The solar dynamo

Arnab Rai Choudhuri
Department of Physics, Indian Institute of Science, Bangalore 560 012, India

It is believed that the magnetic field of the Sun is produced by the dynamo process, which involves nonlinear interactions between the solar plasma and the magnetic field. After summarizing the main characteristics of solar magnetic fields, the basic ideas of dynamo theory are presented. Then an appraisal is made of the current status of solar dynamo theory.

1. Introduction

In elementary textbooks on stellar structure, a star is usually modelled as a spherically symmetric, non-rotating, non-magnetic object. It is mainly the magnetic field which makes our Sun much more intriguing than such a textbook star. Several other reviews in this special section should convince the reader of this. It comes, therefore, as no surprise that one of the central problems in solar physics is to understand the origin of the Sun’s magnetic field. The solar dynamo theory attempts to address this problem. The basic idea of this theory is that the solar magnetic fields are generated and maintained by complicated nonlinear interactions between the solar plasma and magnetic fields.

As we shall see in this review, there are still many difficulties with this theory and we are still far from having a completely satisfactory explanation of why the Sun’s magnetic field behaves the way it does. However, no alternate theory of the origin of solar magnetism has so far been able to explain even a fraction of what dynamo theory has explained. Some of us, therefore, are still struggling to put the solar dynamo theory on firmer footing, with the fond hope that we are probably approximately on the correct path.

The aim of this special section is to make the readers of Current Science aware of the present status of solar physics. The solar dynamo theory is a fairly technical subject. It is next to impossible to write a review that will provide a comprehensive introduction to this subject for an average reader of Current Science and, at the time, survey the research frontiers. Still a partial attempt is made here at this next-to-impossible task of presenting the subject in a way which should be understandable – if not to a general reader of Current Science – at least to a reader with some familiarity in physics and fluid mechanics. It is left to the readers to judge if the author has failed completely or only moderately. Needless to say, no attempt is made at a complete coverage of the fundamentals. After summarizing the relevant observations in the following section, we write just enough about the basics in the next two sections to give a rough idea of what is going on. Then in the last three sections we discuss some of the important issues from current research frontiers.

Dynamo theory is based on the principles of magnetohydrodynamics (MHD), in which hydrodynamics equations are combined with Maxwell’s electrodynamics equations. Comprehensive introductions to MHD can be found in the books by Alfvén and Fälthammar, Cowling, Parker, Priest, and Choudhuri. Some books devoted exclusively to dynamo theory are by Moffatt, Krause and Rädler, and Zeldovich et al. We also refer to the review articles on the solar dynamo by Ruzmaikin, Gilman, Hoyng, Brandenburg and Tuominen, and Schmitt.

2. Relevant observations

In 1908 Hale discovered the first evidence of Zeeman effect in sunspot spectra and made the momentous announcement that sunspots are regions of strong magnetic fields. This is the first time that somebody found conclusive evidence of large-scale magnetic fields outside the Earth’s environment. The typical magnetic field of a large sunspot is about 3000 G.

Even before it was realized that sunspots are seats of solar magnetism, several persons have been studying the occurrences of sunspots. Schwabe noted that the number of sunspots seen on the solar surface increases and decreases with a period of about 11 years. Now we believe that the Sun has a cycle with twice that period, i.e. 22 years. Since the Sun’s magnetic field changes its direction after 11 years, it takes 22 years for the magnetic field to come back to its initial configuration. Carrington found that sunspots seemed to appear at lower and lower latitudes with the progress of the solar cycle. In other words, most of the sunspots in the early phase of a solar cycle are seen between 30° and 40°. As the cycle advances, new sunspots are found at increasingly lower latitudes. Then a fresh half-cycle begins with sunspots appearing again at high latitudes. Individual sunspots live from a few days to a few weeks.

After finding magnetic fields in sunspots, Hale and his coworkers made another significant discovery. They found that often two large sunspots are seen side by side and they invariably have opposite polarities. The line joining the centres of such a bipolar sunspot pair is usually...
nearly parallel to the solar equator. Hale’s coworker Joy, however, noted that there is a systematic tilt of this line with respect to the equator and that this tilt increases with latitude. This result is usually known as Joy’s Law. It was also noted that the sunspot pairs have opposite polarities in the two hemispheres. In other words, if the left sunspot in the northern hemisphere has negative polarity, then the left sunspot in the southern hemisphere has positive polarity. This is clearly seen in Figure 1, which is a magnetic map of the Sun’s disk obtained with a magnetogram. The regions of positive and negative polarities are shown in white and black respectively. The polarities of the bipolar sunspots in any hemisphere get reversed from one half-cycle of 11 years to the next half-cycle.

After the development of the magnetograph by Babcock and Babcock, it became possible to study the much weaker magnetic field near the poles of the Sun. This magnetic field is of the order of 10 G and reverses its direction at the time of solar maximum (i.e. when the number of sunspots seen on the solar surface is maximum). This shows that this weak, diffuse field of the Sun is in some way coupled to the much stronger magnetic field of the sunspots and is a part of the same solar cycle. Low-resolution magnetograms show the evidence of weak magnetic field even in lower latitudes. The true nature of this field is not very clear. It was found that the magnetic field on the solar surface outside sunspots often exists in the form of fibril flux tubes of diameter of the order of 300 km with field strength of about 2000 G (large sunspots have sizes larger than 10,000 km). One is not completely sure if the field found in the low-resolution magnetograms is truly a diffuse field or a smearing out of the contributions made by fibril flux tubes. Keeping this caveat in mind, we should refer to the field outside sunspots as seen in magnetograms as the ‘diffuse’ field. It was found that there were large unipolar matches of this diffuse field on the solar surface which migrated towards the pole. Even when averaged over longitude, one finds predominantly one polarity in a belt of latitude which drifts polewards. The reversal of solar field presumably takes place when sufficient field of opposite polarity has been brought near the poles.

Figure 2 (taken from Dikpati and Choudhuri) shows the distribution of both sunspots and the weak, diffuse field in a plot of latitude vs. time. The colour shades indicate values of longitude-averaged diffuse field, whereas the latitudes where sunspots were seen at a particular time are marked by vertical black lines. The sunspot distribution in a time-latitude plot is often referred to as a butterfly diagram, since the pattern (the vertical black lines in Figure 2) reminds one of a butterfly. Such butterfly diagrams were first plotted by Maunder. Historically, most of the dynamo models concentrated on explaining the distribution of
sunspots and ignored the diffuse field. Only during the last few years, it has been realized that the diffuse fields give us important clues about the dynamo process and they should be included in a full self-consistent theory. The aim of such a theory should be to explain diagrams like Figure 2 (i.e. not just the butterfly diagram).

We have provided above a summary of the various regular features in the Sun’s activity cycle. One finds lots of irregularities and fluctuations superposed on the underlying regular behaviour, as can be seen in Figure 2. These irregularities are more clearly visible in Figure 3, where the number of sunspots seen on the solar surface is plotted against time. Galileo was one of the first persons in Europe to study sunspots at the beginning of the 17th century. After Galileo’s work, sunspots were almost not seen for nearly a century. It may be noted that all the observations discussed above pertain to the Sun’s surface. We have no direct information about the magnetic field underneath the Sun’s surface. The new science of helioseismology, however, has provided us lots of information about the velocity field underneath the solar surface. For an account of this subject, the readers may turn to the reviews by Chitre and Antia, and by Christensen-Dalsgaard and Thompson. We shall have occasions to refer to some of the helioseismic findings in our discussion later.

3. Some basic magnetohydrodynamics considerations

The velocity field $\mathbf{v}$ and the magnetic field $\mathbf{B}$ in a plasma (regarded as a continuum) interact with each other according to the following MHD equations:

$$\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla) \mathbf{v} = \frac{1}{\rho} \nabla \left( p + \frac{B^2}{8\pi} \right) + \frac{\nabla \times \mathbf{B}}{4\pi} + \mathbf{g} + \nu \nabla^2 \mathbf{v},$$

(1)

$$\frac{\partial \mathbf{B}}{\partial t} = \nabla \times (\mathbf{v} \times \mathbf{B}) + \lambda \nabla^2 \mathbf{B},$$

(2)

Here $\rho$ is density, $p$ is pressure, $\mathbf{g}$ is gravitational field, $\nu$ is kinematic viscosity, and

$$\lambda = \frac{c^2}{4\pi \sigma}$$

(3)

is magnetic diffusivity ($\sigma$ is electrical conductivity). Equation (1) is essentially the Navier–Stokes equation, to which magnetic forces have been added. It is clear from eq. (1) that the magnetic field has two effects: (i) it gives rise to an additional pressure $B^2/8\pi$; and (ii) the other magnetic term $(\mathbf{B} \cdot \nabla)\mathbf{B}/4\pi$ is of the nature of a tension along magnetic field lines.

Equation (2) is known as the induction equation and is the key equation in MHD. It has the same form as the vorticity equation in ordinary hydrodynamics (see, for example, § 4.2 and § 5.2 of Choudhuri). If $V$, $B$ and $L$ are the typical values of velocity, magnetic field and length scale, then the two terms on the RHS of eq. (2) are of order $VB/L$ and $\lambda B^2/L^2$. The ratio of these two terms is a dimensionless number, known as the magnetic Reynolds number, given by

$$R_m = \frac{VB/L}{\lambda B^2/L^2} = \frac{VL}{\lambda},$$

(4)

Since $R_m$ goes as $L$, it is expected to be much larger in astrophysical situations than it is in the laboratory. In fact, usually one finds that $R_m \gg 1$ in astrophysical systems and $R_m \ll 1$ in laboratory-size objects. Hence the behaviours of magnetic fields are very different in laboratory plasmas and astrophysical plasmas. For example, it is not possible to have laboratory analogues of the self-sustaining magnetic fields of the Earth or the Sun. If $R_m \gg 1$ in an astrophysical system, then the diffusion term in eq. (2) is negligible compared to the term preceding it. In ordinary hydrodynamics, when the viscous dissipation term in the vorticity equation is neglected, we are led to the famous

Figure 3. The number of sunspots seen in a year plotted against the year for the period 1610–1975. The original figure is due to John A. Eddy. Reproduced from Moffat.
Kelvin’s theorem of vorticity conservation (see, for example, § 4.6 of Choudhuri[13]). Exactly similarly, when the diffusion term in eq. (2) is neglected, it can be shown that the magnetic field is frozen in the plasma and moves with it. This result was first recognized by Alfvén[32] and is often referred to as Alfvén’s Theorem of Flux-Freezing.

It is known that the Sun does not rotate like a solid body. The angular velocity at the equator is about 20% faster than that at the poles. Because of the flux freezing, this differential rotation would stretch out any magnetic field line in the toroidal direction (i.e. the $\phi$ direction with respect to the Sun’s rotation axis). This is indicated in Figure 4. We, therefore, expect that the magnetic field inside the Sun may be predominantly in the toroidal direction.

We have already mentioned in § 2 that energy is transported by convection in the layers underneath the Sun’s surface. To understand why the magnetic field remains concentrated in structures like sunspots instead of spreading out more evenly, we need to study the interaction of the magnetic field with the convection in the plasma. This subject is known as magnetoconvection. The linear theory of convection in the presence of a vertical magnetic field was studied by Chandrasekhar[14]. The nonlinear evolution of the system, however, can only be found from numerical simulations pioneered by Weiss[33]. It was found that space gets separated into two kinds of regions. In certain regions, magnetic field is excluded and vigorous convection takes place. In other regions, magnetic field gets concentrated and the tension of magnetic field lines suppresses convection in those regions. Sunspots are presumably such regions where magnetic field is piled up by surrounding convection. Since heat transport is inhibited there due to the suppression of convection, sunspots look darker than the surrounding regions.

Although we have no direct information about the state of the magnetic field under the Sun’s surface, it is expected that the interactions with convection would keep the magnetic field concentrated in bundles of field lines throughout the solar convection zone. Such a concentrated bundle of magnetic field lines is called a flux tube. In the regions of strong differential rotation, therefore, we may have flux tubes aligned in the toroidal direction. If a part of such a flux tube rises up and pierces the solar surface as shown in Figure 5, we expect to have two sunspots with opposite polarities at the same latitudes. But how can a configuration like Figure 5 arise? The answer to this question was provided by Parker[27] through his idea of magnetic buoyancy. We have seen in eq. (1) that a pressure $B^2/8\pi$ is associated with a magnetic field. If $p_{in}$ and $p_{out}$ are the gas pressures inside and outside a flux tube, then we need to have

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

(5)

to maintain pressure balance across the surface of a flux tube. Hence,

$$p_{in} \leq p_{out}$$

(6)

which often, though not always, implies that the density inside the flux tube is less than the surrounding density. If this happens in a part of the flux tube, then that part becomes buoyant and rises against the gravitational field to produce the configuration of Figure 5 starting from Figure 5.

A look at Figure 4 now ought to convince the reader that the sub-surface toroidal field in the two hemispheres should have opposite polarity. If this toroidal field rises due to magnetic buoyancy to produce the bipolar sunspot pairs, we expect the bipolar sunspots to have opposite polarities in the two hemispheres as seen in Figure 1. We thus see that combining the ideas of flux freezing, magnetoconvection and magnetic buoyancy, we can understand many aspects of the bipolar sunspot pairs. We now turn our attention to the central problem – the dynamo generation of the magnetic field.

### 4. The turbulent dynamo and mean field MHD

We now address the question whether it is possible for motions inside the plasma to sustain a magnetic field. Ideally, one would like to solve eqs (1) and (2) to understand how velocity and magnetic fields interact with each other. Solving these two equations simultaneously in any non-trivial situation is an extremely challenging job. In the early years of dynamo research, one would typically assume a velocity field to be given and then solve eq. (2) to find if this velocity field would sustain a magnetic field. This problem is known as the kinematic dynamo problem. One of the first important steps was a negative theorem due to Cowling[34], which established that an axisymmetric solution is not possible for the kinematic dynamo problem. One is, therefore, forced to look for more complicated, non-axisymmetric solutions.
A major breakthrough occurred in 1955 when Parker\textsuperscript{32} realized that the turbulent motions inside the solar convection zone (which are by nature non-axisymmetric) may be able to sustain the magnetic field. We have indicated in Figure 4 how a magnetic field line in the poloidal plane may be stretched by the differential rotation to produce a toroidal component. Parker\textsuperscript{32} pointed out that the uprising hot plasma blobs in the convection zone would rotate as they rise because of the Coriolis force of solar rotation (just like cyclones in the Earth’s atmosphere) and such helically moving plasma blobs would twist the toroidal field shown in Figure 6\textit{a} to produce magnetic loops in the poloidal plane as shown in Figure 6\textit{b}. Keeping in mind that the toroidal field has opposite directions in the two hemispheres and helical motions of convective turbulence should also have opposite helicities in the two hemispheres, we conclude that the poloidal loops in both hemispheres should have the same sense as indicated in Figure 6\textit{c}. Although we are in a high magnetic Reynolds number situation and the magnetic field is nearly frozen in the plasma, there is some diffusion (especially due to turbulent mixing) and the poloidal loops in Figure 6\textit{c} should eventually coalesce to give the large-scale poloidal field as sketched by the broken line in Figure 6\textit{c}.

Figure 7 captures the basic idea of Parker’s turbulent dynamo. The poloidal and toroidal components of the magnetic field feed each other through a closed loop. The poloidal component is stretched by differential rotation to produce the toroidal component. On the other hand, the helical turbulence acting on the toroidal component gives back the poloidal component. Parker\textsuperscript{32} developed a heuristic mathematical formalism based on these ideas and showed by mathematical analysis that these ideas worked. However, a more systemic mathematical formulation of these ideas had to await a few years, when Steenbeck, Krause and Rädler\textsuperscript{33} developed what is known as mean field MHD. Some of the basic ideas of mean field MHD are summarized below.

Since we have to deal with a turbulent situation, let us split both the velocity field and the magnetic field into average and fluctuating parts, i.e.

\begin{equation}
v = \overline{v} + v', \quad B = \overline{B} + B',
\end{equation}

Here the overline indicates the average and the prime indicates the departure from the average. On substituting eq. (7) in the induction eq. (2) and averaging term by term, we obtain

\begin{equation}
\frac{\partial \overline{B}}{\partial t} = \nabla \times (\overline{v} \times \overline{B}) + \nabla \times \varepsilon + \lambda \nabla^2 \overline{B};
\end{equation}

on remembering that \( \overline{v} = \overline{B} = 0 \). Here,

\begin{equation}
\varepsilon = \overline{v} \times \overline{B}'.
\end{equation}

It should be clear from this that \( \beta \) is the turbulent diffusion. This is usually much larger than the molecular diffusion \( \lambda \) so that \( \lambda \) can be neglected in eq. (13). It follows from eq. (11) that \( \alpha \) is a measure of average helical motion in the fluid. It is this coefficient which describes the production of the poloidal component from the toroidal component by helical turbulence. This term would go to zero if turbulence has no net average helicity (which would happen in a non-rotating...
Equation (13) is known as the *dynamo equation* and has to be solved to understand the generation of magnetic field by the dynamo process. A variant of this equation was first derived by rather intuitive arguments in the classic paper of Parker. The mean field MHD developed by Steenbeck, Krause and Rädler put this equation on a firmer footing. In the kinematic dynamo approach, one has to specify a velocity field $\mathbf{v}$ and then solve eq. (13). Using spherical polar coordinates with respect to the rotation axis of the Sun, we can write

$$\mathbf{v} = \Omega(r, \theta) r \sin \theta \, \mathbf{e}_p + \mathbf{v}_p,$$  \hspace{1cm} (14)

where $\Omega(r, \theta)$ is the angular velocity in the interior of the Sun and $\mathbf{v}_p$ is some possible average flow in the poloidal plane. Until a few years ago, almost all the calculations of the kinematic dynamo problem were done by taking $\mathbf{v}_p = 0$. If this is the case, then one has to specify some reasonable $\Omega(r, \theta)$ and $\alpha(r, \theta)$ before proceeding to solve the dynamo eq. (13). In the 1970s almost an industry grew up presenting solutions of the dynamo equation for different specifications of $\Omega$ and $\alpha$.

The first pioneering solution in rectangular geometry was obtained by Parker himself. He showed that periodic and propagating wave solutions of the dynamo equation are possible. Presumably this offers an explanation for the solar cycle. Sunspots migrate from higher to lower latitudes with the solar cycle because sunspots are produced (by magnetic buoyancy) where the crest of the propagating dynamo wave lies. Parker found that the parameters $\alpha$ and $\Omega$ have to satisfy the following condition in the northern hemisphere to make the dynamo wave propagate in the equatorward direction (so as to explain the butterfly diagram of sunspots):

$$\frac{\alpha \Omega}{\alpha r} \leq 0.$$ \hspace{1cm} (15)

Steenbeck and Krause were the first to solve the dynamo equation in a spherical geometry appropriate for the Sun and produced the first theoretical butterfly diagram of the distribution of sunspots in time-latitude. Then many dynamo solutions were worked out by Roberts, Köhler, Yoshimura, Stix and others. One might have felt complacent about the varieties of butterfly diagrams produced by these authors, however, it has to be admitted that many basic physics questions remained unanswered. Since nothing was known at that time about the conditions in the interior of the Sun, different authors were choosing different $\alpha$ and $\Omega$ subject only to the condition (15), and thereby were trying to fit the observational data better. Eventually it appeared that it was becoming a game in which you could get solutions according to your wishes by tuning your free parameters suitably. Further progress in solar dynamo theory became possible only by asking fundamental questions about the basic physics in the interior of the Sun, rather than by blindly solving the dynamo equation. These efforts will be described in the next section. It may be noted that all the authors of this period focussed their attention on explaining the equatorward propagation of sunspots by assuming that sunspots were produced in the regions where the toroidal component had the peak value. No serious attempt was made to connect the behaviour of the weak, diffuse magnetic field with the dynamo process or to explain the poleward migration of this field, although Köhler and Yoshimura presented some models that show a polar branch, i.e. a region near the poles where the dynamo wave propagates poleward.

5. Dynamo in the overshoot layer?

Where does the solar dynamo work? Since one needs convective turbulence to drive the dynamo, it used to be tacitly assumed in the early 1970s that the dynamo works in the solar convection zone and the different researchers of that period used to take $\alpha(r, \theta)$ non-zero in certain regions of the convection zone. This approach had to be questioned when Parker started looking at the effect of magnetic buoyancy on the solar dynamo. Magnetic buoyancy is particularly destabilizing in the interior of the convection zone, where convective instability and magnetic buoyancy reinforce each other. On the other hand, if a region is stable against convection, then magnetic buoyancy can be partially suppressed there (see, for example, § 8.8 of Parker). Calculations of buoyant rise by Parker showed that any magnetic field in the convection zone would be removed from there by magnetic buoyancy fairly quickly. Hence it is difficult to make the dynamo work in the convection zone, since the magnetic field has to be stored in the dynamo region for a sufficient time to allow for dynamo amplification.

It is expected that there is a thin overshoot layer (probably with a thickness of the order of 10^4 km) just below the bottom of the convection zone. This is a layer which is convectively stable according to a local stability analysis, but convective motions are induced there due to convective plumes from the overlying unstable layers overshooting and penetrating there. Several authors (Spiegel and Weiss, van Ballegooijen) pointed out that this layer is a suitable location for the operation of the dynamo. Although there would be enough turbulent motions in this layer to drive the dynamo, magnetic buoyancy would be suppressed by the stable temperature gradient there. This idea turned out to be a really prophetic theoretical guess, since helioseismology observations a few years later indeed discovered a region of strong differential rotation at the bottom of the solar convection zone. See the review by Christensen-Dalsgaard and Thompson in this issue on this subject. So it is certainly expected that a strong toroidal magnetic field...
should be generated just below the bottom of the convection zone due to this strong differential rotation. It may be noted that there have been other ideas as well for suppressing magnetic buoyancy at the bottom of the convection zone. Parker suggested ‘thermal shadows’, whereas van Ballegooijen and Choudhuri showed that an equatorward meridional circulation at the base of the convection zone can help in suppressing magnetic buoyancy there.

For about a decade starting from the mid-1980s, most researchers in this field believed that the whole dynamo process in the Sun, as summarized in Figure 7, takes place in the overshoot layer. Properties of such a dynamo operating in the overshoot layer were studied by DeLuca and Gilman, Gilman et al., and Choudhuri. If the dynamo operates in the overshoot layer, some new questions arise. Previously when the solar dynamo was supposed to work in the convection zone, the sunspots seen on the solar surface could be regarded as direct signatures of the dynamo process. One could assume that sunspots appeared wherever the dynamo produced strong toroidal fields just underneath the surface. On the other hand, if the dynamo works at the bottom of the convection zone, the whole depth of the convection zone separates the region where the magnetic fields are generated and the solar surface where sunspots are seen. In order to understand the relation between the solar dynamo and sunspots, one then has to study how the magnetic fields generated at the bottom of the convection zone rise through the convection zone to produce sunspots.

The best way to study this is to treat it as an initial-value problem. First an initial configuration with some magnetic flux at the bottom of the convection zone is specified, and then its subsequent evolution is studied numerically. The evolution depends on the strength of magnetic buoyancy, which is in turn determined by the value of the magnetic field. If the dynamo is driven by turbulence, one would expect an equipartition of energy between the dynamo-generated magnetic field and the fluid kinetic energy, i.e.

$$\frac{B^2}{8\pi} = \frac{1}{2}\rho v^2. \quad (16)$$

This suggests $B \approx 10^4$ G on the basis of standard models of convection. Because of the strong differential rotation, we expect the magnetic field at the bottom of the convection zone to be mainly in the toroidal direction. One, therefore, has to take a toroidal magnetic flux tube going around the rotation axis as the initial configuration. The evolution of such magnetic flux tubes due to magnetic buoyancy was first studied by Choudhuri and Gilman and Choudhuri. It was found that the Coriolis force due to the Sun’s rotation plays a much more important role in this problem than what anybody suspected before. If the initial magnetic field is taken to have a strength around $10^3$ G, the flux tubes move parallel to the rotation axis and emerge at very high latitudes rather than at latitudes where sunspots are seen. Only if the initial magnetic field is taken as strong as $10^8$ G, magnetic buoyancy is strong enough to overpower the Coriolis force and the magnetic flux tubes can rise radially to emerge at low latitudes.

D’Silva and Choudhuri extended these calculations to look at the tilts of emerging bipolar regions at the surface. Figure 8 taken from their paper shows the observational tilt vs. latitude plot of bipolar sunspots (i.e. Joy’s law) along with the theoretical plots obtained by assuming different values of the initial magnetic field. It is clearly seen that theory fits observations only if the initial magnetic field is about $10^8$ G. Apart from providing the first quantitative explanation of Joy’s law nearly three-quarters of a century after its discovery, these calculations put the first stringent limit on the value of the toroidal magnetic field at the bottom of the convection zone. Several other groups soon performed similar calculations and confirmed the result. The evidence is now mounting that the magnetic field at the bottom of the convection zone is indeed much stronger than the equipartition value given by eq. (16) (see Schüssler for a review of this topic).

If the magnetic field is much stronger than the equipartition value, it would be impossible for the helical turbulence (entering the mathematical theory through the $\alpha$ term defined in eq. (11)) to twist the magnetic field lines. The dynamo process, as envisaged in Figure 7, is therefore, not possible. We need a very different type of dynamo model. Schmitt and Feriz-Mas et al. proposed that the buoyant instability of the strong magnetic field itself may lead to a magnetic configuration which was previously thought to be created by helical turbulence. Parker suggested an ‘interface dynamo’ in which the helical turbulence acts in a region above the bottom of the convection zone. This idea has been further explored by Charbonneau and MacGregor. In the next section, we discuss what we regard as the most promising approach to build a model of the solar dynamo that can account for the very strong toroidal magnetic field at the bottom of the convection zone.
6. The Babcock–Leighton approach and hybrid models

We saw in § 4 and § 5 that one of the crucial ingredients in turbulent dynamo theory is the role of helical turbulence in generating the poloidal component from the toroidal component, which is mathematically modelled through mean field MHD. This approach to the dynamo problem will be called the Parker–Steenbeck–Krause–Rädler or the PSKR approach. In this approach, the dynamo is supposed to operate within a region where convective turbulence exists and no attention is paid to phenomena taking place at the solar surface. Babcock  and Leighton  in the 1960s developed a somewhat different approach, which we call the Babcock–Leighton or the BL approach. Even in this approach, the toroidal magnetic field is believed to be produced by the differential rotation of the Sun. For the production of the poloidal component, however, a totally different scenario is invoked. The strong toroidal component leads to bipolar sunspots due to magnetic buoyancy. We have noted that these bipolar sunspots have a tilt with respect to latitudinal lines (i.e. Joy’s law). Therefore, when these bipolar sunspots eventually decay, the magnetic flux spreads around in such a way that the flux at the higher latitude has more contribution from the polarity of the sunspot which was at the higher latitude. In this way, a poloidal component arises.

Compared to the PSKR approach, the BL approach was heuristic and semi-qualitative. The mean field MHD is, no doubt, based on some assumptions and approximations, and it is not clear whether these hold in the conditions prevailing in the Sun’s interior. However, for an ideal system satisfying these assumptions and approximations, the mean field MHD is a rigorous mathematical theory. No quantitative mathematical theory of comparable sophistication was developed for the BL approach. Nearly all the self-consistent dynamo calculations in the 1970s and 1980s, therefore, followed the PSKR approach. The BL approach was developed further by a group in NRL  who were studying the spread of magnetic flux from the decay of sunspots. There was growing evidence that there is a general meridional flow with an amplitude of about a few m s^{-1} near the Sun’s surface proceeding from the equator to the pole  . The poloidal magnetic field produced from the decay of tilted bipolar sunspots is carried poleward by this meridional circulation. As we have already pointed out in § 2, the weak diffuse magnetic field on the solar surface migrates towards the pole, in contrast to sunspots which migrate equatorward. One presumably has to identify the weak diffuse field as the poloidal component of the Sun’s magnetic field, whereas sunspots form from the much stronger toroidal component. The main aim of the NRL group was to model the evolution of the weak diffuse field, assuming that this was entirely coming from the decay of bipolar sunspots. No attempt was made to address the full dynamo problem. They even made the drastically simple assumption that the magnetic field is a scalar residing on the solar surface and the appropriate partial differential equation was solved only on this two-dimensional surface.

Dikpati and Choudhuri  and Choudhuri and Dikpati  attempted to make a vectorial model of the evolution of the weak diffuse field and to connect it to the dynamo problem. Since the meridional flow at the surface is poleward, there must be an equatorward flow in the lower regions of the convection zone, rising near the equator. If the dynamo operated at the base of the convection zone, then, in accordance with the ideas prevalent a few years ago, the poloidal component produced by this dynamo would be brought to the surface by the meridional circulation. This can be an additional source of the weak diffuse field at the surface, apart from the contributions coming from the decay of sunspots. Figure 9 from Choudhuri and Dikpati  shows a theoretical time-latitude distribution of the weak diffuse field on the surface, obtained by assuming a dynamo wave at the bottom of the convection zone as given. In other words, to produce this figure – which should be compared with the

![Figure 8](image-url)  

**Figure 8.** Plots of sin(\(\gamma\)) against sin(\(\lambda\)) theoretically obtained for different initial values of magnetic field indicated in kG. The observational data indicated by the straight line fits the theoretical curve for initial magnetic field 100 kG (i.e. 10^5 G). Reproduced from D’Silva and Choudhuri.

![Figure 9](image-url)  

**Figure 9.** A theoretical time-latitude distribution of the weak, strong magnetic fields on the solar surface, with half-butterfly diagrams’ obtained. A running dynamo wave assumed given at the left from the toroidal field by helical turbulence and

1482 CURRENT SCIENCE, VOL. 77, NO. 11, 10 DECEMBER 1999
envisaged in the PSKR approach – takes place at the base of the convection zone. For the generation of the poloidal field, we then invoke the BL idea that it is produced by the decay of bipolar sunspots on the surface. The meridional circulation can then carry this field poleward, to be eventually brought to the bottom of the convection zone where it is stretched by the differential rotation to produce the strong toroidal field. If a mean field formulation is made of the process of poloidal field generation near the surface, this hybrid model of the dynamo will incorporate the best features of both the PSKR and the BL approaches. On the one hand, detailed quantitative calculations will be possible, as in the PSKR approach. On the other hand, the surface phenomena emphasized in the BL approach, are integrated in the dynamo problem. The meridional circulation plays an important role in this hybrid model so that a suitable form of $\alpha$ in eq. (14) is to be specified. It is hoped that this hybrid model will account for both the equatorward migration of the strong toroidal field at the base of the convection zone and the poleward drift of the poloidal field at the surface.

This hybrid model has one other attractive feature. Researchers in 1970s built kinematic models of the solar dynamo by arbitrarily specifying $\alpha(r, \theta)$ and $\Omega(r, \theta)$. These were regarded as free parameters to be tuned suitably so as to give solutions with desired characteristics. In the present hybrid models, these important ingredients to the dynamo process are directly based on observations. Helioseismology has given us $\Omega(r, \theta)$, with its shear concentrated at the base of the convection zone. The $\alpha$ coefficient also arises out of the observed decay of bipolar sunspots on the surface. Previously it used to be even debated whether $\alpha$ is positive or negative. Researchers used to fudge $\alpha$ such that the inequality (15) was satisfied. The direction of tilt of bipolar sunspots on the surface, however, clearly indicates that $\alpha$ arising out of their decay has to be positive in the northern hemisphere: a point made by Stix$^{68}$ long ago. Once these key ingredients are fixed directly by observation, we no longer have the freedom to fudge them according to our wishes, which researchers in 1970s could do. This leads to no problem. Helioseismology shows that $d\Omega/dr$ is positive in lower latitudes where sunspots are seen. If $\alpha$ is also positive in the northern hemisphere, then clearly inequality (15) is not satisfied and dynamo waves are expected to propagate poleward!

The first calculations on the hybrid model were reported by Choudhuri et al.$^{68}$. Dynamo waves were indeed found to propagate poleward, if meridional circulation was switched off. The toroidal field and the poloidal field are respectively produced in layers near the base of the convection zone and near the solar surface. When meridional circulation is switched off, any field can reach from one layer to the other layer by diffusion with a time scale of $L^2/\beta$ where $L$ is the separation between the layers (i.e. the thickness of the convection zone). If the meridional flow has a typical velocity of the order $V$, it takes time $Vt/L$ for the meridional circulation to carry some quantity between the two layers. When the time scale $Vt/L$ is shorter than the diffusion time scale $L^2/\beta$, the problem is dominated by meridional circulation and Choudhuri et al.$^{68}$ found that the strong toroidal component at the bottom of the convection zone actually propagates equatorward, overriding the inequality (15). Thus the inequality (15), which was regarded as sacrosanct for four decades since Parker$^{72}$ obtained it, is found not to hold in the presence of a meridional circulation having a time scale shorter than diffusion time, thereby opening up the possibility of constructing realistic hybrid models of the solar dynamo. Further calculations on hybrid models have been reported in a series of papers$^{66–73}$. It should be emphasized that all these studies are still of rather exploratory nature. They demonstrate the viability of the hybrid models and study their different characteristics. We are, however, still far from building a sufficiently realistic model, putting in all the details, that would account for the observational data presented in Figure 2. Achieving this should be our goal now.

7. Miscellaneous ill-understood issues

Since this is a review in a special section on solar physics, we have primarily discussed those aspects of kinematic dynamo models which directly pertain to the matching of theory with observational data. It should, however, be kept in mind that many fundamental issues of dynamo theory are still very ill-understood. Until we have a better understanding of these issues, the kinematic models can, at best, be considered superficial attempts at a very deep physics problem. The reader may look up the IAU Symposium volume on The Cosmic Dynamo$^{74}$ for several articles dealing with these fundamental issues. Here we make only very brief comments on some of these issues.

We have seen in § 4 that the turbulent dynamo theory is developed by averaging over turbulent fluctuations. The existence of magnetic flux concentrations clearly indicates that the fluctuations are much larger than the average values (often by orders of magnitude). Does a mean field theory make sense in such a situation? Can we trust the perturbative procedures like the first-order smoothing approximation? Hoyng$^{75}$ raised some questions regarding the interpretation of the averaged quantities. The dynamo eq. (13) admits of several possible modes in spherical geometry: the preferred mode seems to be the mode with dipole symmetry, wherein the toroidal component is oppositely directed in the two hemispheres. This mode approximately corresponds to the observational data. However, Stenflo and Vogel$^{76}$ pointed out that one hemisphere of the Sun often has more sunspots than the other, indicating that there may be a superposition with higher modes having different symmetry. Analysing the
statistics of sunspot data for several decades, Gokhale and Javaraiah claimed to have found evidence for multiple modes. If the fluctuations are so large, there is no reason why a particular mode should be very stable, or why higher modes should not be excited. The interference of modes with different symmetry was theoretically studied by Brandenburg et al. employing a nonlinear dynamo model.

Since the toroidal magnetic field is far stronger than the equipartition value, it is certainly not justified to assume that the magnetic fields do not back-react on the flow. One should therefore ideally solve eqs (1) and (2) simultaneously, instead of proceeding with kinematic models. Since this is a fairly difficult job even by the standard of today’s computers, attempts are made to include the back-reaction of the magnetic field within kinematic models instead of going to fully dynamic models. One easy way to incorporate the back-reaction in the dynamo eq. (13) is to make the crucial quantity decrease with the magnetic field, following some prescription like:

\[
\alpha = \frac{\alpha_0}{1 + (B/B_0)^2}.
\]  

(17)

The effect of such \(\alpha\) quenching on the dynamo process has been extensively studied. When the velocity field \(\mathbf{v}\) is specified, the dynamo eq. (13) is a linear equation for the magnetic field \(\mathbf{B}\); we assume the various coefficients in the equation to be independent of \(\mathbf{B}\). If \(\alpha\) is quenched by \(\mathbf{B}\), in accordance with eq. (17), we have a nonlinear problem. One important question is whether the irregularities of the solar cycle, as seen in Figure 3, can be explained with the help of nonlinear models. It seems that the nonlinearity introduced through eq. (17) cannot cause such chaotic behaviour. Since a sudden increase in the amplitude of magnetic field would diminish the dynamo activity by reducing \(\alpha\) and thereby pull down the amplitude again (a decrease in the amplitude would do the opposite), the \(\alpha\) quenching mechanism tends to lock the system in a stable mode once the system relaxes to it. In fact, Krause and Meinel argued that nonlinearities must be what makes one particular mode of the dynamo so stable. Only by introducing more complicated kinds of nonlinearity (with suppression of differential rotation) in some highly truncated dynamo models, Weiss et al. were able to find the evidence of chaos. Jennings and Weiss presented a study of symmetry-breakings and bifurcations in a nonlinear dynamo model. Since \(\alpha\) quenching of the form (17) cannot explain the irregularities of the solar cycle, Choudhuri explored the effect of stochastic fluctuations on the mean equations and obtained some solutions resembling Figure 3. Several subsequent papers explored this possibility further.

Finally we comment on the efforts in building fully dynamic models by solving both eqs (1) and (2) simultaneously. This is a highly complicated nonlinear problem and can only be tackled numerically. Gilman and Glatzmaier presented very ambitious numerical calculations in which convection, differential rotation and dynamo process were all calculated together from the basic MHD equations. These calculations, however, gave results which do not agree with observational data. For example, angular velocity was found to be constant on cylinders, whereas helioseismology found it to be constant on cones. If various diffusivities were set such that the surface rotation pattern was matched, the dynamo waves propagated from the equator to the pole. The codes of Gilman and Glatzmaier naturally had finite grids, and the physics at the sub-grid scales was modelled by introducing various eddy diffusivities. Probably the physics at sub-grid scales is more subtle and the details of it are crucially important in determining the behaviour of the dynamo. This is generally believed to be the reason why these massive codes did not produce agreement with observations. The subsequent approach in numerical modelling has been to do dynamic calculations over cubes which correspond to small regions of the Sun, rather than trying to build models for the whole Sun. Brandenburg et al. and Nordlund et al. have followed this approach.

8. Conclusion

It seems that the solar magnetic fields are generated and maintained by the dynamo process. There is only a small minority of solar physicists who would disagree with this point of view. It is, however, not easy to build a sufficiently detailed and realistic model of the dynamo process to account for all the different aspects of solar magnetism. The 1970s happened to be a period of optimism in dynamo research when various researchers were producing butterfly diagrams by choosing different forms of \(\alpha(r, \Theta)\) and \(\Omega(r, \Theta)\). It was felt that further research would narrow down the parameter space and establish a standard model of the solar dynamo. As we discussed above, this did not happen. In 1993 Schmitt wrote in his review on the solar dynamo: ‘The original hope that detailed observational and theoretical information would yield better results of the dynamo did not prove true, on the contrary, they raised difficulties instead’. Today, a few years afterwards, we can perhaps have a less pessimistic outlook. The hybrid models, which are closely linked to observations and which combine together some of the best ideas that came out of dynamo research in the last few decades, certainly do look promising. Although sufficiently detailed models have not yet been worked out, we hope that we are close to building kinematic models which are much more realistic and sophisticated than the kinematic models of the 1970s. Perhaps other researchers may regard this point of view as a reflection of this author’s personal prejudice. Only time will tell if this prejudice is justified.

Finally we end by cautioning the reader that this article should not be regarded as a comprehensive review of the
solar dynamo problem. Since this article is primarily aimed not at the experts but at more general readers, we have emphasized those aspects of kinematic models which have direct relevance to observations. A very incomplete discussion of various fundamental issues is presented in the § 7. There is no doubt that kinematic models can never fully satisfy us. The ultimate challenge is to build fully dynamic models starting from the basic equations, and then to explain both the fluid flow patterns and magnetic patterns in the interior of the Sun in a grand scheme. As we have pointed out in § 7, the modern computers still seem inadequate for handling this problem. The solar dynamo problem will certainly remain alive for years to come.