central aspect of MHD is not often appreciated (B.C. Low 2012, personal communication). Einstein (1905) boldly relaxed the Galilean transformations at the root of Newton's laws of motion in order to reconcile the

† (Parker, 2007, section 8.1) points out that continuum approximations often apply even in collisionless regimes since collisions conserve mass and momentum. Some of the fluid equations are still valid even in collisionless plasmas, when care is take of higher order quantities such as the pressure tensor!

laws with Maxwell's equations and to account for the experimental results of Michelson and Morley (1887). Partly reversing Einstein's work, Alfvén in the 1940s (Alfven, 1950) formulated MHD within the Galilean/Newtonian framework, but keeping Einstein's relativistic transformations to $\mathcal{O}(v/c)$. The Bard might have remarked

"What's done cannot be undone." – Lady MacBeth

but then Shakespeare preceded Newton and the notion of physical approximations. The electric field $\mathbf{E}$, magnetic field $\mathbf{B}$, charge density ρ and current density $\mathbf{j}$ transform from rest frame to one moving with velocity $\mathbf{u}$ to $\mathcal{O}(u/v)$ as, e.g. Ferraro and Plumpton (1966)

$$\begin{aligned} \mathbf{E}' &\approx \mathbf{E} + \mathbf{u} \times \mathbf{B}; \quad \mathbf{B}' \approx \mathbf{B}, \\ \rho' &\approx \rho; \quad \mathbf{j}' \approx \mathbf{j}. \end{aligned}$$

To $\mathcal{O}(u/c)$, $\mathbf{E}$ is frame-dependent, but we can speak of $\mathbf{B}$, $\mathbf{j}$ and ρ without specifying the frame of reference.

To arrive at the standard MHD equations we add Maxwell's equations, which have source terms ρ and $\mathbf{j}$. On scales larger than (usually small) Debye radii r_D, the charge density $n_p e - n_e e$ scale with $\exp -r/r_D$ (here n refers to particle number densities). However, we cannot neglect macroscopic electric currents $\mathbf{j}$, which arise from small (but finite) differences in flow speeds of the electron and ion fluids. Braginskii (1965) derives equations of motion for electrons and ions in plasmas leads to the current density $\mathbf{j}'$, neglecting electron inertia†. For simplicity assume a hydrogen plasma which consists only of protons (p) and electrons (e) with charge $+e$, $-e$. Then in terms of any residual macroscopic electric field (to be determined) the electron equation of motion reduces to

$$\mathbf{j}' = ne(\mathbf{u}_p - \mathbf{u}_e) = \sigma \mathbf{E}'; \quad \sigma \approx \frac{e^2 n_e \tau_e}{m_e}, \tag{2.16}$$

where $\mathbf{u}_p$ $\mathbf{u}_e$ are proton and electron fluid velocities, and τ_e is the "electron collision time". This equation is generically called "Ohm's law".

We also eliminate $\mathbf{j}$ using $\mathbf{j} = \sigma \mathbf{E} + \mathbf{u} \times \mathbf{B}$ to arrive at equations written

† Electrons are assumed light enough that their momentum changes essentially instantaneously, compared with other terms in the momentum equation for electrons. The inertial term then drops out of the equation.

in terms of **u**, **B** σ with no explicit appearance of **j** or **E**. The equations are (Rempel, 2009)

$$\frac{\partial \rho}{\partial t} = -\nabla\cdot(\rho\mathbf{u}) \tag{2.17}$$

$$\rho\frac{\partial \mathbf{u}}{\partial t} = -\rho(\mathbf{u}\cdot\mathbf{grad}\,)\mathbf{u} - \mathbf{grad}\,p + \rho\mathbf{g} + \frac{1}{\mu_0}(\mathbf{curl}\,\mathbf{B})\times\mathbf{B} + \nabla\cdot\tau \tag{2.18}$$

$$\rho\frac{\partial e}{\partial t} = -\rho(\mathbf{u}\cdot\mathbf{grad}\,)e - p\nabla\cdot\mathbf{u} + \nabla\cdot(\kappa\mathbf{grad}\,T) + Q_\nu + Q_\eta \tag{2.19}$$

$$\frac{\partial \mathbf{B}}{\partial t} = \mathbf{curl}\,\left(\mathbf{u}\times\mathbf{B} - \eta\,\mathbf{curl}\,\mathbf{B}\right)\ . \tag{2.20}$$

Here, τ is a viscosity tensor, Q_ν and Q_η are additions to the thermal energy e from viscous and Joule heating (collisions between particles in the kinetic picture). MHD problems involve the solution of coupled partial differential equations for unknowns ρ, **u**, e, **B**, given an equation of state (relating temperature T, ρ,e, pressure p) and various additional relations written in terms of these variables needed to close the system (expressions for pressure and heat flux tensors, conductivity, various source and sink terms such as radiative or convective gains or losses, even ionization). The equations themselves are non-linear, the equation of motion containing the advection term $\mathbf{u}\cdot\nabla\rho\mathbf{u}$ and the Lorentz force $\mathbf{curl}\,\mathbf{B}\times\mathbf{B}/\mu_0$, and the induction equation containing $\mathbf{curl}\,(\mathbf{u}\times\mathbf{B})$, from the frame transformation.

2.6 Magnetic field (re-) generation

There are (at least) two major challenges in modern solar physics: (1) what causes the global regeneration of large-scale magnetic field with a period of $\approx$ 22 years, (2) why must the magnetic field appear most prominently in spots? In this section, I draw some essential points from the review by Rempel (2009) about the first question.

Dynamo theory studies conditions under which a velocity field of a highly conducting fluid can sustain a magnetic field against Ohmic decay. Rempel adopts "dynamo" to mean a system that has a magnetic energy that does not approach zero as time $t \to \infty$, owing to current systems occurring entirely within a finite volume. In the absence of induction effects, magnetic fields decay on a timescale $\tau_d = L^2/\eta$, for the Sun this is $\sim 10^9$ years, using kinetic values for conductivity (equation 2.16). The Sun is $\sim 4.5\times10^9$ years old, so the mere existence of magnetic field does not by itself require dynamo action, but the observed 22 year cycling behavior show that something other than diffusion is going on. (Rempel, 2009 notes that at least one alternative mechanism has been proposed, but these are in conflict with helioseismology). If one accepts the idea of "turbulent magnetic diffusivities" based upon turbulence, mixing length ideas or "mean field theory", which yield conductivities several orders of magnitude smaller, a dynamo is indeed required to sustain fields against this "enhanced diffusion". This issue is discussed below (section 2.6.1).

The Sun could well have both a "large scale" and a "small scale" dynamo. If most magnetic energy is in scales associated with turbulence (surface granulation for example has scales of ~ 1 Mm compared with $R_\odot = 700$ Mm) we speak of a small-scale dynamo. On the basis of Figure 2.7, 3D numerical simulations with Prandtl numbers $P_m = \nu/\eta \sim 1$ (e.g. Cattaneo, 1999), it is tempting to conclude that a small scale dynamo is active on the Sun. Caution is needed since the convection zone has $P_m \ll 1$. Almost any chaotic velocity field with large R_m produces a small-scale dynamo (Cattaneo, 1999), but a large scale dynamo requires new ingredients, such as a net mean helicity in the turbulence induced by, say Coriolis forces (Parker, 1955) and/or large scale shears.

However, MHD dynamos require at least *some* diffusion to work at all. This is clearly seen by taking the limit of ideal MHD, ($\sigma \to \infty$), in which the magnetic field is frozen to the plasma. In this case the magnetic

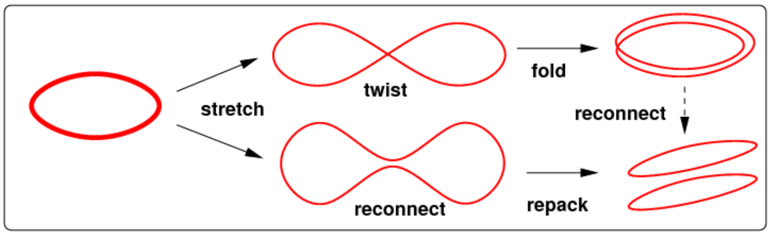


Fig. 2.13. A figure from Rempel (2009) showing two possible dynamos. Amplification occurs during stretching, the twist-fold and reconnect-repack steps remap the amplified flux-rope into the original volume element so that the process can be repeated. Both three dimensions and magnetic diffusion are *required to allow for the topology change needed to close the cycle.*

fields entrain the same elements of plasma for ever. Thus, the *topology* of the magnetic fields is fixed, once and for all in this limit, all that can be done is to entangle tubes of plasma-entraining flux between one another, like braiding hair in which the bundles of hair are separate entities. But this can only be achieved so far as the Lorentz force ($\mathbf{j}\times\mathbf{B}$) does not act to oppose the braiding. Consider Figure 2.13 from Rempel (2009), showing the operation of "flux rope" dynamos. In the absence of diffusion, repeated stretch-twistfolds simply generate more complexity (more turns of the rope in a given volume). Eventually such a state will need such strong Lorentz forces that the limited driving energy can generate no further changes. This behavior is not capable of explaining solar magnetic behavior. With diffusion, the topology can change (the dashed "reconnection" step in the figure), allowing field lines to slip across the fluid, as is required to explain, for example, the 11 year flip in magnetic polarity of global solar magnetic fields.

The role of stretching, among other velocity gradient effects, can be seen from the ideal induction equation

$$\frac{\partial \mathbf{B}}{\partial t} = \mathbf{curl}\,(\mathbf{u} \times \mathbf{B}) = -(\mathbf{u} \cdot \mathbf{grad}\,)\mathbf{B} + (\mathbf{B} \cdot \mathbf{grad}\,)\mathbf{u} - \mathbf{B}\,\nabla \cdot \mathbf{u}. \tag{2.21}$$

The first term on the right hand side describes advection of magnetic field, the second amplification by shear (including stretching) and third by compression.

Helioseismology has revealed large-scale velocity shears in the solar interior. Radial shear exists near the interface of the core and convection zone ("the tachocline"), near the surface; latitudinal shear exist at intermediate latitudes throughout the convection zone (Schou *et al.*, 1998). Such sheared regions naturally lead to field amplification through equation (2.21). Figure 1 of Spruit (2011) is reproduced in Figure 2.14. It shows the linear growth rates of toroidal field starting from a uniform poloidal field, driven by a fit to surface differential rotation (a rough approximation to interior rotation through the solar convection zone). But, as Spruit (2011) emphasizes:

> "The common ingredient . . . is the generation of a toroidal (azimuthally directed) field by stretching (winding-up) of the lines of a poloidal field. . . This is "the easy part". . . To produce a cyclic, self-sustained field as observed there must be a second step that turns some of the toroidal field into a new poloidal component, which is again wound up, completing a field-amplification cycle that becomes independent of initial conditions."

(Rempel, 2009, sections 3.3.7 - 3.4.3) summarizes these ideas using an example of a cylindrically symmetric non-dynamo and dynamos based upon scale separation, the "mean field" dynamos, based on work by Rädler (1980). Equations for the evolution of the large-scale or "mean" fields $\overline{\mathbf{u}}$ and $\overline{\mathbf{B}}$ are written in terms of the mean fields themselves and the correlations between the small scale velocity and magnetic fields, $\overline{\mathbf{u}', \mathbf{B}'}$. The latter are, in turn, written in terms of the large scale fields to provide closure, based upon several strong assumptions, but drawing upon several important symmetry

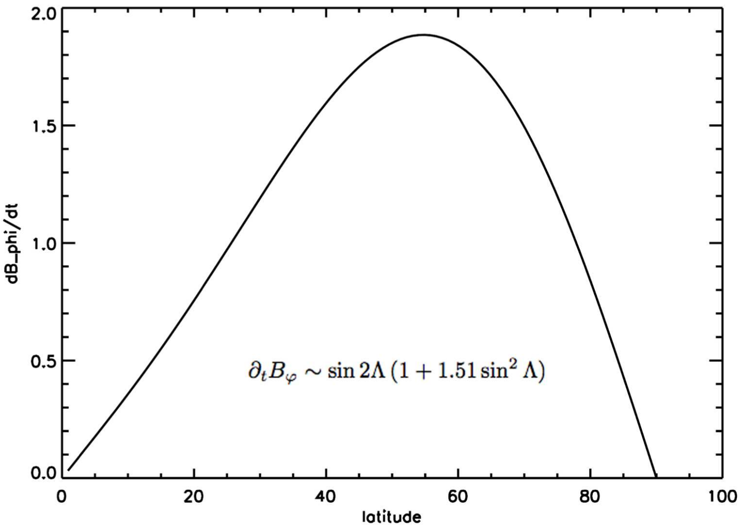


Fig. 2.14. A figure from Spruit (2011) showing the growth rates $\frac{\partial \mathbf{B}}{\partial t}$ of toroidal field, starting from a uniform poloidal field stretched by observed surface latitudinal differential rotation, neglecting any Lorentz force. Is this simple figure a key ingredient in the solar dynamo?

properties. In a nutshell, the mean field equations yield

$$\frac{\partial \overline{\mathbf{B}}}{\partial t} = \mathbf{curl}\left(\overline{\mathbf{u}' \times \mathbf{B}'} + \overline{\mathbf{u}} \times \overline{\mathbf{B}} - \eta\, \mathbf{curl}\, \overline{\mathbf{B}}\right) , \tag{2.22}$$

where a "new" electromotive force (EMF) term arises compared with the full induction equation:

$$\overline{\mathcal{E}} = \overline{\mathbf{u}' \times \mathbf{B}'} \,. \tag{2.23}$$

Taking into consideration symmetries, the EMF can be written in terms of tensor turbulent transport coefficients $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ and $\boldsymbol{\delta}$ as

$$\overline{\mathcal{E}} = \boldsymbol{\alpha}\overline{\mathbf{B}} + \boldsymbol{\gamma} \times \overline{\mathbf{B}} - \boldsymbol{\beta}\, \mathbf{curl}\, \overline{\mathbf{B}} - \boldsymbol{\delta} \times (\mathbf{curl}\, \overline{\mathbf{B}}) + \ldots \tag{2.24}$$

Making further physical assumptions with various levels of justifiability, the tensors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ become diagonal. With τ_c the correlation time of components of $\mathbf{u}'$ Rempel writes (his eqns. 3.59, 3.60)

$$\alpha = \frac{1}{3}\alpha_{ii} = -\frac{1}{3}\tau_c\, \overline{\mathbf{u}' \cdot (\mathbf{curl}\, \mathbf{u}')} \;, \;\; \eta_T = \frac{1}{3}\beta_{ii} = \frac{1}{3}\tau_c\, \overline{\mathbf{u}'^2} \;, \;\; \boldsymbol{\gamma} = -\frac{1}{2}\mathbf{grad}\, \eta_T \;. \tag{2.25}$$

The α-effect is related to the kinetic helicity of the flow, H_k:

$$H_k = \mathbf{u} \cdot (\mathbf{curl}\, \mathbf{u}) \ , \tag{2.26}$$

while the turbulent magnetic diffusivity is proportional to the intensity of the turbulence. In fact, $\eta_T \sim Lu_{rms} \sim R_m \eta \gg \eta$ (Rempel 2009, his equation 3.63), and η_T describes the *transport* of magnetic energy via *advection* and some *reconnection* of field (section 2.7), the latter requiring the development of small scales to permit ultimate dissipation via Ohmic collisions.

Under the influence of rotation at angular velocity $\mathbf{\Omega}$, and when stratification exists (under a gravity vector $\mathbf{g}$), then

$$\alpha \approx \alpha_0(\mathbf{g} \cdot \mathbf{\Omega}) = \tau_c^2 v_{\mathrm{rms}}^2 \mathbf{\Omega} \cdot \mathbf{grad} \ \ln(\varrho v_{\mathrm{rms}}) \tag{2.27}$$

Note that $\Omega\tau_c$ is the Rossby number shown on the abscissa of 2.12.

In essence, the α-effect "fixes" the problem quoted from Spruit above by adding the following term $\boldsymbol{\alpha}\mathbf{B}$ to the induction equation

$$\frac{\partial \overline{\mathbf{B}}}{\partial t} = \ldots + \mathbf{curl} \ \left(\alpha \overline{\mathbf{B}}\right) = \ldots + \alpha \mu_0 \overline{\mathbf{j}} \ , \tag{2.28}$$

i.e. the induced magnetic field is proportional to the mean current. The new term converts poloidal magnetic field into toroidal field and vice-versa, so there exists in principle the following dynamo scenario:

$$\mathbf{B}_t \xrightarrow{\alpha} \mathbf{B}_p \xrightarrow{\alpha} \mathbf{B}_t \ , \tag{2.29}$$

called an "α^2-dynamo". Preferred α^2-dynamo modes are stationary solutions with poloidal and toroidal field of comparable amplitude, clearly not compatible with solar data.

Returning to the Sun, its known interior shear profiles and essential properties of time-dependent large scale magnetic fields (2.4.4), we can see that a combination of shear-driven amplification (Ω-effect) and an α effect might in principle contain essential ingredients of a solar dynamo: to deliver strong toroidal fields at the surface in the form of emerging ropes of flux generated by the Ω effect, and an α effect to complete the cycle of events needed for field reversals. Such models are epitomized by those of (e.g. Dikpati and Charbonneau, 1999). In this case, the α effect producing toroidal field is assumed smaller than the Ω effect, so that we have

$$\mathbf{B}_t \xrightarrow{\alpha} \mathbf{B}_p \xrightarrow{\Omega} \mathbf{B}_t \ , \tag{2.30}$$

In such a dynamo, "waves" of amplified field propagate along $\mathbf{grad}\,\mathbf{\Omega}$- to explain the latitudinal migration of spots (butterfly diagram) this requires a radial gradient in $\mathbf{\Omega}$. This basic property of $\alpha\Omega$ dynamo-waves is pitted against recent arguments by Spruit in the next section. Lastly, the "tight relationship" with Rossby number plotted in figure 2.12 should not be regarded as supporting mean field dynamo models *per se*, given the underlying assumptions as stressed already by (Noyes *et al.*, 1984).

2.6.1 What might be the ingredients in the solar dynamo?

In order that an $\alpha\Omega$ dynamo be consistent with the observed 22 year periodic change in global solar magnetic field, the cycle of events must take place *fast*: the speed at which the cycle eq. (2.30) is limited by the slowest critical physical process. The differential rotation operates fast in the Ω part of the cycle, the ratio $\Delta\Omega/\Omega$ is ~ 0.1, so the time t where $\Delta\Omega t \sim 1$ is about 10 rotations, the best part of a year. A critical part of all dynamos is not just the existence of an α-effect but the accompanying β-effect which helps promote the changes of topology by generating small scales needed for kinetic diffusivities to become important. In the mean field theory this process is modeled through the "turbulent diffusivity" (equation 2.25), which is equivalent to an "eddy diffusivity" of $\frac{1}{3}\lambda v'$. Conveniently, the "mixing length" λ and v' values yield $\eta_T \sim 10^{12}$ cm^2 s^{-1}, orders of magnitude larger than kinetic values. Such values are close to those needed in dynamo models to produce cycles near 11 years in length, in this formalism.

So is everything in order? Not according to Parker (2009); Spruit (2011), who have independently re-assessed the ingredients needed to make a dynamo compatible with physical principles and observations. Both have pointed to problems with the mean field/ turbulent dynamo concept. Parker points out that the concept of a turbulent diffusivity at high R_m, valid in the kinematic regime, is likely invalid in the dynamic regime. By considering the magnetic fields to be frozen up til the development of scales small enough for molecular diffusion to arise, Parker argues that the diffusion occurs when the field strength has increased by $R_m^{1/2}$. However, by the time this occurs, the magnetic stresses ($\propto B^2$) will have increased by R_m. Thus, for fields of interest, the magnetic stresses overwhelm any turbulence long before resistive diffusion can obliterate the field. In the language of mean field theory, Parker concludes that

> "The idea that the azimuthal [i.e. toroidal] magnetic field is subject to ordinary turbulent diffusion, with $\eta = \frac{1}{3}u\lambda$, seems unjustified ... There is no way to account for the value $\eta_T \sim 10^{12}$ cm^2 s^{-1}, suggesting that it is necessary to re-think the $\alpha\Omega$-dynamo for the Sun."

It should be noted however that magnetic stresses can both hinder and enhance small scale dynamics, indeed there is evidence from numerical work that on small scales there may not be such a dynamical reduction of diffusive-like dynamics (Rempel, personal communication 2014).

Spruit (2011) offers a scathing critique of mean field dynamo theory, arguing instead for a return to models of the 1960s (Babcock, 1961, 1963; Leighton, 1969). In the 1960s nothing was known about interior rotation profiles of the Sun, in these early works the surface (i.e. latitudinal) rotation profile was adopted for simplicity. In Spruit's view, latitudinal differential rotation generates toroidal field in the interior, which becomes unstable to buoyancy. The dynamic buoyant rise of these toroidal ropes of magnetic flux is subject to the Coriolis force generating automatically poloidal field. As shown in Figure 2.14 the first flux is expected to emerge near 55° latitude, and later flux emerges polewards and equatorwards. Some properties of Figure 2.10 might be considered qualitatively (if not quantitatively) compatible with this picture. Spruit cites evidence (albeit indirect) for such behavior, other support has recently been reported by McIntosh (private communication 2014). Convective turbulence plays no active role in the induction equation, instead it provides stresses needed to maintain latitudinal differential rotation.

Spruit's concerns can be summarized as follows:

- The observed evolution of sunspots points to processes which are not diffusive in nature (see quote in section 2.4.3). The observations reveal intense bundles of flux that must have been assembled by deterministic processes from below.
- Various observed properties of emerging flux appear compatible with the simple notion that the solar cycle reflects the rate of generation of toroidal field by latitudinal differential rotation (Figure 2.14). There remain important questions (why do sunspots form only at low latitudes?).
- The radial gradient ("tachocline") in rotation profile at the base of the convection zone can do little work to stretch the magnetic field, the lower boundary being stress-free in the radiative core.
- Instead the latitudinal gradient found in helioseismology – similar to the surface differential rotation profile known by Babcock and Leighton in the 1960s – is a more promising source for the Ω-effect.
- The physical justifications underlying mean field theories are weak.
- Hale's and Joy's laws imply, albeit indirectly, that a strong (super-equipartition) field is required to survive in the convection zone.
- Such a field, roughly 10^5 G in the depths of the convection zone, is also an estimate of the field strength needed for buoyant eruption (e.g. Schüssler,

1996). (Recent work has relaxed this estimate to a few tens of kG, Weber *et al.*, 2013). Several observed properties of sunspots seem compatible with such "initial conditions" for the formation of sunspots.

At least two open questions remain, one being how to maintain the toroidal field in the latitudinal shear layer as it builds up slowly to become buoyant, against the strong convective eddy motions. Spruit cites another puzzle:

> The most challenging problem may well be finding a satisfactory description for the process by which the mass of buoyant vertical flux tubes resulting from a cycle's worth of eruptions gets 'annealed' back into a simpler configuration... appeal to traditional convective 'turbulent diffusion' will not work (even if the concept itself is accepted), since the fields are now much stronger than equipartition with convection (at least near the base of the convection zone where this annealing has to take place).

I conclude that while we have different cookbooks and different recipes to make a solar cycle-like model, we still seem to be missing some key ingredients in explaining how the Sun and other stars re-generate their global magnetic fields.

2.7 Magnetic field dissipation, topology, reconnection

By "dissipation", I refer to the conversion of ordered forms of energy to disordered, thermal energy. In MHD, magnetic energy is dissipated directly through Ohmic dissipation (the ∇^2 term in the induction equation) in which the ordered differential motion of electrons and ions (or other neutral species when present), i.e. electric currents, is converted to random motion, heat, via particle collisions. The rate of conversion of this energy, $\mathbf{j} \cdot \mathbf{E}$ is merely

$$\frac{1}{2\mu_0}\frac{\partial B^2}{\partial t} = j^2/\sigma = (\mathbf{curl}\,\mathbf{B})^2/\mu_0^2\sigma \tag{2.31}$$

Of course the rate of destruction of magnetic field is

$$\frac{\partial \mathbf{B}}{\partial t} = \frac{1}{\mu_0\sigma}\nabla^2\mathbf{B} = \eta\nabla^2\mathbf{B} \tag{2.32}$$

Manifestly both processes involve the generation of small scales to be effective. But to dissipate energy on large scales requires a mechanism naturally generating such small scales. In the limit $R_m \to \infty$, *topology* becomes terribly important, because, owing to Alfvén's theorem plasma is nearly frozen to specific tubes of flux. One can imagine an initial condition of a star threaded by magnetic field, say in a simple dipolar configuration. Draw a closed path around some magnetic flux, the field lines on this path trace out a tube which will close back on itself (there are no monopoles). As time goes on, convection, rotation, shear tries will make this field very untidy, but Alfvén's theorem implies that although plasma may be forced to move, the topology of this tube cannot change- to make knots in this requires one to cut the flux tube, but this is forbidden in ideal MHD. The topological constraints are of a global, not local, nature, in contrast with the local terms in the equation of motion.

To change magnetic topology requires diffusion, i.e. the development of smaller physical scales. *Most of the solar- and solar-terrestrial phenomena of interest involve the evolving magnetism and its changing topology*, even on the largest scales. Like cutting through rubber bands, the "snip" only needs to occur *somewhere* along a tube of magnetic flux, not along the entire length. Enormous changes $\frac{\partial \mathbf{B}}{\partial t}$ can occur because of a "reconnection" (a "snip" that satisfies $\operatorname{div} \mathbf{B} = 0$) somewhere else along a tube, allowing the system to reach a lower energy state in a state with a quite different topology.

Magnetic reconnection in MHD is the study of changing topology of magnetic fields. Accordingly, one would think that the diffusivity η plays a central role. But in fact η plays a secondary role, something that became clear when discussing the β effect in mean field dynamos. In that case, correlations between velocity and magnetic perturbations lead to values η_T that are associated with the "turbulence"- parcels of fluid are moved bodily with speed u a mixing length λ to produce $\eta_T = \frac{1}{3} u\lambda$ to produce changes in the magnetic field orders of magnitude faster than are possible using kinetic values for η. In this picture, field lines frozen to fluid elements are advected by them without hindrance by the Lorentz force (in the kinetic regime of mean field theory), shuffling them around. No reconnection is implied in the mean field theory by η_T, this term originates from the $\mathbf{curl}\,(\mathbf{u} \times \mathbf{B})$ term. It is merely a re-shuffling, which, in MHD, can lead to reconnection if and when it generates scales small enough for genuine diffusion to occur via the kinetic value of η.

Important first steps in magnetic reconnection were taken by Dungey (1953), Sweet (1958), Parker (1957). The famous 2D "Sweet-Parker" model, based upon a stationary flow bringing opposite polarity, collision-dominated plasmas together, yields a diffusion rate $\sim V_A/\sqrt{R_m}$, which, through R_m, depends explicitly on η. A modification by Petschek (1964) introduced slow shocks to allow the bulk of the frozen field in the Sweet Parker picture to avoid having to flow through the diffusion region, permitting the diffusion region to be much shorter. In Petschek's scenario, much faster reconnection rates $\sim V_A/\log_e R_m$, almost independent of η, were achieved! Such reconnection scenarios are called "fast" reconnection (i.e.. almost independent of

R_m), they are *required* to account for the sudden release of energy seen in solar flares, coronal mass ejections, and indeed to explain the operation of the solar dynamo. For some time the ability of space plasmas like the solar corona to be able to reconnect at the Petschek rate (i.e. "fast" enough) was in doubt, it requiring a somewhat specific configuration and prescribed steady state in the velocity field. Curiously, tearing instabilities in MHD has been put forward by Bhattacharjee *et al.* (2009), may provide MHD systems with a natural tendency to approach the Petschek rate of reconnection. Reconnection physics is a huge subject and so I simply refer readers to texts by Biskamp (2005) and Priest and Forbes (2007), and a tutorial article by Kulsrud (2011).

Below, I point out some problems I find interesting concerning the evolution and dissipation of magnetic fields under conditions where turbulent and Reynolds stresses and gas pressure dominate (the high β regime), a regime that occurs in the solar interior and most parts of the solar photosphere. I also discuss Parker's "fundamental theorem of magnetostatics" as it pertains to the solar corona, which is mostly in the opposite, low plasma-β regime. Some surprises will arise. Figure 2.15 shows the dramatic change from hydrodynamically dominated turbulent structure characteristic of the dense photospheric plasma, to the magnetically dominated corona in the tenuous plasma there.

2.7.1 "High β" plasma

Surface magnetic fields on the Sun exhibit behavior that seem to imply enormous diffusivities at and near the visible surface. Generally speaking, direct observations of $\frac{\partial \mathbf{B}}{\partial t}$ are not made, instead "tracers" of magnetic fields measured through "bright points" in the photosphere or overlying coronal structures, for example. An implicit assumption is that fields evolve across some two dimensional surface, there is no flux of magnetic field through the atmosphere. In one example, Abramenko *et al.* (2011) studied trajectories of tiny bright points of magnetic origin in photospheric inter-granular lanes and derived diffusion coefficients $\propto \lambda^\gamma u$ where $\gamma = 1$ for "normal " (thermal-like) diffusion. These concentrations of magnetic field are close to the "high β" regime, so that the surface magnetic fields are assumed not only to be advected only across the surface but that they are unimpeded by magnetic forces. These studies typically yield $\eta_T \sim 3 \times 10^{12}$ cm^2 s^{-1}. Rather astonishingly, this surface diffusion coefficient is of the same magnitude needed to make a (3D) interior dynamo work in the kinematic regime (Parker, 2009)!

Does this (coincidence?) lend credibility to the idea of "turbulent dif-

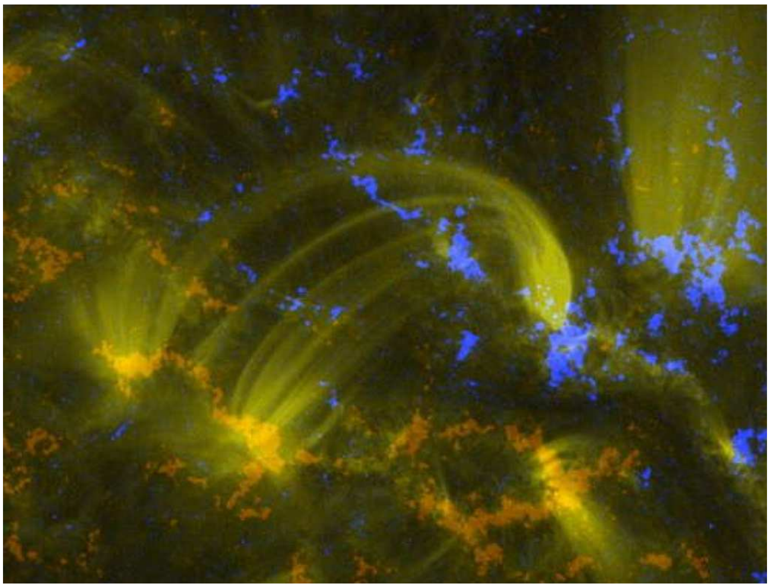


Fig. 2.15. An image showing a longitudinal magnetogram (blue and orange with opposite polarities) and some associated hot plasma in the overlying corona (green). The switch from dominant "turbulent" fluid motions seen in in the photospheric magnetogram, an example of a "high β plasma, to the far more ordered coronal loop structures arising from the dominant magnetic stresses, a "low β plasma", is dramatic. But the switch occurs in a geometrically thin, poorly understood stratified layer called the "chromosphere" which is not seen in this image. The magnetograms trace out "supergranule" cells of 20-30 Mm diameter.

fusion"? This remains an open question. The argument of Parker (2009) says "no", others (e.g. Lazarian, 2013) believe turbulent diffusion a central physical process in all plasmas with small diffusive terms. The resolution of this difference will probably come from detailed studies of direct numerical simulations, to look at when and where the small-scale Lorentz forces act so as to oppose reconnection, since they can also enhance it. Rempel (2009) discusses some of these issues in the context of mean field theory in section 3.5.2.

Curiously, it has become customary to invoke "turbulent diffusivities" to explain observations of the rotation and evolution of coronal features, such as coronal holes, in higher layers of the Sun's atmosphere where the magnetic stresses tend to dominate. It is important to understand the meaning of such work since the literature invokes diffusion and reconnection sometimes interchangeably, and at various levels in the sun's atmosphere (photosphere, corona).

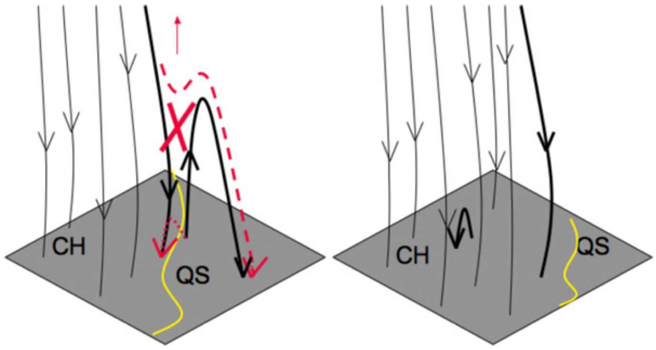


Fig. 2.16. Interchange reconnection: the location of a coronal hole and quiet Sun boundary is indicated as a line. In the left figure, a quiet Sun loop located near a CH open field. Following the reconnection (at the "x") a small loop is created inside the CH and the open field line relocated, in the right figure. This process has been invoked to explain the rigid rotation as well as the growth of coronal holes. From Krista *et al.* (2011).

On the one hand, in the work of Wang and Sheeley (1993) diffusion occurs (implicitly) in the photosphere. They solve a 2D surface induction equation using a diffusivity of 6×10^{12} cm^2 s^{-1} and macroscopic velocities based upon photospheric measurements. Consider an area originally of normal "quiet Sun" (i.e. mixed polarity) that is evolving into a coronal hole (mostly one polarity). The change in polarity must occur in surface transport models via some "cancellation" or reconnection at photospheric heights of field of opposite polarity to that of the coronal hole. Since there are no monopoles the net flux remains the same but the boundaries move. This can occur just as fast as the flows drive opposite polarities together since there is nothing to halt mutual annihilation. In this model, the corona is dealt with via a potential-field extrapolation (including a source surface at $2.5R_\odot$ for "open" field lines), in which the state of the corona is set instantaneously by the boundary conditions. By assumption no electrical currents are permitted in the coronal plasma. If an MHD calculation were performed instead using the same boundary conditions, one would impose a large value of η to dissipate all currents on timescales short compared with dynamical times, i.e. $R_m \lesssim 1$. In turn this implies very fast reconnection everywhere in the corona. Astonishingly, this simple model is found to capture important features of the evolving coronal holes under study. In the next section I discuss more recent work on effective diffusion in the corona.

2.7.2 Low β plasma: diffusion

As well as in high-β plasmas, it has become popular to describe coronal magnetic field evolution in terms of diffusive processes. For example, superficially similar work to that of Wang and Sheeley (1993) has been presented by Fisk and Schwadron (2001), followed by later work (e.g. Krista *et al.*, 2011). In these studies, observations *in the coronal plasma* have been used to infer effective diffusivities. In their studies of coronal holes, observations indicate to Fisk & Schwadron that

> "Diffusion by random convective motions in supergranules† is quite slow and will prove to be inadequate for the transport of open magnetic flux on the Sun. We introduce, therefore, a faster process."

The proposed process is called "interchange reconnection", where $\frac{\partial \mathbf{B}}{\partial t}$ changes locally at a point in the corona by the process illustrated in Figure 2.16 from Krista *et al.* (2011). It is quite a different process than the convective diffusion, it has a diffusion coefficient at least an order of magnitude larger. To describe the evolution of coronal hole boundaries, Krista *et al.* (2011) find $\eta_T \lesssim 3 \times 10^{13}$ cm^2 s^{-1}. Fisk and Schwadron (2001) however found $\eta_T \sim 3.5 \times 10^{13}$ and 1.6×10^{15} (50× larger!) cm^2 s^{-1}, based upon the apparent "jumping around" movement of closed loops, within coronal holes and near the equator respectively. The picture is that this kind of topology change corresponds to a diffusion coefficient which can approach $\frac{1}{3}\lambda v_A$ for a "jump" of length λ at the enormous Alfvén speed v_A. The new idea is that *fast reconnection occurs somewhere well above the photosphere where the Alfvén speed is high*, between open and closed regions brought slowly together by the relatively slow photospheric footpoint motions. Instead of annihilation of fields at the photosphere, a smaller loop structure, not reaching coronal heights, remains (Figure 2.16).

This is an interesting departure from the "mean field" picture since magnetic forces ostensibly cannot be neglected, so this effect cannot be described by kinematics of hydrodynamic turbulence. What, then, does a "turbulent diffusivity" mean in the low-β regime? This example differs not only in plasma β from the photosphere, but also in that implies that *reconnection is very fast*.

There are observable consequences of such large values of η occurring near the coronal base. Using $\eta \sim \frac{1}{3}L^2/t$ we can turn this around and say that such diffusion will limit the lifetime of any solar features to $t \lesssim L^2/\eta$. Modern telescopes resolve features down to about 10^7 cm on the Sun. With $\eta = 3 \times 10^{13}$ cm^2 s^{-1}, the lifetime is $\lesssim$ 3s. There are many fine scale

† Simon *et al.* (1995) find $\eta_T \sim 5 - 7 \times 10^{12}$ cm^2 s^{-1} for supergranules.

structures living far longer than this seen near the coronal base in quiet and coronal hole regions (e.g. de Pontieu *et al.*, 2007). Additionally, if we allow that say 1 day is needed for a $L \sim 20$ Mm-sized active region to build up enough free energy within the corona so that a large reconnection event can trigger a flare, then $\eta \lesssim \frac{1}{3}L^2/(1 \text{ day})$, or about 10^{13} cm^2 s^{-1}. It seems that interchange reconnection, if real, does not happen everywhere, all the time, in the Sun's corona. If it does happen, then we must explain how it occurs so rapidly in the low-β environment at far higher rates than can be driven by photospheric diffusion driven by turbulence there. Is there a process which can drive such dynamic changes?

2.7.3 Low β plasma: reconnection, flares, heating

The "coronal heating problem" is a 70+ year old problem that is alive and well. It is a long recognized "grand challenge" for astronomy (e.g. Hoyle, 1955). Several lines of argument indicate the magnetic field as a prime source of energy (e.g., see Figure 2.15). Why is this problem so "refractory", resistant to a clean solution? After all, usually a mere 1 part in 10^6 or so of the Sun's energy flux is used to maintain a corona. Part of the problem is precisely *because* of this small energy requirement. The problem is "ill posed" in that unobservably small changes in regions of the atmosphere where we can accurately measure components of energy fluxes (via hydrodynamic or electrodynamic processes) can lead to order zero changes in the energy flux into the corona! Indeed, photospheric changes associated with large flares has only in recent years been detected through spectropolarimetry.

We have an embarrassment of riches in ideas to convert ordered magnetic energy into heat (e.g. Parnell and De Moortel, 2012). But the bulk of the mechanisms rely on the same process- the development of small scale structures in the large coronal volume in order to permit either MHD (Ohmic heating, viscous heating) or non-MHD processes (wave particle interactions?) to dissipate magnetic energy. Instead of focusing on individual mechanisms, I draw attention to a general property of low β, highly conducting plasma, first discovered by Parker (1972); Parker (1994).

In the limit of "zero β", the Lorentz force is the only game in town. Parker asks us to consider that a force-free state exists between two infinitely conducting plates at $z = 0$, $z = L$, filled with an infinitely conducting plasma, then

$$\mathbf{j} \times \mathbf{B} = 0 \quad \text{(force} - \text{free condition)} \tag{2.33}$$

which is non-trivially satisfied when

$$\mathbf{curl\,B} = \alpha(\mathbf{r})\mathbf{B} \tag{2.34}$$

where $\alpha(\mathbf{r})$ is a measure of twist. Boundary values of B_z are specified. The ideal induction equation endows the magnetic field with a specific, fixed topology. An operational approach is taken, first to obtain the set of all *continuous* force-free fields satisfying the boundary conditions. These fields typically have different topologies, one of the fields being the unique potential field ($\alpha = 0$). The second step is to search the set $S\{\mathbf{B}_\alpha\}$ for the solution $\mathbf{B}_\alpha^{sol}$ with the correct topology τ.

Then Parker's "Fundamental Theorem of Magnetostatics" says that, in general, the solution $\mathbf{B}_\alpha^{sol}$ must be *discontinuous.* The relevance to the Sun is that, on the one hand, one can imagine the two plates being the solar "photosphere" and the intermediate plasma as the "corona", the corona living long enough for magnetostatic equilibrium to be a reasonable approximation outside of obvious dynamic events. The relevance to the heating, reconnection, flaring problems should be clear also, in that the theorem enforces the formation of magnetic discontinuities in anything other than highly tidy geometries. This is precisely the property – the systematic development of small from large scales – needed to account for dissipation and reconnection. In the physical situation on the Sun, this implies that other physics must be included- *ideal MHD causes its own demise.* In trying to become singular to satisfy the partial differential equations as well as the integral equations imposing the field-line topology, ideal MHD must break down.

On the other hand, solar plasmas are not infinitely conducting, the photosphere-corona interface involves the complications of the chromosphere, which is not force-free. Much has been discussed regarding this theorem far beyond these simple comments, but closer inspection of mathematics (Low, 2010; Janse *et al.*, 2010) and some highly non-dissipative numerical simulations (Bhattacharyya *et al.*, 2010) have revealed evidence in support of the theorem. The reader should refer to the cited papers for the formal proofs, but they can also be understood using two heuristic arguments. Using the abstract of the paper by Janse *et al.* (2010),

"A magnetic field embedded in a perfectly conducting fluid and rigidly anchored at its boundary has a specific topology invariant for all time. Subject to that topology, the force-free state of such a field generally requires the presence of tangential discontinuities (TDs). This property proposed and demonstrated by Parker [Spontaneous Current Sheets in Magnetic Fields (Oxford University Press, New York, 1994)] is explained in terms of (i) the over-determined nature of the magnetostatic partial differential equations nonlinearly coupled to the integral equations imposing the field topology and (ii) the hyperbolic nature of the partial differential equation for the twist function α of the force-free field."

To see the origin of point (ii), simply take the divergence and curl of equation (2.34).

$$\mathbf{B} \cdot \mathbf{grad}\, \alpha = 0 \tag{2.35}$$

$$\nabla^2 \mathbf{B} + \alpha^2 \mathbf{B} = \mathbf{B} \times \mathbf{grad}\, \alpha \tag{2.36}$$

The second equation has complex characteristics, but the first implies that α is real along field lines, showing that along field lines the characteristics are real. There is therefore *no requirement that* $\alpha(\mathbf{r})$ *be continuous from one field line to the next.* Different neighboring values of α clearly must lead to some discontinuity since the twist is different at neighboring points.

In 1988, Parker proposed that this tendency for natural MHD systems near force-free equilibrium to form discontinuities is an essential ingredient in the physics of coronal heating. The consequence of this tendency is, Parker argues, a theory for coronal heating – "nanoflares"(Parker, 1988). This tendency may explain why "potential fields" (Wang and Sheeley, 1993) can often resemble the solar corona, in spite of the fact that there is no free energy in the potential field to produce the corona. The nanoflares might just release this energy on small scales, giving the appearance of a near-potential configuration on observable scales.

Lastly, Parker's work suggests the omnipresence of small scale reconnections in untidy physical systems. As such it may provide some rationale behind at least some puzzles such as "interchange reconnection" which appear to have neither a strong physical basis nor critical observational support.

2.8 Concluding remarks

I leave it to the reader to decide if the *history* of dynamo theory, as related by Parker (2009) and Spruit (2011), is a *tragedy*, or perhaps a *comedy.* The Parker theorem on magnetostatics reminds me personally of a classic, tragic Shakespearean hero, being defeated by its own, fundamental character flaw. What is different from MacBath and Lady MacBeth, is that the "flaw" in ideal MHD might just lead to the correct explanation of much "coronal heating", a problem of a mere 70 years, but a significant one.

For my own part, I have learned to take seriously the weight of observational evidence and the ingenuity of theoretical scientists to understand the true implications of what the truly critical observations imply. This is one way of describing the "Scientific Method". I believe solar physics is, and will remain for some time, an "observationally-driven" area of research. Either way, it is an ongoing, entertaining drama.

"I wish your horses swift and sure of foot; And so I do commend you to their backs. Farewell." – MacBeth

Acknowledgments

I am grateful to Roberto Casini, Matthias Rempel and Boon Chye Low for sharing material and discussions, and to the organizers for inviting me and supporting my participation in the school. Matthias Rempel carefully read the manuscript, for which I am thankful.

Bibliography

Abramenko, V. I., Carbone, V., Yurchyshyn, V., Goode, P. R., Stein, R. F., Lepreti, F., Capparelli, V., and Vecchio, A.: 2011, *Astrophys. J.* **743**, 133

Alfven, H.: 1950, *Cosmical electrodynamics*, International Series of Monographs on Physics, Oxford: Clarendon Press

Babcock, H. W.: 1953, *Astrophys. J.* **118**, 387

Babcock, H. W.: 1961, *Astrophys. J.* **133**, 572

Babcock, H. W.: 1963, *Ann. Rev. Astron. Astrophys.* **1**, 41

Babcock, H. W. and Babcock, H. D.: 1952, *Publ. Astron. Soc. Pac.* **64**, 282

Baliunas, S. L., Donahue, R. A., Soon, W. H., Horne, J. H., Frazer, J., Woodard-Eklund, L., Bradford, M., Rao, L. M., Wilson, O. C., Zhang, Q., Bennett, W., Briggs, J., Carroll, S. M., Duncan, D. K., Figueroa, D., Lanning, H. H., Misch, T., Mueller, J., Noyes, R. W., Poppe, D., Porter, A. C., Robinson, C. R., Russell, J., Shelton, J. C., Soyumer, T., Vaughan, A. H., and Whitney, J. H.: 1995, *Astrophys. J.* **438**, 269

Beer, J., McCracken, K., and von Steiger, R.: 2012, *Cosmogenic Radionuclides*

Bhattacharjee, A., Huang, Y.-M., Yang, H., and Rogers, B.: 2009, *Physics of Plasmas* **16(11)**, 112102

Bhattacharyya, R., Low, B. C., and Smolarkiewicz, P. K.: 2010, *Physics of Plasmas* **17(11)**, 112901

Birkeland, K. and Muir, J. t.: 1908, *On the Cause of Magnetic Storms and the Origin of Terrestrial Magnetism*, Vol. 1 of The Norwegian Aurora Polaris Expedition, 1902-03, H. Aschehoug

Biskamp, D.: 2005, *Magnetic Reconnection in Plasmas*, Cambridge University Press, Cambridge, UK

Braginskii, S. I.: 1965, *Reviews of Plasma Physics.* **1**, 205

Carrington, R. C.: 1859, *Mon. Not. R. Astron. Soc.* **20**, 13

Casini, R. and Landi Degl'Innocenti, E.: 2008, *Plasma Polarization Spectroscopy*, Chapt. 12. Astrophysical Plasmas, 247, Springer

Cattaneo, F.: 1999, *Astrophys. J. Lett.* **515**, 39

Comte, A.: 1835, *Traité de philosophie positive,* available at gallica.bnf.fr, Vol. 2

de Pontieu, B., McIntosh, S., Hansteen, V. H., Carlsson, M., Schrijver, C. J., Tarbell, T. D., Title, A. M., Shine, R. A., Suematsu, Y., Tsuneta, S., Katsukawa, Y., Ichimoto, K., Shimizu, T., and Nagata, S.: 2007, *Publ. Astron. Soc. Japan* **59**, 655

Dikpati, M. and Charbonneau, P.: 1999, *Astrophys. J.* **518**, 508

Dungey, J. W.: 1953, *Phil. Mag.* **44**, 725

Eddy, J. A.: 1976, *Science* **192**, 1189

Edlén, B.: 1943, *Zeitschrift für Astrofysik* **22**, 30

Einstein, A.: 1905, *Annalen der Physik* **322**, 891

Ferraro, V. C. A. and Plumpton, C.: 1966, *An introduction to magneto-fluid mechanics*, Clarendon Press, Oxford, 2nd edition

Fisk, L. A. and Schwadron, N. A.: 2001, *Astrophys. J.* **560**, 425

Gandorfer, A. M. and Povel, H. P.: 1997, *Astron. Astrophys.* **328**, 381

Grotrian, W.: 1939, *Naturwissenschaften* **27**, 214

Hale, G. E.: 1908, *Astrophys. J.* **28**, 315

Hale, G. E., Ellerman, F., Nicholson, S. B., and Joy, A. H.: 1919, *Astrophys. J.* **49**, 153

Hanle, W.: 1924, *Zeitschrift fur Physik* **30**, 93

Hathaway, D. H.: 2010, *Living Reviews in Solar Physics* **7**, 1

Hoyle, F.: 1955, *Frontiers of Astronomy*, ??

Janse, A. M., Low, B. C., and Parker, E. N.: 2010, *Physics of Plasmas* **17(9)**, 092901

Jeffries, J., Lites, B., and Skumanich, A.: 1989, *Astrophys. J.* **343**, 920

Judge, P.: 2002, in *ASP Conf. Ser. 277: Stellar Coronae in the Chandra and XMM-NEWTON Era*, 45

Judge, P. G. and Thompson, M. J.: 2012, in C. H. Mandrini and D. F. Webb (Eds.), *IAU Symposium*, Vol. 286 of *IAU Symposium*, p. 15

Kirchhoff, G. and Bunsen, R.: 1860, *Annalen der Physik* **186**, 161

Kleint, L., Berdyugina, S. V., Shapiro, A. I., and Bianda, M.: 2010, *Astron. Astrophys.* **524**, A37

Krista, L. D., Gallagher, P. T., and Bloomfield, D. S.: 2011, *Astrophys. J. Lett.* **731**, L26

Kulsrud, R. M.: 2011, *Physics of Plasmas* **18(11)**, 111201

Landi Degl'Innocenti, E.: 2013, *Memorie della Societa Astronomica Italiana* **84**, 391

Lazarian, A.: 2013, *Space Sci. Rev.*

Leighton, R. B.: 1969, *Astrophys. J.* **156**, 1

Lites, B. W.: 1987, *Applied Optics* **26**, 3838

Lites, B. W.: 2000, *Reviews of Geophysics* **38**, 1

Lites, B. W., Elmore, D. F., and Streander, K. V.: 2001, in M. Sigwarth (Ed.), *Advanced Solar Polarimetry – Theory, Observation, and Instrumentation*, Vol. 236 of *Astronomical Society of the Pacific Conference Series*, 33

Low, B. C.: 2010, *Astrophys. J.* **718**, 717

Maunder, E. W.: 1904, *Mon. Not. R. Astron. Soc.* **64**, 747

McCracken, K., Beer, J., Steinhilber, F., and Abreu, J.: 2013, *Space Sci. Rev.* **176**, 59

Michelson, A. A. and Morley, E. W.: 1887, *Sidereal Messenger, vol. 6, pp.306-310* **6**, 306

Noyes, R. W., Hartmann, L. W., Baliunas, S. L., Duncan, D. K., and Vaughan, A. H.: 1984, *Astrophys. J.* **279**, 763

Parker, E. N.: 1955, *Astrophys. J.* **122**, 293

Parker, E. N.: 1957, *J. Geophys. Res.* **62**, 509

Parker, E. N.: 1965, *Plan. Space Sci.* **13**, 9

Parker, E. N.: 1972, *Astrophys. J.* **174**, 499

Parker, E. N.: 1988, *Astrophys. J.* **330**, 474

Parker, E. N.: 1994, *Spontaneous Current Sheets in Magnetic Fields with Application to Stellar X-Rays*, International Series on Astronomy and Astrophyics, Oxford University Press, Oxford

Parker, E. N.: 2007, *Conversations on Electric and Magnetic Fields in the Cosmos*, Princeton University Press, Princeton, NJ USA

Parker, E. N.: 2009, *Space Sci. Rev.* **144**, 15

Parnell, C. E. and De Moortel, I.: 2012, *Royal Society of London Philosophical Transactions Series A* **370**, 3217

Petschek, H. E.: 1964, *NASA Special Publication* **50**, 425

Priest, E. and Forbes, T.: 2007, *Magnetic Reconnection*, Cambridge University Press, Cambridge, UK

Rädler, K.-H.: 1980, *Astronomische Nachrichten* **301**, 101

Rempel, M.: 2009, *Heliophysics I: Plasma Physics of the Local Cosmos*, Chapt. 3. Creation and destruction of magnetic fields, 42, Cambridge University Press

Ribes, J. C. and Nesme-Ribes, E.: 1993, *Astron. Astrophys.* **276**, 549

Rosner, R., Tucker, W. H., and Vaiana, G. S.: 1978, *Astrophys. J.* **220**, 643

Saha, M. N.: 1937, *Proc. R. Soc. Lond. A.* **160**, 155

Schou, J., Antia, H. M., Basu, S., Bogart, R. S., Bush, R. I., Chitre, S. M., Christensen-Dalsgaard, J., di Mauro, M. P., Dziembowski, W. A., Eff-Darwich, A., Gough, D. O., Haber, D. A., Hoeksema, J. T., Howe, R., Korzennik, S. G., Kosovichev, A. G., Larsen, R. M., Pijpers, F. P., Scherrer, P. H., Sekii, T., Tarbell, T. D., Title, A. M., Thompson, M. J., and Toomre, J.: 1998, *Astrophys. J.* **505**, 390

Schrijver, C. J. and Harvey, K. L.: 1989, *Astrophys. J.* **343**, 481

Schüssler, M.: 1996, in K. C. Tsinganos (Ed.), *Solar and Astrophysical Magneto-hydrodynamic flows*, Kluwer, p. 17

Schwabe, H.: 1844, *Astron. Nachr.* **21(495)**, 233

Seagraves, P. H. and Elmore, D. F.: 1994, *Proc. SPIE* **2265**, 231

Shibata, K., Isobe, H., Hillier, A., Choudhuri, A. R., Maehara, H., Ishii, T. T., Shibayama, T., Notsu, S., Notsu, Y., Nagao, T., Honda, S., and Nogami, D.: 2013, *Publ. Astron. Soc. Japan* **65**, 49

Simon, G. W., Title, A. M., and Weiss, N. O.: 1995, *Astrophys. J.* **442**, 886

Spruit, H. C.: 2011, in Miralles, M. P. & Sánchez Almeida, J. (Ed.), *The Sun, the Solar Wind, and the Heliosphere*, IAGA Special Sopron Book Series, Vol. 4. Berlin: Springer, 39

Sweet, P. A.: 1958, in B. Lehnert (Ed.), *Electromagnetic Phenomena in Cosmical Physics*, Vol. 6 of *IAU Symposium*, 123

Wang, Y.-M. and Sheeley, Jr., N. R.: 1993, *Astrophys. J.* **414**, 916

Weber, M. A., Fan, Y., and Miesch, M. S.: 2013, *Solar Phys.* 287

Young, C. A.: 1892, *The Sun*, Kegan Paul, Trench, Trübner & co. Ltd., London, 4th edition