Magnetic Fields in The Sun

D. J. Mullan

Bartol Research Foundation of The Franklin Institute
Swarthmore, Pennsylvania 19081

Abstract

The observed properties of solar magnetic fields are reviewed, with particular reference to the complexities imposed on the field by motions of the highly conducting gas. Turbulent interactions between gas and field lead to heating or cooling of the gas according as the field energy density is less or greater than the maximum kinetic energy density in the convection zone. The field strength above which cooling sets in is 700 gauss.

A weak solar dipole field may be primeval, but dynamo action is also important in generating new flux. The dynamo is probably not confined to the convection zone, but extends throughout most of the volume of the sun. Planetary tides appear to play a role in driving the dynamo.
I. Introduction

As far as inhabitants of the Earth are concerned, the most important characteristic of the sun is that it provides a source of energy which has remained essentially constant (apart from short interruptions during ice ages) for at least a billion years. Solar energy is generated in thermonuclear processes near the center of the sun, and this energy is carried outwards towards the surface by radiation for most of the way. Near the surface, convection becomes the principal energy carrier, and the visible surface of the sun is covered with convection cells. The whole process of energy transfer is spherically symmetric, and steady in time (at least in a statistical sense).

There are, however, more or less transient localized phenomena which appear at the solar surface from time to time. Among these are explosions called solar flares, dark areas called sunspots, and others. These transient features, which are described by the general title of "solar activity" owe their existence to the presence of magnetic fields in the sun. If such fields did not exist, the most energetic feature observable on the sun would be the hydrodynamic convection. There would be no spots, no flares, and no radio bursts; and we on Earth would never observe aurorae, nor experience radio communications blackouts. But there are magnetic fields in the sun, and in order to understand transient solar phenomena and how they affect life on Earth, we must first understand the properties of the solar field.
II. Complexity of the Solar Field

As far as the physics of the solar field is concerned, the most important properties of the sun are (i) high electrical conductivity, and (ii) gas motions, both (a) small-scale turbulent, and (b) large-scale ordered. The electrical conductivity of gas near the surface of the sun is so high that the field and the gas interact strongly with each other. The length scales involved are much larger than those which can be studied in laboratory conditions, so the physics of the field-gas interactions in laboratory plasma cannot always be extrapolated in a simple way to predict what will happen in solar conditions. But to a first approximation, we can consider the two types of gas motion (a) and (b) separately, and understand to some extent how they interact with the field.

(a) In the case of turbulent gas flow, the relative energy densities of the field \( E_B \) and of the gas motions \( E_K \) are important in determining the interaction. An example of small scale turbulent gas flow is the circulation in short-lived (5-10 minutes) convection cells called granules. It appears that when the kinetic energy density of the granule motions \( E_K \) exceeds \( E_B \), the presence of a field causes the gas to be locally hotter than it would be in the absence of the field. On the other hand, if \( E_B \) exceeds \( E_K \), the interaction is such as to make the gas locally cooler than normal (see below).

(b) A different type of interaction occurs when the gas flow is ordered on a large scale, such as in the horizontal flows observed at the tops of large convection cells called supergranules.
This persistent flow sweeps up magnetic field lines and flux tubes and pushes them to the peripheries of the supergranules. The latter lose their identity after a lifetime of about one day, so the field lines are pushed around today by different supergranules from those which pushed them around yesterday. But at any one time this type of field-gas interaction gives rise to a very characteristic network of vertical field lines concentrated around the edges of supergranules. This in fact is the most characteristic feature of the solar field in quiet regions of its surface: a network of field enclosing cells of relatively field-free gas (1). When the gas motions are of the ordered type, there is no clear relation between the energy density of the field at supergranule peripheries and the kinetic energy density of the supergranule flows (2-4). The reason is that the gas flows are persistent, and, given enough time, the flows can move field lines even if the field energy density is large. This is particularly well demonstrated by the fact that even the very large fields in sunspots are gradually eroded by the persistence of supergranule flow at the edge of the spot. Computer simulations suggest that in two-dimensional flow, the maximum $E_B$ at the edges of supergranules is 4-10 times greater than the maximum $E_K$ in the cell (4). However, the strongest fields are concentrated in an area much smaller than the cell area, so that when the field is averaged over the entire cell area, the mean $E_B$ turns out to be no larger than the mean $E_K$.

These are only the first steps in the complex interactions between fields and motions. An example of the next highest order of
interactions is provided by the differences between the gas motions permitted above supergranule peripheries, where the fields are strong, and those permitted above the cells, where fields are weak. In the weak field regions, the gas moves freely in a wide variety of flow configurations, including large-amplitude oscillations at a period of 5 minutes. However, in the strong field regions, chromospheric gas flows are restricted to essentially one unique flow pattern, and the 5-minute oscillations are greatly reduced in amplitude (5, 6).

III. Observations of the Solar Magnetic Field
(a) The "General" Field

At first sight, eclipse photographs of the solar corona suggest that the sun has a large scale magnetic field which is similar (at least in the vicinity of the rotation poles) to a magnetic dipole. Field strengths on the solar surface can be obtained by means of a magnetograph, and near the rotation poles of the sun, the surface fields are found to be about 1 gauss (7). However, the polarities of the fields on the surface are not consistent with those of a dipole field of strength 1 gauss, for there are long periods of time (e.g., June 1957-August 1958) when both north and south rotation poles of the sun have the same magnetic polarity (8). Moreover, from 1964 up to at least 1970, both poles had a predominance of south-seeking flux (9). At times, the general field at one of the rotation poles disappears for periods of some months, although the field at the other pole persists. The mean field strength varies as
a function of latitude in a way which is inconsistent with that expected of a dipole, for the mean field changes sign very rapidly with latitude, first changing sign (from that at the pole) at latitudes as high as 55-70°. There are irregular time fluctuations in the general field near the poles, on time scales as short as one day, with almost synchronous appearance of peaks in the field at both rotation poles (9). The sun at times appears to have the properties of a magnetic quadrupole, and it is in fact this characteristic which is imprinted on the field which escapes from the sun into interplanetary space (see below).

However, it must be stressed that the sensitivity of current magnetographs is not sufficient to rule out the possibility that the sun does indeed have a true dipole field with strength of order 0.1-1.0 gauss, or weaker, at the poles.

(b) Localized Fields

Far from the rotation poles, the solar field no longer shows any similarity to a dipole field. In a broad band of solar latitudes centered on the equator, all vestiges of any dipole field which the sun might possess are overwhelmed by localized areas of enhanced field strength called active regions. These regions are transient, surviving for periods of a few weeks or months, after which they gradually become more diffuse, and spread their field lines over progressively larger areas of the sun. As the fields spread out, the mean field strength decreases, and the fields of order 1 gauss observed at the rotation poles are believed to be the relics of old
active regions, transported from low to high solar latitudes by a random walk diffusion along supergranule peripheries (10).

The simplest description of active regions is that they are generally elongated from east to west, with opposite magnetic polarities at the eastern and western ends. At any given time, most active regions in the northern hemisphere of the sun have the same ordering of magnetic polarities, in the sense that the eastern end of most active regions would have, say, south-seeking polarity, while the western end would have north-seeking polarity. At the same time, most active regions in the southern hemisphere would have the reverse polarities. The number of active regions on the surface of the sun varies with a period of about 11 years; the most recent maxima occurred in 1957-58, and 1968-69, with a minimum in 1964-65. magnetically, the period is not 11 years, but 22 years, because the ordering of magnetic polarities in active regions in a given hemisphere reverses from one 11-year half-cycle to the next.

More detailed examination of active regions indicates the east-west separation of polarities is a greatly over-simplified description. The fields are actually extremely complex, with new flux tubes emerging from time to time, causing the intrusion of new bipolar pairs into the older fields. Elements of one polarity sometimes disappear with the simultaneous appearance of the opposite polarity. At other times, elements of one polarity disappear or appear without any accompanying changes in the opposite polarity (11). Once new field lines or bipolar pairs have appeared, gas motions move the field lines around on a large scale (>20 thousand km). The more complex the intrusions of opposite polarity in an active region are,
the more likely is the region to be an active one, for it is in regions where elements of opposite polarity are pushed close together that the most energetic phenomena are observed, such as flares. In regions where activity is not so high, the boundaries separating areas of opposite polarity are relatively stable, and are often strikingly delineated by long-lived dark filaments. At the limb of the sun, these filaments appear as prominences, and they are thought to indicate cool gas trapped in a depression at the top of arched field lines (3).

Active regions are characterized in optical radiation by bright patches of enhanced emission called plages, where the magnetic fields are of order several hundred gauss or less. Some active regions also develop dark patches called sunspots, where the fields are of order 1-5 kilogauss (12). Fields in spots are the strongest known fields on the surface of the sun. At Bartol Observatory, spot fields are measured by photoelectric polarimetry of one of the strong absorption lines of sodium, which is formed over a wide range of altitudes in a sunspot atmosphere. Observations of the polarimetric profile of such a strong line can lead to information on the gradient of field strength as a function of altitude above the spot.

The most striking feature of solar magnetic fields, apart from their existence, is the fact that they can lead to both brightening and darkening of the solar surface. If the field is less than 700 gauss (13), increasing field strength leads to increased brightness (Fig. 1), especially in regions where the field lines are predominantly horizontal. But if the field exceeds 700 gauss, and
particularly if it exceeds 1200-1400 gauss (12), the brightness falls off dramatically, especially if the field lines are vertical. In one of the spots observed at Bartol, for example, the brightness of the spot at 4000 Angstroms was only 2% of the normal solar brightness. The transition from brightening to darkening as $B$ increases through 700 gauss is related to the maximum kinetic energy density $E_K$ of the turbulent gas motions in the solar convection zone. Using appropriate densities and convective velocities (14), it is found that the maximum $E_K$ is $1.8 \times 10^4$ ergs/cm$^3$. This corresponds to the energy density $E_B$ of a field of $B = 670$ gauss. Hence when $B \leq 700$ gauss, $E_K > E_B$, and the turbulent gas can jostle the field lines, generating hydromagnetic waves of relatively short period. These waves are rapidly dissipated in regions of weak field (dissipation length $\approx B^3$ (15)), particularly if the local field lines are horizontal, so that the waves can be trapped at low altitudes, and prevented from propagating upwards into the low density gas at higher altitudes. It is the dissipation of the waves which heats the gas locally and causes the plage to appear brighter than normal. However, in sunspots, where $B > 1200$ gauss, $E_B > E_K$, and so the field controls the turbulent gas motions, thereby impeding the normal process of convective energy flow upwards to the solar surface. Convection is not altogether suppressed, but gas velocities are reduced, and the circulation time around a convection cell becomes longer. The field lines in the spot are distorted by these slow convective motions, and the Lundqvist number (16) is so high that copious fluxes of hydromagnetic waves are generated in the spot.
These relatively long period waves are hardly dissipated at all, partly because the field strengths in the spot are large, partly because the field lines are vertical so that the waves are not easily trapped, and partly because the periods $P$ of the waves are long (dissipation length $\propto P^2$ (15)). The waves can therefore escape from the regions in the spot where they are generated, and carry away a large fraction of the solar thermal energy flux in invisible form. The spot therefore remains dark as long as some of the energy flux can be transported away by the waves. At Bartol, a steady-state model of a sunspot has been developed which describes quantitatively the extent to which convection is impeded in a spot, on the assumption that the invisible flux is in the form of Alfvén waves (17). This model has also been successful in explaining the properties of enormous dark spots which have been detected recently on the surfaces of nearby red dwarf stars (18).

The invisible energy flux in a spot is presumably redistributed over an area in the vicinity of the spot. Part of the observing program at Bartol is aimed at measuring both the coolness of a spot, and also the distribution of brightness around the spot. For the latter type of measurement, the Bartol telescope is especially suited because photometric accuracy of order 0.1% is attainable. This is at least an order of magnitude better than has been achieved in the past using photographic techniques, and our sensitivity is now sufficient to detect excess flux around spots even if the redistributed flux is spread over an area as large as 100-1000 times the area of the spot.
(c) Short-lived Events

Sunspots and plages are relatively long-lived phenomena in the life of an active region. The presence of a magnetic field also gives rise to events which are short-lived, lasting only 100-1000 seconds. These events are called flares, and they involve the release of a large amount of energy ($10^{32}$ ergs) from a small volume ($10^{27}$ cm$^3$) (19). Flares are generally observed to occur near sunspots, and it is believed that only large magnetic fields, such as occur in spots, can provide sufficient energy density to energize a flare. The importance of the magnetic field in the flare process is indicated by the fact that flares tend to lie close to horizontal field lines (20), especially if the horizontal gradient of vertical field strength near the flare is large. It has been believed that there is direct observational evidence for conversion of energy stored in the magnetic field into flare energy (21), and this belief has given rise to a class of flare theories which depend on building up a store of magnetic energy for some time (of order one day or longer), until conditions become favorable for explosive release of the stored energy in some type of event which rearranges field lines (22,23). However, there are many flares where no measurable change of magnetic energy in the active region can be detected (24), and there are theoretical reasons for doubting that energies as large as $10^{32}$ ergs could be stored without making the fields unstable (3). These facts suggest that other flare models, which do not rely on stored magnetic energy directly, cannot be excluded. It is possible that
the flare energy emerges from the photosphere either immediately preceding, or even during, the course of the flare (25,26).

Whatever the source of flare energy, the products of flares give rise to certain characteristic phenomena on the Earth, such as blackouts in short-wave radio, aurorae, cosmic ray decreases, and fluctuations in the Earth's magnetic field. (During the large flares of August 1972, fluctuations in the Earth's field were sufficiently large to cause concern in military circles about possible triggering of magnetic mines in the harbors of Hanoi and Haiphong.) These phenomena represent our most direct connection with magnetic fields in the sun.

(d) The Field Above the Solar Surface

Eclipse photographs of the corona suggest that the solar field extends for some distance beyond the surface of the sun. How far out does the field extend? Very far indeed, it turns out. Once the surface fields are known, the field equations can be integrated upward (assuming that the currents are either zero or parallel to the field lines) to determine the fields in the corona (27). It appears that vertical field lines in the photosphere fan out only slightly all the way up through the chromosphere, so that the mean field strength falls off by a factor of only 1.5 between the photosphere (where the 5250Å line of FeI is formed) and the chromosphere (where Hα is formed) (28). Thus there is very good correlation between features observed in photospheric magnetograms and those observed in Hα magnetograms (29,30). However, when the
field lines reach the corona, the lines fan out very markedly, so that at the level of formation of the MgX line at 625Å (corresponding to mean temperatures of $1.6 \times 10^6$°K), details of the photospheric field pattern are washed out (31). The widely spreading fans of field lines in the corona either arch back to the surface to link up with distant active regions, or else open out into interplanetary space (32). The hydrodynamic expansion of the corona then carries the field away from the sun with velocities of several hundred km/sec. By the time the field has been dragged out as far as the Earth's orbit, direct satellite measurements of field strength are possible. The field is found to be very weak, $10^{-5}-10^{-4}$ gauss, although this is still sufficiently large to control the propagation of most of the cosmic rays which are ejected from the sun during large flares. The field can impose anisotropies on the cosmic rays which are detectable by ground-based neutron monitors such as the one at Bartol (33).

However, the most significant result to emerge from the satellite measurements was not so much the strength of the field as the direction of the field at the Earth's orbit. The field was found to retain an overall direction either towards or away from the sun within large sectors extending 60-90° in solar longitude (34). At solar minimum, four sectors persisted in a quasi-stationary configuration for approximately one year (35). The sector boundaries can be traced back onto the solar surface, where they are found to correlate well with boundaries between regions of opposite magnetic polarity in the large scale photospheric field (36,37). These re-
remarkable results suggested that, unknown to previous investigators, there is hidden in the complex fields near the solar equator a large scale background weak field (about 1 gauss) which is more important in determining the properties of the field escaping from the sun than is the stronger field which appears to dominate photospheric magnetograms. The large scale field appears to be composed of unipolar magnetic regions which are the remains of old active regions (38). It is the large scale of these weak fields which allow them to extend from the photosphere upward to a "source surface" (39) at which coronal expansion begins to drag the field outward from the sun.

On the other hand, the strong fields which have been concentrated in the network by supergranule flow have length scales which are so small that the field lines cannot reach up to the source surface, so that at the Earth's orbit, we cannot detect the effects of network field. The weak background fields are the components which have escaped the concentrating action of supergranule flow (40). (The fact that some field must escape this concentrating action is also a necessary feature of one of the most detailed models of the solar magnetic cycle (57).) The network component of the field, although it is certainly stronger than the weak background component, may in fact have a pattern imposed on it by the action of the large-scale component of the field. Thus active regions are unusually active if they lie to the east of a sector boundary, while on the western side of the sector boundary, active regions are unusually quiet (41).

The large scale field seems to be a more fundamental formation than the network component (42), and the large scale field has the peculiar property that it rotates rigidly with a period of
27 days (43) rather than partaking in the differential rotation which is observed in the surface layers of the sun. The 27-day period may represent a rigidly-rotating subsurface layer which has a measurable effect on the magnetic pattern at the photosphere (43). If it is true (as has been proven to a certain approximation (44)) that the differential rotation observed at the surface of the sun must persist all the way to the bottom of the convection zone, then the rigidly-rotating large scale field must originate in the gas beneath the convection zone.

When the solar field has been dragged far past the orbit of the Earth, out to 10-100 times the Earth's orbital radius, the field must adjust itself across a shock front to merge with the magnetic field in interstellar space. The region within the shock front is called the solar cavity, with a volume exceeding the volume of the sun by a factor of more than $10^9$. As far as cosmic rays from interstellar space are concerned, the sphere of influence of the sun is the solar cavity, for such cosmic rays can reach the orbit of the Earth only to the extent that the solar magnetic field within the cavity permits them to diffuse inwards. The presence of the solar field therefore extends the sun's cosmic sphere of influence by more than $10^9$ compared with that which the sun would have if the sun had no magnetic field.

(e) Small-Scale Features in the Field

Magnetographs currently in operation cannot resolve directly magnetic features smaller than about 2 arc seconds in diameter.
However, by an ingenious choice of Zeeman-sensitive absorption lines, Stenflo (45) showed that field strengths determined by magnetographs interpreted in the conventional way are actually only large scale averages of the true field. The field at the solar surface was found to consist of small clumps of field lines, no larger than 100-300 km in diameter (0.1-0.4 arc sec). At the centers of the clumps, the field strengths must be very large, at least 2000 gauss. Magnetically, the best description of the sun now is a "magnetic pin-cushion" (46), with magnetic pins protruding all over the surface. The entrance aperture of current magnetographs covers areas typically 50-500 times larger than a magnetic pin, so if one magnetic pin falls within the entrance aperture, the mean field in that particular area of the surface will appear to have a mean value of about 10 gauss. (Very small clumps, with diameters about 100 km or less, have been observed by optical interferometry in sunspot umbrae (47), but the relation between these elements and Stenflo's clumps is not yet clear.) Active regions are areas where the surface density of magnetic pins is higher than normal, with plage granules (48) possibly to be identified with individual pins. Supergranule peripheries may also be the loci of high-density magnetic pin regions. (It is not clear whether supergranules should be considered as agents to concentrate magnetic field lines, thereby amplifying the field strength, or as sweepers of magnetic pins, where the field strength is already high. In the literature, there are contradictory claims as to whether supergranule flow is or is not sufficient to build up the fields observed at supergranule
peripheries (2,4,49).) Around sunspots, magnetic pins are observed to stream horizontally outwards (50). Even though the field strengths in the magnetic pins are as large as those in some sunspots, the gas in the magnetic pins is not darker than normal, because the pins have diameters which are almost an order of magnitude smaller than the observed diameters of convection cells near the solar surface (diameters 1000-2000 km). Convection is therefore not significantly impeded by the strong fields in the magnetic pins.

The essential point made by Stenflo (45) is that Zeeman splitting of a solar absorption line is not easily interpretable. Factors other than the field affect the line profile, and these factors must be taken into account. For example, magnetic fields of a few hundred gauss cause local heating, and this heating can have a very significant effect on the absorption line profile formed in the local gas. Hence, in order to extract information about magnetic field strength from the line profile, the magnetograph must be calibrated with an absorption line formed in gas of the same temperature in field-free conditions. The difficulty is that it is not obvious where one should look in order to observe hot gas which is free of magnetic field, since the heating is essentially due to the field itself.

These difficulties have implications also for our understanding of solar flares. It is certainly true that Zeeman-sensitive lines have been observed to change following a flare (21), and these changes have traditionally been interpreted as indicating a change in magnetic field strength. Now, however, it must be admitted
that part (if not all) of the changes observed in absorption line profiles following flares must be due to changes in the temperature of the local gas. It is no longer clear whether the photospheric magnetic field really changes at all during a flare, except insofar as the energy released by the flare may force the field to change. The traditional theories, in which changes in the field cause the flare, would then need to be reversed, and a flare would be considered as the prime cause of changes in the field, rather than vice versa.

IV. Why Does the Sun Have a Magnetic Field?
(a) Primeval (or "Fossil") Fields

It is highly probable that the sun must have had a magnetic field ever since it became a start. According to current ideas about star formation (51), the sun formed when a cloud of interstellar gas contracted as a result of thermal or gravitational instability. Interstellar gas clouds have densities of typically 10 hydrogen atoms/cm$^3$, and temperatures of about 100$^\circ$K. (For a review of the properties of the material in interstellar space, see (52).) The radius of the cloud which eventually formed the sun must have been of order 3x10$^{18}$ cm in order to have a mass equal to the solar mass (2x10$^{33}$ gm). Such a cloud would occupy almost all of the volume between the present sun and the nearest star. As the cloud contracts to the present radius of the sun (7x10$^{10}$ cm), the radius of the cloud must decrease by a factor of 4x10$^7$. This contraction has an effect on the properties of the interstellar cloud, such as the magnetic field and the angular velocity of rotation. The fact that a mag-
Magnetic field threads through interstellar space has been established by several independent methods, such as Zeeman splitting of the 21-cm hyperfine transition of hydrogen, synchrotron radiation from cosmic ray electrons, Faraday rotation of the plane of polarization of radiation from distant radio sources, and alignment of dust grains in space. The most probable magnetic field strength in the interstellar gas is 1-10 microgauss.

Consider now what happens to this field as the cloud contracts to the present radius of the sun. The time-scale for contraction is (53) of the order of $10^7$ years ($t = \sqrt{3/32\pi G \rho}$, $G =$ gravitational constant, $\rho =$ gas density). During this time, the field lines do not have time to slip appreciably away from the gas, if the electrical conductivity is assumed to have the classical value (54), $\sigma = 10^{-14} T^{3/2}$ emu. The amount of slippage permitted in $10^7$ years ($L = \sqrt{t/4\pi \sigma}$) in a gas with $T = 100^\circ$K is less than one-millionth of the initial radius of the gas cloud. The field is therefore essentially frozen in, and the magnetic flux linking the cloud must remain constant. With an initial field $B_1$, when the cloud radius is $R_1$, the field must grow to $B_2 = B_1 (R_1/R_2)^2$ when the radius has contracted to $R_2$. With $R_1 = 3 \times 10^{18}$ cm, $R_2 = 7 \times 10^{10}$ cm, and $B_1 = 1-10$ microgauss, the present sun should have fields of order $10^9-10^{10}$ gauss. (By a similar argument concerning conservation of angular momentum, any initial angular velocity the cloud possessed would have been amplified by $(R_1/R_2)^2$ during collapse. If the initial angular velocity is of the order of galactic rotation, $10^{-15}$ sec$^{-1}$, the present sun should rotate in about 1 second, if nothing had
acted to brake the rotation.)

Fields as strong as $10^9-10^{10}$ gauss have never been observed in the sun, and if they were to be present, they would have a significant effect on the internal structure of the sun. For example, a field of order $10^9$ gauss at the center of the sun leads to an increased solar neutrino flux which is about an order of magnitude larger than the observed upper limit on the neutrino flux (55). There are other reasons why the field inside the sun is probably not as large as $10^9$ gauss. Thus, in view of the fact that the larger the scale on which a field is ordered, the longer it persists, it is to be expected that any vestiges of a primeval field which have managed to survive to the present time would be ordered on large scales. Such a field would, however, cause the sun to be oblate by a factor which exceeds the observed oblateness by a factor of $10^6-10^8$ (56). Moreover, in discussing the collapse of an interstellar cloud to form the sun, it is not correct to assume that the electrical conductivity has the classical value. During the collapse, the sun has probably passed through a completely convective phase (53), during which the field lines are tangled by turbulent gas motions, thereby greatly accelerating field dissipation. The internal field would not then have grown as large as $10^9$ gauss, and in fact the oblateness of the sun sets an upper limit of $2\times10^6$ gauss on the permissible large scale field inside the sun (56). An indication of the great rapidity with which fields can be dissipated in turbulent gas is provided by the surface convection zone in the sun, where diffusion is so rapid that fields are obliterated in a
time of about $10^3$ years \(57\). Hence, if the primeval field persists at all, it must lie below the convection zone, and the only time we would see primeval field would be when a flux tube would rise into the convection zone and break through the surface. The fact that sunspot fields are never observed to exceed several thousand gauss indicates that at least near the surface of the sun, the internal field strength is much less than the permissible upper limit of $2 \times 10^6$ gauss. Nevertheless, it is an unfortunate fact of life that an extremely large field could exist inside the sun without ever manifesting itself as a field at the surface \(58\).

(b) Fields Generated Inside the Sun

So far, we have considered only the possibility that the sun has retained primeval field. However, as well as primeval field, the sun may also generate its own field after it has settled down into its existence as a normal star. Field generation can be achieved by at least two mechanisms: (i) a battery effect in which field energy is derived from thermal energy in the electron gas; this mechanism encounters great difficulties \(59\) and will not be considered further here; (ii) a self-exciting dynamo, in which fluid motions are such that fields created by electric currents in the gas provide the correct forces to sustain these currents against ohmic dissipation. The problems of dynamo models were first discussed in connection with the Earth's magnetic field by Elsasser \(60\), and a heuristic physical solution for the Earth's field and for the solar field were obtained by Parker \(61\). (See also a Review by Parker \(62\).) In order to excite a dynamo, it is sufficient to satisfy two conditions: (a) non-uniform rotation,
which creates toroidal field from an initial poloidal field; (b) motions of field lines along the local vertical direction; these motions are influenced by coriolis forces in such a way as to regenerate poloidal field from the toroidal component. We will consider (a) and (b) in turn.

(a) The angular velocity of rotation must vary as a function of depth inside the sun, or as a function of latitude. In fact, the solar surface does indeed rotate at different rates at different latitudes. The equatorial regions rotate in about 25 days, while at latitudes of 70°, the rotation period is about 10 days longer than at the equator. The slower rotation of the poles is probably related to the fact that the magnetic field lines emerging from the poles are almost always open, extending out into interplanetary space, whereas many field lines in the equatorial regions are closed. The open field lines at the poles allow the solar wind to escape most freely there, and since it is the escape of the solar wind which is eventually responsible for slowing down the sun (63), the slowing down is expected to be most effective at the poles (64).

Much less certain is the question of the variation of angular velocity with depth beneath the surface of the sun. A radial gradient of angular velocity would be much more efficient as a dynamo energizer than a latitudinal gradient (65), and such a gradient would arise if the magnetic torque applied by the solar wind were to affect only the outer layers of the sun, leaving the inner regions to rotate at a faster rate. Possible evidence for the existence of a rapidly rotating solar core has been discussed by several
authors (57,64-67), although there are theoretical difficulties in understanding how a fast core could persist in equilibrium with a slowly rotating outer convection zone (56). As far as solar neutrinos are concerned, it appears that a fast core might help in explaining why the neutrino flux is observed to be so small (68).

However, whether or not there is a radial gradient of rotation, there is at least differential rotation at the surface as a function of latitude. Hence, if this latitudinal shear persists at great depths below the surface (and Iroshnikov (44) has shown that the shear is preserved at least to the bottom of the convection zone), then the first requirement for a solar dynamo is satisfied. In fact, Babcock (69) achieved remarkable success in explaining many of the observed features of the solar cycle by assuming only latitudinal shear. Conversely, the Babcock mechanism cannot amplify a rigidly-rotating field pattern such as the large-scale background field which is responsible for interplanetary sector structure. According to the Babcock model, such a field would simply decay monotonically with time (35), and therefore should be considered as a truly primeval field.

(b) The second requirement for a self-exciting dynamo is the presence of vertical fluid motions. The ratio of the relative strength of these motions to the non-uniform rotation determines the ratio of poloidal to toroidal field strengths in the dynamo (62). Are there suitable gas motions in the sun? At first, it seemed that an obvious source of such motions would simply be the hydrodynamic convection, and models of the solar dynamo were computed on
the basis of detailed numerical models of the solar convection
zone (70). One of the difficulties with this approach is that
the evolution of turbulent magnetic fields is such an extremely
complex subject that it is not possible to rely on simple arguments
such as equipartition of energy, or on kinematic considerations
such as the analogy between field strength and vorticity (71).
Contrary to the simplest expectations of vorticity theory, it ap-
ppears that in evolving non-equilibrium hydromagnetic turbulence,
magnetic energy cascades both to larger and to smaller wave-numbers.
This is reminiscent of similarly unexpected cascading of energy to
small wave-numbers in hydrodynamic turbulence in the Earth's at-
mosphere and oceans (72,73). Cascading of magnetic energy to small
wave-numbers is of crucial importance in regenerating large mag-
netic loops, and thereby sustaining the dynamo (71).

A further difficulty which arises if the convection zone is
assumed to drive the dynamo is that the structure of the solar con-
vection zone is currently one of the least well known aspects of
solar structure. This is so because the sun unfortunately lies on
a part of the main sequence where the depth of the convection zone
is especially sensitive to the model which is used to describe
convective heat transfer (18). Thus, different models of convective
transfer lead to convection zone models which differ in depth
by factors of more than $10^{14}$. If the convection zone turns out
to be much shallower than the usual model predicts, then dynamo
generation in the convection zone will be impossible. For example,
in the author's model of the solar convection zone (14), the dynamo
number (74) has the value 0.07, which is four orders of magnitude too small to allow a self-excited dynamo (74).

It is therefore necessary to consider sources of vertical gas motion other than convection. One source which has already been incorporated into semi-quantitative models of the solar magnetic cycle is magnetic buoyancy of flux tubes (57,62,69). A flux tube is expected to be buoyant because the field contributes to the total pressure inside the tube. Hence if the tube is to be in pressure equilibrium with outside gas, gas pressure inside the tube must be lower than gas pressure outside. Unless the temperature inside the flux tube is significantly lower than outside, the gas density inside the flux tube must be lower than that outside, and so the flux tube is buoyant. The buoyancy is sufficient to bring the flux tube up towards the surface of the sun if the length of the flux tube exceeds a critical value (75), or if the field strength inside the tube exceeds a critical value (57).

Another source of vertical motion which has not yet been explored quantitatively concerns planetary tidal influences on the sun. Tidal amplitudes at the solar surface are ±600 km (76,77), and although it is not clear what effects tides might have on magnetic flux tubes, it is possible that the tides might trigger instabilities which would help to raise the flux tubes towards the surface. The tides would then be only indirectly responsible for energizing the dynamo. Tidal periodicities appear to have an important effect in forcing the sun to vary magnetically with a period close to 11 years. And tidal effects seem to be correlated
with both the mean area of sunspots and the frequency of eruption of new active regions (see below). It is also interesting that flare activity in a nearby binary star system is correlated with tidal action (18).

Both magnetic buoyancy and tidal effects are effective throughout almost the entire volume of the sun, so that the solar magnetic field need not be confined to the convection zone. It has already been noted above that the rigid rotation of the large-scale background field suggests that at least this component of the field originates in gas beneath the convection zone. In fact, the magnetic field may not be allowed to exist in a steady state in the convection zone at all, on account of rapid turbulent dissipation in the supergranule layer (78). This leads to the conclusion that the solar field must on the average be confined to great depths below the surface, below the supergranule layer (78), with only sporadic field emergence into the supergranule layer. Then if the convection zone does indeed extend downwards only as far as the base of the supergranule layer (14, 44), the convection zone does not contribute significantly to the solar dynamo.

IV. The Cycle of Solar Magnetic Activity

Field generation is only the first step in explaining the solar cycle. The second step is to have the field vary with a double cycle in 22 years. Models which have been proposed to explain the cycle begin with a poloidal field which is sheared by differential rotation to give rise to a toroidal field. As time
progresses during a cycle, the toroidal field becomes increasingly sheared, forming flux ropes of greater and greater field strength, with the field lines aligned almost east-west. The energy appearing as magnetic energy is thought to be derived from the differential rotation, although it is important to realize that there is at present no observational evidence to suggest that as the field energy grows, the energy in latitudinal differential rotation decreases. As far as is known at present, the differential rotation of the sun remains the same at all times during the cycle, although the observations are sufficiently difficult to make that the possibility cannot be excluded that the differential rotation does indeed vary during the solar cycle.

The toroidal field grows in strength until something causes it to rise to the surface. The cause of the rise may be either magnetic buoyancy when the flux tube length of the field strength exceeds critical values (57,75), or instability due to loop formation, when the twisting of the flux rope has reached a critical value. (For a description of loop instability, see (79).) Whatever the cause, a rope of toroidal flux arches upwards, breaking through the surface of the sun as an active region, with the characteristic pattern of opposite magnetic polarities aligned more or less in an east-west direction. If fields in the flux ropes are large enough, spots form in the active region. If there are spots, the spots form at the feet of the arches, and the plages will lie underneath the arches (48). An active region is therefore to be considered as the surface manifestation of a flux rope, with indi-
individual strands of the rope possibly to be identified with the clumps of magnetic field (magnetic "pins") discovered by Stenflo (45), or else to be identified with plage granules (48). Sunspots appear if the flux rope includes especially strong fields spread over sufficiently large areas to affect the convective circulation within granules. The clumps of magnetic field which are observed to stream horizontally outward from sunspots (50) are presumably strands of flux torn off as a result of the fraying of the flux ropes (80). The rope is frayed by the eroding action of the surrounding convection zone, which erodes normal convection cells (supergranules) in times of order one day. It is a measure of the cohesiveness of a magnetic flux rope that a spot can survive the effects of this erosion for a period which is 10-100 times longer than a non-magnetic cell can survive.

The fact that spots generally appear in configurations which have an east-west orientation is considered to be evidence that a toroidal field has been formed by stretching a poloidal field. Then the reversal of the polarity in a pair of spots every eleven years requires that the toroidal field must change sign periodically. One way to make this happen is to change the sign of the poloidal field. This occurs in the models of Babcock (69) and of Leighton (57). In these models, the reversal of the poloidal field is accomplished by insisting on just the right amount of cyclonic twisting in the gas motions which raise the toroidal flux tubes upwards towards the surface, so that rising tubes of toroidal flux arrive at the surface with a component of poloidal field. Babcock and
Leighton succeeded in explaining a great many of the observed properties of the sunspot cycle. Their models are however not entirely free from criticism (81), especially at times when both rotation poles of the sun have the same magnetic polarity (8). At such times, it is not clear what the sign of the poloidal field is.

On the other hand, if the poloidal field of the sun is truly a primeval field, then there can be no question of reversing the polarity. Instead, reversals of the toroidal field must depend on reversing the direction in which the differential rotation stretches the poloidal field (82). Consider, for example, the case that solar rotation varies with depth inside the sun. Then in steady state the solar magnetic field lines must lie on surfaces of constant rotation (Ferraro's law of iso-rotation (93)). If a mechanism can be found which will tilt the field lines away from the surfaces of constant rotation, stretching of the field lines will result, and the sign of the stretched toroidal field will depend on the direction of tilt of the field lines. Piddington (82,83) suggests that the tilt varies from one eleven-year interval to the next as a result of a hydromagnetic oscillation, for it appears that the period of free oscillations of the solar field might be about 22 years (see below). The general solar field which is involved in the oscillation is found to be quite weak. The estimate of the field strength begins by noting that sunspot motions have a 22-year period, during half of which they drift towards the equator, and towards the pole during the other half (84,85). The velocity amplitude of the oscillation is observed to be 60 cm/sec. Repeating Piddington's numerical
estimates (Eq. (3) in Ref. 83), but inserting the gas density at the base of the author's model of the solar convection zone-(14) ($\rho = 3 \times 10^{-6} \text{ gm/cm}^3$ at a depth of $10^4 \text{ km}$), we find that the amplified meridional field strength $B_g$ has the value 0.4 gauss. The amplification of the meridional field arises because meridional motions (as deduced from spot motions) have amplitudes of $d = 6 \times 10^4 \text{ km}$ (83), so that if the general radial field $B_g$ is rooted in gas at depth $D$, the amplification factor for meridional field, $K = B_g/B_g$ is approximately $K = d/(d^2+D^2)^{1/2}$. If $K$ is of order 1, the general solar field $B_g$ is about 0.4 gauss, and is therefore too weak to be detected by current magnetographs, although future improvements in sensitivity will be of crucial importance in looking for such a weak general field. The 1-2 gauss fields currently observed at the poles of the sun would then have no direct relation whatever with the "general" solar field, according to this model. Such fields observed at the poles are merely the surface remnants of old active regions transported poleward from the sunspot latitudes by random walk of fields lines diffusing through the supergranule cell network (10). The fields at the poles can therefore be expected to have polarities which are determined by the history of poleward migration of active regions during the preceding 5-10 years. Statistical fluctuations in the numbers, areas, and field strengths of these active regions can be assumed to be responsible for the fact that the magnetic polarity of the polar caps does not vary in a strictly regular way. In particular, both polar caps can have the same polarity at the same time.
Repeating Piddington's estimate of the period of the hydro-magnetic oscillation, we use Eq. (6) in Ref. (83):

\[ P = 4 \pi^{3/2} \left( \rho \frac{d H}{B_r B_\theta} \right)^{1/2} \]

where \( B_r \) and \( B_\theta \) are radial and meridional fields, \( d = 6 \times 10^4 \) km is the amplitude of the oscillation, and \( H \) is the scale height of density in the region threaded by the general field. If this region lies at depths of \( 10^4-10^5 \) km, where temperatures are in the range \( 10^5-10^6 \) K, we find \( H = 5-50 \) thousand km. Then if we set \( B_r = B_\theta \), \( B_\theta = KB_g \), and \( \rho = 3 \times 10^{-6} - 3 \times 10^{-5} \) gm/cm\(^3\), we find \( P = (5-50)K^{1/2} \) years. If \( K \) is of order unity, the mean period of oscillation is between 16 years (geometric mean) and 27 years (arithmetic mean). These limits are satisfactorily close to the observed period of 22 years.

Piddington's model can be summarized as follows. A weak poloidal field (\( \leq 0.4 \) gauss, too weak to be currently detected) is sheared first of all by a meridional oscillation with a period of 22 years. The meridional field thereby formed does not in general lie on the surfaces of constant rotation, and is therefore sheared by radial differential rotation into a toroidal field. The sign of the latter is determined by the sign of the meridional field, which in turn reverses every eleven years. The toroidal field therefore also reverses every eleven years. The model depends on the assumption that the hydromagnetic approximation is valid. In view of the temporal variations of the large scale field, and in
view of the sector structure of the field, the validity of this assumption has been questioned (86). The dynamo field may be non-stationary, oscillating with complex multiple periodicities.

The question now arises: what excites the 22-year oscillation? A statistical analysis of sunspot numbers (87) suggests that there are actually two periodicities in the sunspot cycle. One is a free-running time for the solar cycle, 11.8 years, and the second is a "well-defined periodicity which excites the new solar cycle", 10.45 years. Piddington (83) also believes that as well as a free-running oscillation (period 22 years), there is a forced oscillation with a period of about 17 years. Since the statistical analysis (87) ignores polarity reversals, it is necessary to double the period of the free-running time determined from that analysis in order to compare with Piddington's model. The solar cycle should then be considered to have a free-running period of 22-23.6 years, combined with a forced oscillation with a period of 10.45-11 years.

In explaining the forcing term, it is tempting to consider the correlation between sunspot activity and planetary tidal influences, especially due to Venus, Earth, and Jupiter. The correlation between tidal periodicities (mainly, but not entirely, tied to the orbital period of Jupiter, 11.86 years) and sunspot activity is so high that the probability of chance occurrence is only one in 1230 (77). Trellis (88,89) has shown that planetary tides on the sun are well correlated with the area of sunspots and with the rate of formation of active regions. Moreover, active regions and sunspots tend to occur in regions separated by 180 degrees of solar longitude (90-92), an effect which is typical of the behavior expected
if tidal influences are important.

V. Conclusion

Our understanding of solar magnetic fields is changing, both observationally and theoretically. The large scale dipole solar field is probably quite weak (<0.4 gauss), and therefore not directly related to any currently measured fields. The most accurate description of the sun now seems to be that the sun is a magnetic pin-cushion, with strong fields (>2000 gauss) clumped into small regions (100-300 km diameter). These clumps occur all over the surface of the sun, but are more densely packed in active regions, where large magnetic flux ropes have erupted through the surface from the interior of the sun. Sunspots are magnetic flux ropes where both the field strength and the diameter are sufficiently large to allow the field to interfere with local hydrodynamic convection cells, and thereby impede the normal convective transfer of energy upwards towards the surface. The reduction in convective efficiency in spots is compensated by the transfer of energy in an alternate form, such as Alfvén waves. Outside spots, clumps of magnetic field are observed to stream horizontally away from the spot, and these clumps are probably individual strands of field torn away from the flux rope by the erosion of the surrounding convection zone. The solar convection zone, which was at one time thought to play an essential role in solar magnetic activity, may after all not be responsible for generating the field, nor for reversing it every eleven years. This is a welcome improvement, for our knowledge of
the solar convection zone is one of the least well known aspects of solar structure at the present time. Deeper layers in the sun, below the convection zone, can now be considered as the site of dynamo action (if the observed fields are not primeval), or as the site of a primeval field which is subject to hydromagnetic oscillations having a period of about 22 years. The role of planetary tides in the solar cycle of activity is not yet clear, but the correlation between tides and solar activity is high enough that the correlation is unlikely to be due to chance.

However, although the convection zone may no longer be considered to play a dominant role in solar activity, this is not to say that its role is now negligible. Far from it. The convection zone in conjunction with local magnetic fields at the surface of the sun, probably provides the mechanical energy which is responsible for heating the corona. This is the first step in the process which eventually drags the solar magnetic field away from the sun, out into the interplanetary medium, and eventually into interstellar space. Then if the solar field is indeed a primeval field, the convection zone is instrumental in bringing the field full circle from its condition in interstellar space long before the sun formed 5 billion years ago. Out of interstellar space the solar field originally came, and back to interstellar space the solar field eventually goes.

Acknowledgment

This work has been supported by NASA Contract NGR-39-005-066.
References


(10) R.B. Leighton, "Transport of Magnetic Fields on the Sun", 

(11) D.L. Schatz, "Evolution of the Magnetic Field Configuration 
in an Active Region" in Solar Magnetic Fields (Ed. R. Howard) 

(12) N.V. Steshenko, "Magnetic Fields of Small Sunspots and Pores", 


(14) D.J. Mullan, "Cellular Convection in Model Stellar Envelopes", 

(15) D.E. Osterbrock, "The Heating of the Solar Chromosphere, 
Plages, and Corona by Hydromagnetic Waves", Astrophys. J. 


(17) D.J. Mullan, "Sunspot Models with Alfvén Wave Emission", 


(19) C. de Jager, "Solar Flares and Space Research", (Ed. C. de 
dam, 1969.

(20) R. Michard, "Solar Magnetic Fields in Association with Flares", 
in Solar Magnetic Fields, (Ed. R. Howard) p. 359, Reidel 


Legend for Figure

Figure 1  The ordinate is the continuum intensity of features on the surface of the sun, expressed in units of the solar intensity in a region which is free of magnetic fields. The abscissa is field strength along the line-of-sight; at the center of the solar disk, this corresponds to the vertical field component. Facular data are taken from Frazier (94). The transition from brighter-than-normal to darker-than-normal occurs at 700 gauss (13). Sunspots are not observed if B is smaller than 1200-1400 gauss (12).
Presence of a magnetic field causes both heating and cooling.

BRIGHTER THAN NORMAL  DARKER THAN NORMAL

FACULAE

SUNSPOTS

INTENSITY (Field-free Sun = 1.0)

LONGITUDINAL FIELD STRENGTH

B=700

B=1200