The Solar Dynamo and our attempts to understand it

Paul Bushby and Joanne Mason

April 29, 2004

Ever becoming more tightly constrained by observations and theory, the solar dynamo model that we have now bears little resemblance to the model studied just 30 years ago, and furthermore, it is likely to be modified in the near future. Solar dynamo theorists have learnt much over recent years, however we are still far from a complete understanding of how the Sun generates its magnetic field. Moreover, based upon our understanding of turbulence and chaotic systems, it seems unlikely that we will ever be able to reliably predict the Sun's magnetic behaviour more than a few years in advance. So what is the solar dynamo, how much do we understand, and why is it so difficult to model?

Solar magnetic activity is observed on a wide range of scales, from the intriguingly regular well-ordered large-scale field, to much more complex small-scale structures. In this article we shall primarily be concerned with the large-scale features, and the associated large-scale dynamo which is responsible for their occurrence. The small-scale magnetic field at the photosphere may also owe its existence to a dynamo mechanism; however, dynamo action of this form is highly localised and is very different in character from the large-scale dynamo, and so the two issues are largely addressed separately. More details regarding small-scale dynamo action can be found in the recent *Astronomy & Geophysics* review article by Cattaneo and Hughes (2001).

Sunspots are perhaps the most widely studied and well-documented surface manifestations of the Sun's magnetic field; due to their intense magnetic field, they appear as dark regions on the solar surface (see figure 1). Sunspots have been studied for several centuries, and from very early on were found to emerge in a fascinatingly regular pattern. This can be illustrated by plotting the latitudinal position at which sunspots occur as a function of time – the resulting butterfly diagram is shown in figure 2. It is apparent that sunspots are confined to low latitudes, and regions of sunspot emergence, which start each cycle at latitudes of around $\pm 30^\circ$, appear to migrate towards the equator over a period of approximately 11 years (this is known as the Schwabe cycle). Sunspots often appear in pairs, and are believed to be the result of a deep-seated band of azimuthal magnetic field becoming buoyantly unstable and piercing the solar surface, generating two regions of opposite polarity where it emerges. The line segment joining the centres of these two regions tends to be slightly tilted with respect to the east-west direction, and the spots are known to follow Hale's law, which states that leading spots of pairs in the northern hemisphere tend to have the same polarity, with the reverse polarity for leading spots in the southern hemisphere. After a time period of approximately 11 years the sunspot field reverses, so that the opposite pattern of polarities is observed in each hemisphere. It is also known that the polar field reverses every 11 years, but is out of phase with the sunspot field and reverses at sunspot maximum. Thus, accounting for the reversal, the solar magnetic field follows a 22 year cycle.

Delving further back in time, time-dependent variations have been observed in the basic 22 year magnetic cycle. Figure 3 illustrates the amplitude modulation of the sunspot number (an arbitrarily defined measure of magnetic activity) over the past 4 centuries. Clearly marked is the absence of sunspots during the late seventeenth century – a period of time known as the Maunder Minimum. Even though there was little surface activity at this time, it is believed that cyclic dynamo action persisted during this period, it was merely that the field strength fluctuated below the amplitude required for the production of active regions. In support of this idea are analyses of ^{10}Be and ^{14}C – these are terrestrial isotopes whose abundance is anti-correlated with solar magnetic activity. It is clearly seen from figure 4 that the cycles in ^{10}Be persisted throughout the Maunder Minimum (shaded area), whilst the ^{14}C data indicates that such periods of low activity are a characteristic feature of solar magnetic activity, with a grand minimum period occurring (on average) about every 200 years.

The Solar Dynamo

It is believed that the large-scale solar magnetic activity is caused by the operation of a dynamo. The idea of a hydromagnetic dynamo is based upon the concept that the motion of an electrically conducting fluid across a magnetic field will induce a current, which (in turn) will generate more magnetic field. This regeneration process works against the continual drain of magnetic energy owing to the resistance of the fluid, and amplification of the total magnetic energy (ie. dynamo action) will occur if the inductive process is more efficient than magnetic diffusion. Since the complex physical processes that occur within the solar dynamo can be described by a set of non-linear partial differential equations, it is possible to attempt to investigate aspects of the global dynamo process by carrying out massive large-scale numerical simulations. However, although this has been successfully carried out for the Earth's dynamo (see, for example, Glatzmaier and Roberts 1995), the vast range of scales and extreme parameter regimes required in the solar context mean that not only is this technique not viable now, but it will be far into the future before we have the required power to model the whole solar dynamo computationally. It should be stated that even if computational resources allowed us do this, it may not be the best way to proceed. With such a complicated turbulent process, we are likely to achieve nothing but confusing non-interpretable results by simply solving equations without any prior understanding of the physics. As such, the majority of the work on the solar dynamo has been geared towards attempting to understand the crucial physical mechanisms involved by investigating simpler, more idealised models.

The dynamo problem is often simplified by neglecting the non-linear feedback of the magnetic field upon the flow (via the Lorentz force) – it is then *simply* a case of finding a prescribed velocity field that is capable of amplifying a seed magnetic field. This is known as the kinematic problem. However, it soon became apparent that even the kinematic problem is far from straightforward. In 1934, Cowling established the first so-called 'antidynamo' theorem, stating that a steady axisymmetric magnetic field cannot be maintained by dynamo action. Many other antidynamo theorems soon followed, which ruled out the possibility of dynamo action in many other simplified cases. For a long time it was feared that there was no hope of making the dynamo work.

A better understanding of the kinematic dynamo problem can be obtained by describing the magnetic field in terms of its poloidal and toroidal components. For an axisymmetric magnetic field, the poloidal component of the magnetic field lies in the meridional plane, whilst the toroidal component is purely azimuthal. Kinematic dynamo action is then possible if we can find a velocity field that is capable of regenerating both the toroidal and the poloidal components of the magnetic field. Historically, the first key ingredient that was identified was the observed differential rotation at the solar surface, which is known to stretch-out an initially poloidal field to produce a toroidal component (assuming that the plasma is highly conducting) – see figure 5(a). This process became known as the ω-effect. Having found this mechanism, the challenge was then to find a physical process that is capable of completing the cycle by regenerating the poloidal field.

The second part of the cycle is much more complex. In a ground breaking paper by Parker (1955), it was suggested that small-scale helical motions (resulting from convection in a rotating body) could twist segments of a toroidal field into loops of field in the meridional plane – see figure 5(b). The net effect of many of these (non-axisymmetric) small-scale events would then give rise to a large-scale meridional field, thereby completing the dynamo cycle. A decade later, a mathematical formulation of this argument, known as *mean-field electrodynamics* (see box), was developed by Steenbeck, Krause and Rädler (1966). This theory more formally describes the way in which small-scale magnetic and velocity fluctuations combine in order to generate a large-scale poloidal magnetic field. This mechanism has subsequently become known as the α-effect.

Having established the physical processes that enable the regeneration of magnetic field, we now need to relate these ideas to the Sun, and determine where the dynamo process is actually taking place. Given that Parker's α-effect relies upon convective motions, and that the Sun is observed to be rotating differentially at the surface, the solar convection zone would appear to be a viable location for the dynamo. Many early theoretical models, where dynamo action was distributed throughout the convection zone, were successful in reproducing several of the main qualitative features of the large-scale solar magnetic field (see, for example, Stix 1976). This type of dynamo model is, however, not without its problems. In particular, it is difficult for a dynamo that operates solely within the convection zone to produce the strong magnetic fields that are found within active regions. Regions of concentrated magnetic flux tend to be less dense than their surroundings, and will therefore rise buoyantly up to the photosphere on a time-scale that is short when compared to the solar cycle period (Parker 1979). It is therefore doubtful that magnetic flux could be held within the convection zone long enough to be amplified to the required field strength.

Turbulent motions will also tend to expel magnetic flux from the convection zone. Like magnetic buoyancy, this will inhibit the operation of any dynamo that is acting solely within this region. However, as a result of these convective motions, magnetic flux will be concentrated into a thin layer in the convectively stable region just below the base of the convection zone (Spiegel and Weiss 1980). This suggests an alternative idea for the solar dynamo, namely that the dynamo may be located in the region around the base of the solar convection zone (Galloway and Weiss 1981). Magnetic flux within this stably stratified region will not be as susceptible to magnetic buoyancy instabilities; therefore, if this is where the bulk of the flux is stored, stronger fields may be able to develop before they become buoyantly unstable. It should also be stressed that a further problem associated with strong magnetic fields is that they are likely to resist deformation by convective upwellings, which will reduce the efficiency of the α-effect and hamper the generation of magnetic field. However, if the bulk of the magnetic flux is stored beneath the convection zone, away from where the α-effect is presumably operating, then this will also reduce this (so-called) α-quenching problem.

The idea that the solar dynamo may be operating around the base of the convection zone has been further reinforced by helioseismological findings. Given the importance of differential rotation to the dynamo process, it is clearly crucial to know the spatial dependence of differential rotation within the solar interior. Early dynamo models were constructed on the basis that the only observational information available was for the solar surface differential rotation. Helioseismology has now provided a great deal of information concerning differential rotation within the Sun (see the article by Thompson, in this issue, for further details – figure 4 of that article shows the inferred differential rotation profile). Many of these helioseismological findings were surprising and of great significance to solar dynamo theory. The layer of pronounced radial shear around the base of the convection zone, commonly referred to as the tachocline (Spiegel and Zahn, 1992), is of particular interest to dynamo theorists. The tachocline acts as a transition region between the (relatively weakly) differentially rotating convection zone and the almost rigidly rotating region below, and appears to be the site of the strongest differential rotation within the Sun.

Following this discovery, and with the aim of circumventing some of the problems associated with the implied presence of a strong magnetic field situated at the base of the convection zone, Parker (1993) formulated a new model known as the *interface dynamo*. In this model, the two generation effects are spatially separated, with the α-effect operating in the turbulently convective layer and the ω-effect in the shear layer below. An essential part of the dynamo's success is the transport of flux between the two regions, the role of which was taken by diffusion. Since turbulence enhances the effects of diffusion, the diffusion within the tachocline region is assumed to be significantly smaller than that found within the convection zone proper. Parker's simple model is not only effective in generating magnetic fields, but also addresses the α-quenching issue by allowing strong toroidal fields to be generated in the tachocline, away from the region where the α-effect is operating (see also Charbonneau and MacGregor 1996).

The interface dynamo model has since grown in popularity, and evolved in a manner that not only makes use of a deeper understanding of the main physical mechanisms (α, ω and diffusion), but also incorporates additional important effects. In particular, it has long been known that the magnetic buoyancy instability is responsible for the transport of toroidal field from the tachocline to the convection zone, and it is now believed that magnetic pumping (the downwards expulsion of flux by turbulent convection) is an effective mechanism for returning poloidal field to the tachocline. Recent numerical simulations of turbulent, penetrative, compressible convection by Tobias *et al.* (2001), in which an unstable region (the convection zone) overlies a stable overshoot region, have examined the fate of an initially horizontal layer of magnetic field inserted into the unstable region. As shown in figure 6, the strong vortical downflows are efficient at wrapping up the magnetic field and dragging it downwards with them as they penetrate the stable layer. Furthermore, the pounding of the overshooting convection tends to confine the flux there, offsetting the effects of magnetic buoyancy. The effects of incorporating the pumping mechanism into a model of the interface dynamo have also recently been investigated, revealing that there can exist a preferred magnitude of pumping for which the dynamo is most efficient (Mason *et al.* 2004, in preparation).

Our increased understanding of the processes operating in the solar dynamo has led us to the scenario illustrated in figure 7. Building upon Parker's original model, it is believed that the toroidal field is generated via the shearing of the poloidal field in the tachocline. The newly generated toroidal field is then susceptible to the magnetic buoyancy instability and rises into the convection zone, where the poloidal field is regenerated via the α-effect. The convection zone acts as a filter, allowing only the strongest field to continue to rise to the surface and appear as active regions. The weaker field is churned up by the convection, and is recycled, being transported back to the tachocline by the turbulent pumping, where the cycle repeats. In current solar dynamo research, each individual process that occurs in the dynamo, and the interaction of these processes, is actively being investigated, generating many unresolved issues and areas of debate (see Ossendrijver 2003 for a thorough review of the subject).

Current areas of debate

The site of strongest differential rotation, and the hence the location for the operation of the ω-effect, is now well established. Dynamo theorists have moved away from the surface as the site for the generation of the toroidal field and, on the basis of the recent helioseismology results, have now pinned down the ω-effect to the tachocline at the base of the convection zone. In many recent dynamo models, differential rotation within the Sun has been represented by choosing, as an imposed velocity field, an analytic fit to the solar differential rotation profile. An example of such an analytic fit, similar to that used by (for example) Dikpati and Charbonneau (1999), is shown in figure 8. Whilst the details of the ω-effect are now relatively well known, many aspects of the α-effect have not yet been agreed upon.

As discussed above, Parker's suggestion was that the poloidal field is regenerated by the twisting action of cyclonic convection upon an initially toroidal field. It is generally accepted that this mechanism will be suppressed as the magnetic field increases in strength; however, how strong the field has to be before this quenching prevents regeneration is less clear. Recent numerical simulations (Cattaneo and Hughes 1996) suggest that the quenching may be severe even for weak fields (with ξ in equation (7) being comparable to R_m – a dimensionless measure of the efficiency of advective effects relative to diffusion, which is approximately $10^8 - 10^{10}$ in the convection zone). Thus, with such extreme quenching, it is difficult to generate fields of the observed strength, and this problem has led to the search for different physical mechanisms that are capable of regenerating poloidal field.

An alternative model that has recently regained popularity is the Babcock-Leighton dynamo (Babcock 1961; Leighton 1969). In these models it is the decay of tilted bipolar active regions that leads to a source of poloidal flux, so that the 'α-effect' only operates at the solar surface. At times when only the surface differential rotation was known these models were very attractive. With the discovery of the tachocline, the models have evolved into so-called *flux transport dynamos*, in which the ω-effect now operates in the tachocline and a meridional circulation is invoked in order to couple the two generation regions. These models are not plagued by the same α-quenching problem previously described – in fact they require a strong field in order to operate – and they are capable of generating results in qualitative agreement with sunspot observations, and may also explain polar reversals (see Dikpati *et al.* 2004 and references therein). However, as the dynamo relies upon the presence of sunspots, it cannot operate in times of grand minima, and we must appeal to another dynamo mechanism to account for the persistence of the solar cycle throughout these periods. The flux-transport models also rely upon a strong steady flow to transport field between the surface and base of the convection zone – although a weak polewards flow is observed at the surface (Hathaway 1996), the details of the sub-surface flow are very uncertain. Indeed, the effects of compressibility will mean that any equatorwards flow in the vicinity of the tachocline will be significantly weaker than the surface flow, casting doubts upon its ability to transport strong fields.

Most notably, due to the large separation of the two generation regions, the models are much less effective than those in which both the α and ω-effects operate within the same region, or near to one another (Mason *et al.* 2002).

Thus, with the toroidal field generated in the tachocline, it is more natural and more efficient to locate the α-effect nearby. Tachocline based 'α-effects' have been the subject of much recent work, and can arise through different mechanisms – one of the most plausible of these is the magnetic buoyancy instability. The effect from magnetic buoyancy is easiest to visualise if the large-scale magnetic field in the neighbourhood of the tachocline is assumed to be a collection of many individual flux tubes. Undular buoyancy instabilities of these flux tubes will create loops of magnetic field which will tend to twist under the influence of the Coriolis force. This macrodynamic process is (in some sense) analogous to the microdynamic convectively-driven α-effect and is therefore assumed to be capable of producing a similar effect (Ferriz-Mas *et al.*, 1994). An alternative (and probably more likely) picture is that the large-scale magnetic field at the base of the convection zone is in the form of a continuous layer. In this case, the magnetic buoyancy instability in the presence of rotation can result in unstable waves which can also give rise to an α-like effect (see, for example, Moffatt, 1978; Thelen 2000).

This buoyantly-driven α-effect differs in several important ways from an α-effect that is driven by turbulent convection. In particular, the magnetic buoyancy instability actually requires relatively strong fields in order to operate, so an α-effect that is driven by this mechanism is not subject to the same quenching problems as the turbulent α-effect. Having said that, there is probably still some quenching for very strong fields, where the magnetic buoyancy instability may be so efficient that flux escapes from the tachocline region before it can contribute to the dynamo process. In most mean-field simulations, α-quenching is usually represented in a simple parameterised way. The relative importance of different non-linear quenching mechanisms remains an open question, so parameterised quenching mechanisms are probably best viewed as a convenient way of forcing the dynamo to saturate in the non-linear regime.

In order to apply the idea of a buoyantly driven α-effect to a mean-field solar dynamo model, we need to determine its region of operation within the Sun. This process is not yet well understood and so we have to rely primarily upon physical intuition. Clearly, it must be concentrated around the base of the convection zone, where the bulk of the magnetic flux is located, however, the latitudinal variation of this kind of α-effect is much harder to predict – although it should be antisymmetric about the equator in order to reflect the equatorial antisymmetry of the Coriolis twisting effect. An approach that is commonly used when modelling the solar dynamo is to fix the radial distribution of the α-effect and then vary the latitudinal dependence until results are obtained that are consistent with observations. In this way, observational details are used to constrain the α-effect.

The natural assumption to make is that the α-effect is strongest at high latitudes (due to its dependence upon the Coriolis force). However, when coupled with the strong negative radial shear there within the tachocline, this tends to lead to oscillatory dynamo action only at high latitudes, where active regions are never observed on the Sun. One way to resolve this problem is if the α-effect is somehow suppressed in this region. Very recently it has emerged that a strong radial shear may inhibit non-axisymmetric buoyancy instabilities in a magnetic layer (Tobias and Hughes, 2004). If the α-effect is driven by magnetic buoyancy, this then provides possible justification for prescribing an α-effect, in a mean-field model, that is confined to lower latitudes (where the radial shear is weaker).

By restricting the α-effect to low latitudes in mean-field dynamo simulations, it is possible to produce results that are in greater agreement with observations. Figure 9 shows contours of the azimuthal component of the dynamo-generated field at the base of the convection zone for a model of this form, in which the sole non-linearity is due to a parameterised α-quenching mechanism. For this solution, dynamo activity is restricted to low latitudes. Since the radial shear at low latitudes is positive, the equatorwards propagation of magnetic activity is a consequence of the fact that α has been chosen to be negative in the Northern hemisphere (Parker, 1955). This solution qualitatively matches the pattern of sunspot activity shown in figure 2. Like the sunspot magnetic field, the azimuthal field shown in figure 9 is antisymmetric about the equator. Given that the meridional field is symmetric about the equator, this is consistent with the dynamo having dipolar symmetry, which is actually what is observed on the Sun (although it should be noted that the symmetry of this simulated solution may depend upon the choice of non-linearity). It is therefore possible to closely match the main qualitative features of the solar magnetic cycle with a mean-field model, provided that the α-effect is confined to low latitudes at the base of the convection zone.

Periods like the Maunder minimum, and other time-dependent features, suggest that there is more to the solar dynamo than the basic solar cycle. If this modulation is deterministic, then it should be possible to adapt mean-field models so as to produce time-dependent behaviour that resembles the long term solar magnetic activity. One way to do this is by considering a mean-field model which includes the back-reaction of the azimuthal component of the Lorentz force upon the differential rotation profile (Malkus and Proctor, 1975). This involves solving an evolution equation for the velocity perturbation; however, this dynamical non-linearity is probably more physical than an arbitrarily parameterised α-quenching mechanism. By introducing a separation in scales between the magnetic diffusion time and the viscous dissipation time for the fluid, it is possible to produce time-dependent behaviour in such models (Tobias 1997; Moss and Brooke 2000). Figure 10 shows a dynamo solution from a simplified Cartesian model, which exhibits strong amplitude modulation with pronounced "Grand Minimum" phases.

The induced angular velocity perturbation is an interesting aspect of these models. Dynamical variations in the solar differential rotation profile were first detected as a surface pattern of alternating bands of faster and slower than average local rotation, which migrate from mid to low latitudes with an 11 year periodicity (Howard and LaBonte 1980). This 11 year cycle is strikingly similar to the sunspot cycle and is consistent with the idea that these (so-called) torsional oscillations may be magnetically driven. Figure 11 illustrates the oscillatory part of the total velocity perturbation, taken from a mean-field model which incorporates the non-linear feedback of the magnetic field upon the flow. This oscillatory pattern bears a clear resemblance to the observed torsional oscillations, at least at low latitudes, on the surface of the Sun (the equatorwards migration of the low-latitude branch of the torsional oscillations can be seen in figure 5 in the article by Thompson, this issue).

Torsional oscillations raise several other interesting questions regarding the solar dynamo. In the models described above, all the dynamo action and torsional oscillations are primarily confined to low latitudes around the base of the convection zone. Recent observations not only suggest that the strongest oscillations occur at the surface, but also that there may be an additional band of torsional oscillations at high latitudes (Vorontsov 2002; Thompson, this issue). A possible explanation for one of these discrepancies lies in the fact that the solar convection zone is highly stratified (Covas *et al.* 2004). The very low fluid density at the photosphere will mean that even a relatively small perturbation to the local angular momentum at the surface would be able to produce a large angular velocity perturbation. The existence of a high-latitude branch to the torsional oscillations presents a more interesting problem, since it is difficult to see how such torsional oscillations could be magnetically driven without the presence of strong magnetic fields at high latitudes, where active region are never observed. If, however, there is dynamo action at high latitudes, it is possible to produce a high-latitude branch to the torsional oscillations that is consistent with observations (Bushby 2004, in preparation). This is an interesting open question.

The future

Since direct numerical simulations of the solar dynamo are currently not feasible, much of the recent progress in the subject has relied upon mean-field dynamo theory. Using the mean-field approach, it is possible to reproduce many of the observed features of the solar dynamo in relatively simple numerical models. By imposing a solar-like rotation law, and then choosing an appropriate spatial dependence for the α-effect, it is possible to find model solutions that are dipolar, confined to low latitudes and migrate equatorwards during each cycle. Mean-field theory is therefore capable of reproducing observations provided that the parameters of the model are chosen accordingly.

Another successful aspect of mean-field modelling is that it has highlighted several key areas for future research. The tachocline is obviously of great importance to the solar dynamo, but many theoretical issues, particularly surrounding the formation and stability of the tachocline, remain poorly understood. There are also issues that still need to be resolved concerning the α-effect. In particular, there is a need to determine the relative importance of the various physical mechanisms that may be responsible for regenerating poloidal magnetic field.

It seems likely that an α-effect that is driven by cyclonic convection may be quenched by relatively weak magnetic fields. Unless we can overcome this issue, a tachocline-based α-effect, driven by the magnetic buoyancy instability (which is not subject to the same quenching problems), seems to be the most plausible alternative mechanism. Having said that, when applying this idea in a mean-field model, a (buoyancy-driven) tachocline-based α-effect must be confined to low latitudes in order to reproduce solar-like behaviour. At present, there is little justification for the suppression at high latitudes, and hence further work is needed in order to fully assess the interaction of the magnetic buoyancy instability with the strong differential rotation that is found in the tachocline. Finally, it is still unclear which of the various possible non-linear quenching mechanisms is actually of primary importance to the solar dynamo. Most basic models rely upon simple parameterised mechanisms, which give an instantaneous quenching effect. The existence of torsional oscillations is evidence of the fact that dynamical non-linearities are likely to be very important.

Mean-field theory has enabled significant progress to be made in the theoretical understanding of the solar dynamo, although it is only possible to use this theory to produce a qualitative picture of the dynamo process. Until the physical issues that are described above are better understood, it seems likely that it is not going to be possible to make significant further progress using this approach. It is very important to understand the limitations of mean-field models. Certain assumptions that are required for the mathematical formulation of the theory, including the separation of spatial scales, are not wholly justifiable in the solar context. Mean-field models have been constructed to give an insight into the dynamo mechanism by parameterising the essential physical processes. It is therefore not possible to make real quantitative predictions using a mean-field model. It should be mentioned however that attempts are being made to predict solar magnetic behaviour using a flux-transport dynamo (see Dikpati *et al.* 2004; Byrne 2004), in which the key ingredient is believed to be a meridional flow, perceived to be responsible for transporting sunspots towards the equator in the 11 year cycle, and also accounting for polar reversals. The ability of the model to reproduce these solar cycle features has led to the suggestion that by using available observational data for the meridional flow, future solar cycle behaviour may be predicted. It is our belief however that this type of calculation is rather ambitious. In particular, not only is the structure of the required meridional flow somewhat speculative, but there are many physical aspects of the dynamo problem that we only understand in a qualitative sense. Until solar dynamo theory has evolved beyond this stage, quantitative predictions cannot yet be made with any degree of confidence.

Highly illustrative models can be used to describe very simple physical problems or even very complicated aspects of dynamical behaviour. Simplified models of this kind have already enhanced our understanding of the solar dynamo, and there is every reason to suppose that this approach will continue to be a useful one. Although it is not yet possible to investigate the large-scale solar dynamo through direct numerical simulation, it is possible to simulate directly more localised processes. An example of this is the simulation of small-scale dynamo action by turbulent convection in the solar photosphere (see, for example, Cattaneo and Hughes 2001 and references therein). This is an interesting problem, many aspects of which have yet to be explored.

Finally, it is worth considering to what extent it is possible to extend these ideas to other stars. It may seem likely that dynamos in other solar-type stars can be described in a very similar way; however, it should be stressed that much of this theory has become very specific to the Sun. Dynamos in fully convective stars or very rapidly rotating late-type stars will probably be very different in character from the solar dynamo. Although plausible assumptions can be made in order to construct dynamo models for such stars (see, for example, Bushby 2003), such models are inevitably somewhat speculative.

It should be possible to make significant advances in solar dynamo theory over the next few years. Through increased computational power, higher resolution observations, and progress in our theoretical understanding of the dynamo mechanisms, we are likely to move nearer to our goal of understanding how magnetic fields behave in the turbulent Sun. At present the solar dynamo is far from being a solved problem, and is unlikely to be so for the foreseeable future.

Acknowledgements: We would like to thank David Hughes, Michael Proctor, Steven Tobias and Nigel Weiss for many useful discussions.
Paul Bushby is a postdoctoral research associate (supported by PPARC) at DAMTP, University of Cambridge; Joanne Mason is a postdoctoral research fellow at the High Altitude Observatory/NCAR, Boulder, CO.

References

Babcock H W 1961 *ApJ* **133** 572–587.
Beer J, Tobias S M and Weiss N O 1998 *Sol. Phys.* **181** 237–249.
Beer J 2000 *Space Sci. Rev.* **94** 53–66.
Bushby P J 2003 *MNRAS* **342** L15–L19.
Byrne G 2004 *New Scientist* **181** 2437.38-41.
Cattaneo F and Hughes D W 1996 *Phys. Rev. E* **54** 4532–4535.
Cattaneo F and Hughes D W 2001 *A&G* **42** 3.18–3.22.
Charbonneau P and MacGregor K B 1996 *ApJ* **473** L59–L62.
Covas E, Moss D and Tavakol R 2004 *A&A* **416** 775–782.
Dikpati M and Charbonneau P 1999 *ApJ* **518** 508–520.
Dikpati M, De Toma G, Gilman P A, Arge C N and White O R 2004 *ApJ* **601** 1136–1151.
Ferriz-Mas A, Schmitt D and Schüssler M 1994 *A&A* **289** 949–956.
Galloway D J and Weiss N O 1981 *ApJ* **243** 945–953.
Glatzmaier G A and Roberts P H 1995 *Nature* **377** 203–209.
Hathaway D H 1996 *ApJ* **460** 1027–1033.
Howard R and LaBonte B J 1980 *ApJ* **239** L33–L36.
Leighton R B 1969 *ApJ* **156** 1–26.
Malkus W V R and Proctor M R E 1975 *J. Fluid. Mech.* **67** 417–444

Mason J, Hughes D W and Tobias S M 2002 *ApJ* **580** L89–L92.
Moffatt H K 1978 *Magnetic Field Generation in Electrically Conducting Fluids*, CUP, Cambridge.
Moss D and Brooke J 2000 *MNRAS* **315** 521–533.
Ossendrijver M 2003 *Astron. Astrophys. Rev.* **11** 287–367.
Parker E N 1955 *ApJ* **122** 293–314.
Parker E N 1979 *Cosmical Magnetic Fields*, Clarendon Press, Oxford.
Parker E N 1993 *ApJ* **408** 707–719.
Spiegel E A and Weiss N O 1980 *Nature* **287** 616–617.
Spiegel E A and Zahn J-P 1992 *A&A* **265** 106–114.
Steenbeck M, Krause, F and Rädler K-H 1966 *Zeitscr. Naturforsch.* **21a** 369–376. (English Translation by Roberts P H and Stix M, 1971 *The turbulent dynamo* NCAR Technical Note TN/IA-60)
Stix M 1976 *A&A* **47** 243–254.
Thelen J-C 2000 *MNRAS* **315** 155–164.
Tobias S M 1997 *A&A* **322** 1007–1017.
Tobias S M, Brummell N H, Clune T L and Toomre J 2001 *ApJ* **549** 1183–1203.
Tobias S M Hughes D W 2004 *ApJ* **603** 785–802.
Vorontsov S V, Christensen-Dalsgaard J, Schou J, Strakhov V N and Thompson M J 2002 *Science* **296** 101–103.

Mean-field electrodynamics

Building upon the pioneering ideas of Parker (1955), Steenbeck, Krause and Rädler (1966) formulated a mathematical theory to describe the way in which small-scale effects play a crucial role in the generation of large-scale magnetic fields (see also Moffatt 1978). The starting point is the induction equation,

$$\frac{\partial \mathbf{B}}{\partial t} = \nabla \times (\mathbf{U} \times \mathbf{B}) + \eta \nabla^2 \mathbf{B}, \tag{1}$$

that governs the evolution of a magnetic field $\mathbf{B}$ within an electrically conducting plasma of magnetic diffusivity η, moving with a velocity $\mathbf{U}$. Progress is made by assuming that we can decompose the magnetic and velocity fields into a mean part ($\mathbf{B_0}, \mathbf{U_0}$ say) varying on a length-scale L, and a randomly fluctuating part ($\mathbf{b}, \mathbf{u}$ say) varying on a much smaller length-scale, $l \ll L$. Defining averages, denoted by $\langle . \rangle$, over an intermediate length-scale, we have

$$\mathbf{B} = \mathbf{B_0} + \mathbf{b}, \quad \mathbf{U} = \mathbf{U_0} + \mathbf{u}, \tag{2}$$

where $\langle \mathbf{b} \rangle = \langle \mathbf{u} \rangle = \mathbf{0}$. Substituting this decomposition into the induction equation (1) and averaging, we obtain an equation for the mean field

$$\frac{\partial \mathbf{B_0}}{\partial t} = \nabla \times (\mathbf{U_0} \times \mathbf{B_0}) + \nabla \times \boldsymbol{\epsilon} + \eta \nabla^2 \mathbf{B_0}, \tag{3}$$

where $\boldsymbol{\epsilon} = \langle \mathbf{u} \times \mathbf{b} \rangle$ is the all-important new quantity, and is known as the mean electromotive force. In order to proceed we must be able to express $\boldsymbol{\epsilon}$ in terms of the mean quantities alone. With this in mind we note that the equation for the fluctuating field, $\mathbf{b}$, may be obtained by subtracting (3) from (1), yielding a linear relationship between $\mathbf{B_0}$ and $\mathbf{b}$. Hence it is reasonable to assume an expression for $\boldsymbol{\epsilon}$ of the form

$$\epsilon_i = \alpha_{ij} B_{0_j} + \beta_{ijk} \frac{\partial B_{0_j}}{\partial x_k} + \cdots \tag{4}$$

where the tensors α_{ij} and β_{ijk} depend upon the fluctuating velocity and the diffusivity, and all higher order derivatives in (4) are negligible under the assumption that L is sufficiently large. In order to understand the effects of the electromotive term we consider the simplest case of homogeneous isotropic turbulence, so that α_{ij} and β_{ijk} are isotropic tensors and can be written as

$$\alpha_{ij} = \alpha \delta_{ij}, \quad \beta_{ijk} = \beta \epsilon_{ijk}. \tag{5}$$

In this case, equation (3) becomes

$$\frac{\partial \mathbf{B_0}}{\partial t} = \nabla \times (\mathbf{U_0} \times \mathbf{B_0}) + \nabla \times (\alpha \mathbf{B_0}) + (\eta + \beta) \nabla^2 \mathbf{B_0}. \tag{6}$$

Thus we see that β represents a turbulent enhancement of the magnetic diffusivity, whilst α represents the ability of the underlying turbulence to act as a source for the mean field. In order for the α-effect to be non-zero, it follows that the turbulent velocity field must lack reflectional symmetry – a notion that is often associated with fluid motion within rotating bodies.

The idea of the kinematic dynamo problem is to solve equation (6) for $\mathbf{B_0}$, given a prescribed velocity $\mathbf{U_0}$ and a chosen diffusivity and α-effect. Any solutions of this linear problem are either exponentially growing or exponentially decaying. In the (more complicated) dynamic problem, in which we would solve equation (6) together with an equation for the fluid flow, the feedback effects of the magnetic field on the fluid would limit the amplitude of the generated field. An attempt at including such effects into the kinematic model has been made by making the dynamo coefficients dependent on the magnetic field strength. Algebraic, parameterised expressions are often used as an ad hoc representation of the tendency of strong magnetic fields to resist the α-effect:

$$\alpha \approx \frac{\alpha_0}{1 + \xi B^2}, \tag{7}$$

where α_0 is the value in the absence of a magnetic field, and ξ is a parameter that determines the strength of the quenching (see the discussion in the main text). Similar expressions are often used to represent the quenching of β by strong magnetic fields.

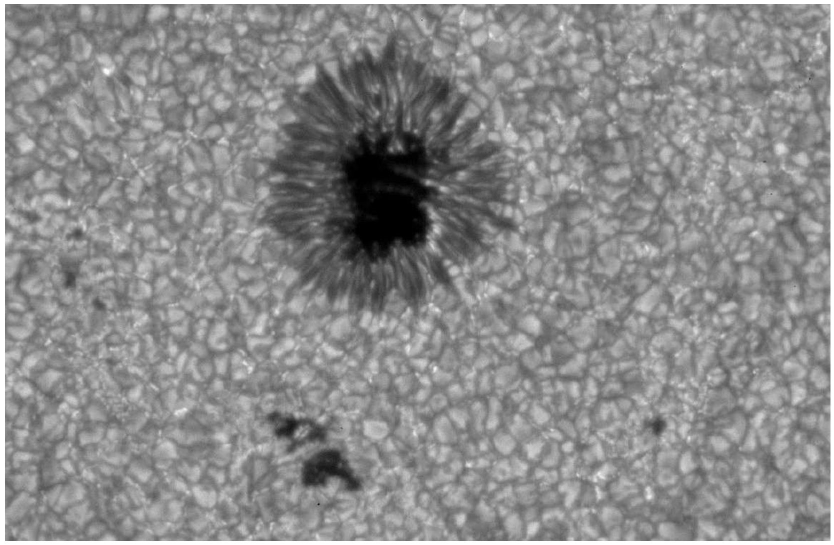

Figure 1: A sunspot, with the dark central umbra and outer filamentary penumbra, and the surrounding granulation pattern. Sunspots are sites of intense magnetic field, and are one of the most important indicators of solar magnetic activity. (Courtesy of P. Charbonneau and O. R. White, HAO/NCAR. Source: Kiepenheuer/Uppsala/Lockheed (P.Brandt, G. Simon, G. Scharmer, D. Shine).)

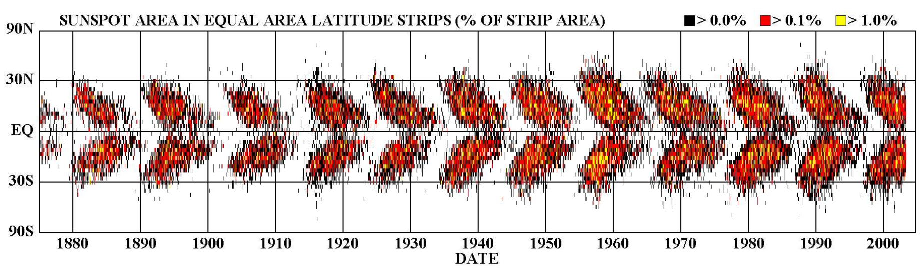

Figure 2: The Butterfly diagram: a latitude versus time plot of the sites at which sunspots appear. At the beginning of each cycle, sunspots emerge at mid latitudes. As the cycle progresses, the locations at which emerging spots appear move towards the equator. After a period of about 11 years the magnetic field reverses and the sunspots reappear at higher latitudes. (Courtesy of D. H. Hathaway, NASA/NSSTC.)

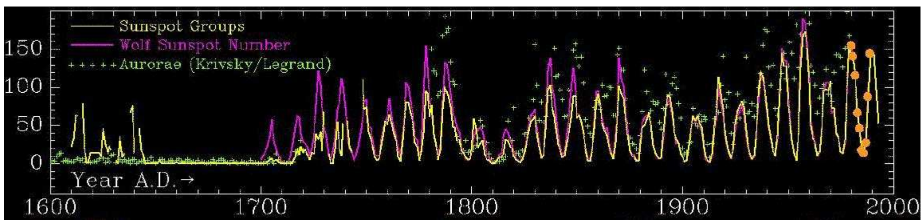

Figure 3: The amplitude modulation of the basic 11 year sunspot cycle. Notice in particular the absence of surface magnetic activity during the late seventeenth century, a period of time known as the Maunder Minimum. (Courtesy of P. Charbonneau and O. R. White, HAO/NCAR. Source: NOAA+Zürich+RDC (D.V.Hoyt)+CNRS/INSU (J.-P. Legrand)+Ondrejov Obs. (K. Krivsky).)

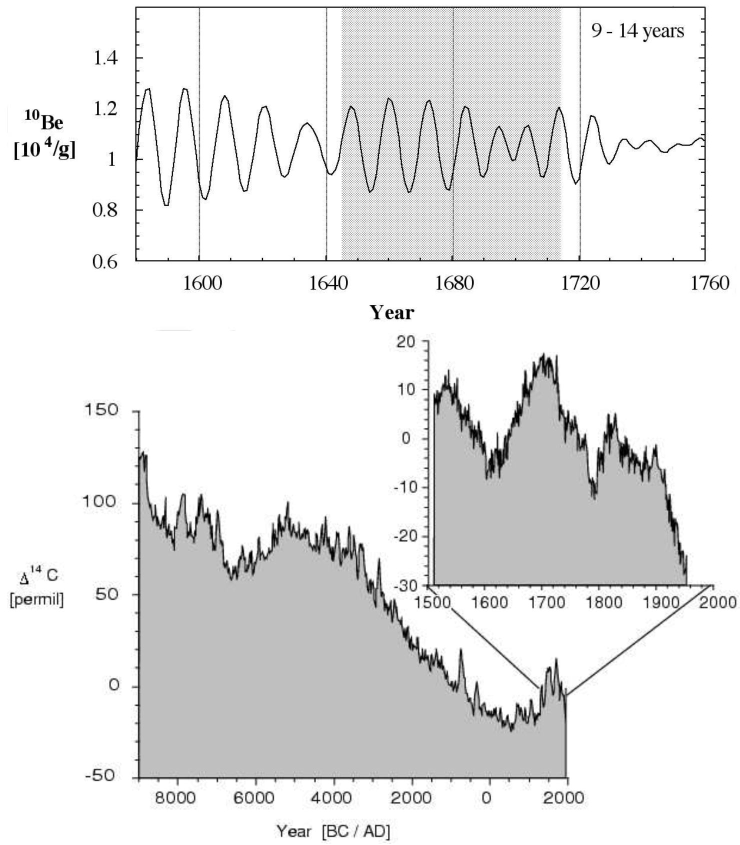

Figure 4: Solar activity is known to be anti-correlated with the abundance of the terrestrial isotopes ^{10}Be and ^{14}C. The persistence of the solar cycle throughout Maunder Minimum is clearly shown in the ^{10}Be data (the shaded area of the top figure). The local maximum in the ^{14}C data (bottom) at around 1700 corresponds to the Maunder Minimum. The appearance of many other local maxima indicates that the Maunder Minimum is not an isolated event. (From Beer *et al.* 1998 and Beer 2000.)

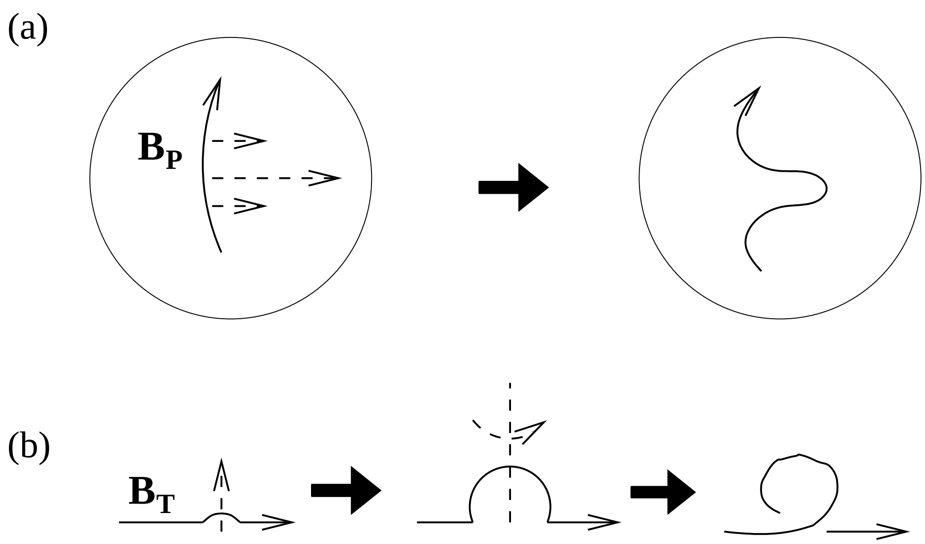

Figure 5: The two regeneration mechanisms: (a) Differential rotation stretches out an initially poloidal field ($\mathbf{B_p}$) generating a toroidal component (the ω-effect). (b) The 'rise and twist' motion of cyclonic convection generates small-scale meridional loops from an initially toroidal field ($\mathbf{B_T}$) (the α-effect).

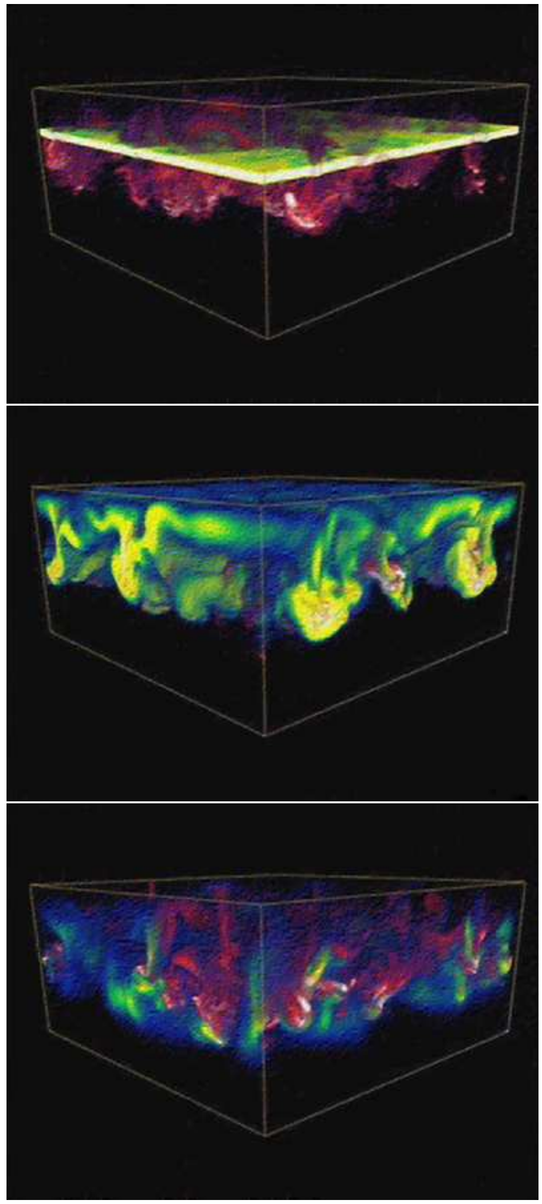

Figure 6: Volume renderings of enstrophy density (purple-white) and of magnetic energy density (yellow-green-blue). Top: Initial configuration with a layer of magnetic field inserted into the unstable region. Middle: Later time, showing pumping of flux by the downflows. Bottom: Final state, in which majority of the magnetic flux has been pumped into the stable region, and is held there by the overshooting convection. (Courtesy of S. M. Tobias.)

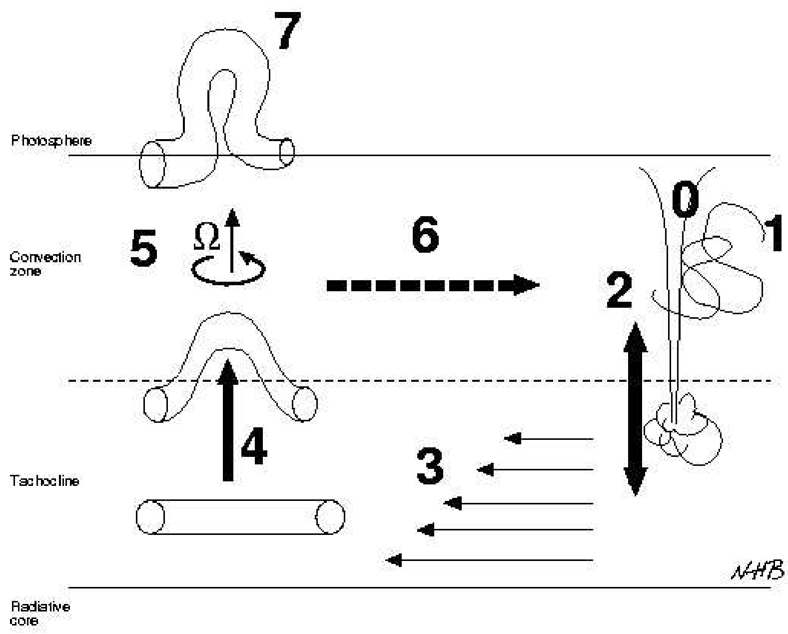

Figure 7: Schematic diagram of the elements of the solar dynamo. (0) Convection dominated by strong down-flowing plumes. (1) Generation/shredding of magnetic field. (2) Transport of magnetic field from the convection zone into the tachocline. (3) Omega effect: Conversion of poloidal field into toroidal field. (4) Formation of structures and magnetic buoyant rise. (5) Dynamic alpha effect: Regeneration of poloidal field. (6) Small-scale alpha effect: Recycling of field. (7) Emergence of structures. (Courtesy of N. H. Brummell.)

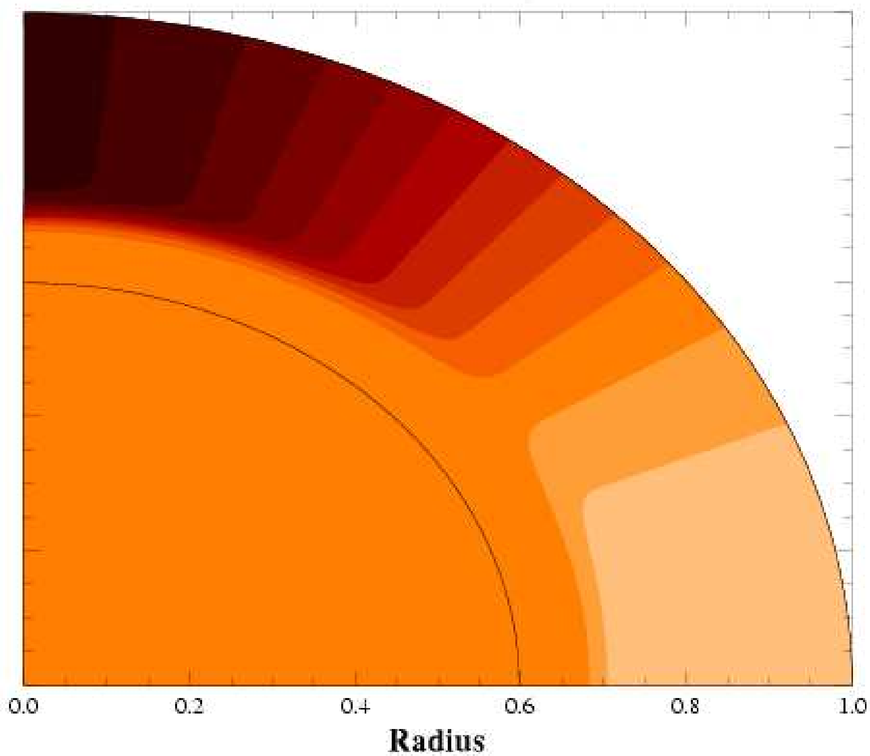

Figure 8: Contours of constant angular velocity for an analytic fit to the solar differential rotation profile. Although only the northern hemisphere is shown here, this profile is assumed to be symmetric about the equator. Darker colours indicate slower rotation rates.

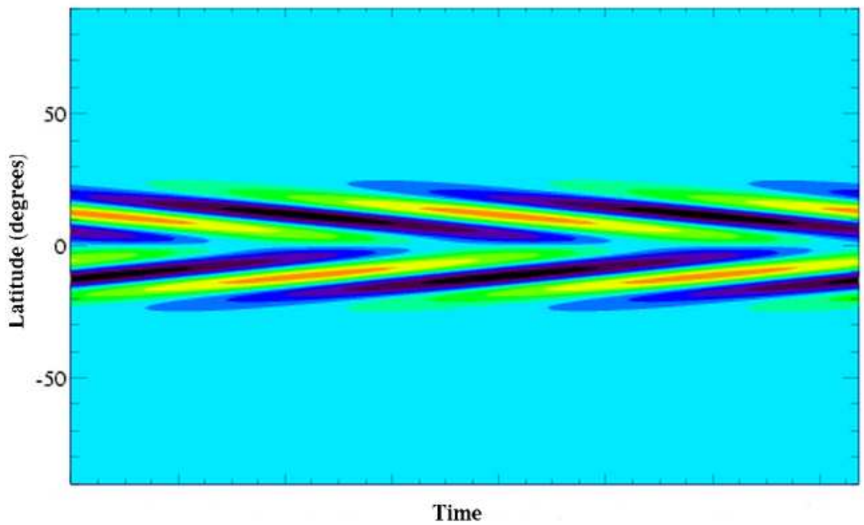

Figure 9: Azimuthal magnetic field, at the base of the convection zone, as a function of latitude and time. Like the observed large-scale field on the Sun, activity is dipolar, confined to low latitudes and migrates equatorwards.

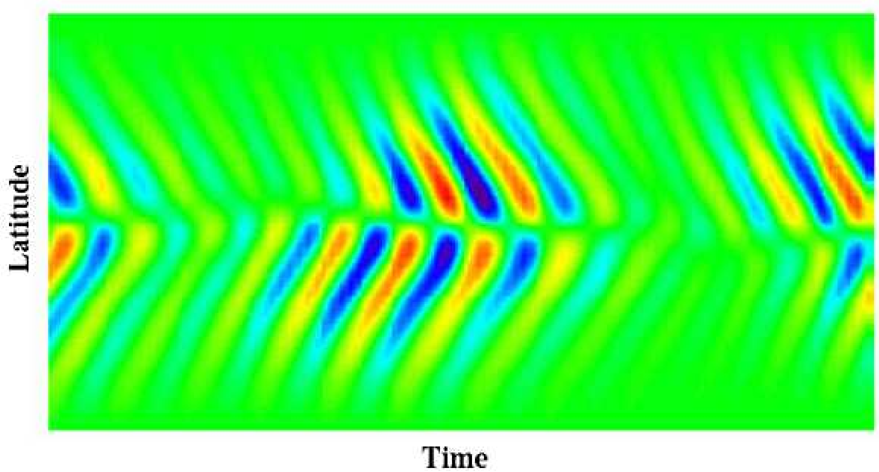

Figure 10: Contours of azimuthal field, from a simplified Cartesian model, as a function of latitude and time. Note the strong variation in the amplitude of the magnetic activity, with pronounced grand minima-like phases. (After Beer *et al.* 1998.)

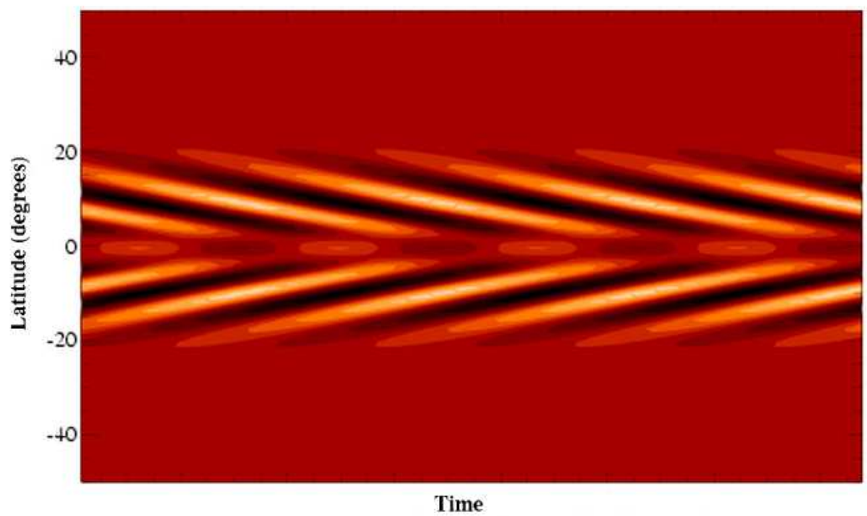

Figure 11: Torsional oscillations, at the base of the convection zone, as a function of latitude and time. Like the associated magnetic field, these oscillations are confined to low latitudes and migrate equatorwards.