THE ORIGIN OF THE SOLAR CYCLE

R. ROSNER
Department of Astronomy and Astrophysics, University of Chicago, 5640 S. Ellis Ave., Chicago, IL 60637

N. O. WEISS
Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Silver St, Cambridge CB3 9EW, UK

ABSTRACT We review the progress of hydromagnetic dynamo theory as it applies to the Sun. Recent progress in a variety of aspects of dynamo theory – from new ideas concerning the nature of turbulent diffusion to more powerful methods of numerical simulation as well as striking new observations of the rotation rate in the solar interior – has led to major advances in our understanding of how magnetic fields are generated in the interior of stars. Nevertheless, the solar dynamo problem remains unsolved: We still do not possess a deductive theory of how the Sun produces the magnetic fields that we observe at its surface.

1. INTRODUCTION

At the first Solar Cycle Workshop this topic was reviewed by Parker (1987a). Since then, helioseismology has revealed the internal rotation of the sun and the subject has consequently been transformed. The new picture has already been discussed in a number of reviews (e.g. Weiss 1989, Belvedere 1990, Hoyng 1990, Brandenburg and Tuominen 1991, Stix 1991), so there is no need to produce another exhaustive account of the processes that maintain the solar cycle. The available evidence points to a dynamo mechanism but the case is not yet cut and dried. Our aim here is to provide a critical survey that exposes the strengths and weaknesses of dynamo theory, and considers alternative oscillator mechanisms.

To set the scene we first summarize the relevant observational evidence. In the next section we introduce the kinematic dynamo problem; then, in Section 3, we point out the limitations of this approach, before discussing oscillators in Section 4. Recent work on nonlinear dynamos is described in Section 5 and we comment on current research in Section 6. In conclusion we put forward a self-consistent picture of the solar dynamo that is compatible with present observational and theoretical knowledge.

We begin by exploiting the solar-stellar connection and considering stars with differing properties in order to learn more about the sun. There are two important groups of magnetic stars. The first are the so-called Ap stars,
Fig. 1. Rotational evolution of the Sun: Sketch showing the variation with age of the ratio of the equatorial angular velocity $\Omega$ to its present value $\Omega_0$ (after Rosner and Weiss 1985). Rapid spin-up as the Sun collapsed on to the main sequence is followed by spindown owing to magnetic braking at an ever-decreasing rate.

with strong fields modulated only by rotation. The second, which concern us, are late-type stars with deep outer convection zones. Cool stars on the main sequence display irregular magnetic activity (as described by Saar and Radick in these Proceedings), and this activity can be monitored by measuring chromospheric Ca$^+$ emission (Baliunas and Vaughan 1985, Hartmann and Noyes 1987, Noyes, Baliunas and Guinan 1991) which is correlated with magnetic flux on the sun. The faster a star of given mass rotates the more active it is; the Ca$^+$ emission depends on the inverse Rossby number $\sigma = \Omega \tau_c$, where $\Omega$ is the angular velocity and $\tau_c$ the convective turnover time (Noyes et al. 1984a). Figure 1 shows the rotational history of a G-star like the sun (Rosner and Weiss 1985, Weiss 1989). $\Omega$ initially increased as the sun collapsed on to the main sequence, conserving its angular momentum; young G-stars are very active but rapidly spin down owing to magnetic braking (Mestel and Spruit 1987, Kawaler 1988) and consequently grow less active. The rate of spindown decreases monotonically with time and there is no evidence that very old stars are conspicuously less active than the sun (Baliunas and Jastrow 1990). Cyclic activity has only been detected in middle-aged stars like the sun, which are comparatively slow rotators; although the available sample is small it suggests that the cycle frequency increases with increasing $\sigma$ (Noyes, Weiss
and Vaughan 1984b, Noyes et al. 1991). We are concerned with the mechanism responsible for this behavior: it is by no means clear that the same process operates in more rapidly rotating and more active stars, where Coriolis and Lorentz forces may alter the pattern of convection (Knobloch, Rosner and Weiss 1981, Jones, Roberts and Galloway 1990).

Cyclic activity has been studied in detail on the sun. The familiar butterfly diagram shows both the irregular 11-year cycle (corresponding to a 22-year magnetic cycle) and the equatorward drift of active regions. This sequence is punctuated by grand minima, when activity is drastically reduced. The dearth of sunspots during the Maunder minimum in the 17th Century was established by direct observation; proxy evidence comes from fluctuations in the abundance of $^{14}$C, in trees, and $^{10}$Be, in ice cores, since the galactic cosmic rays that lead to the formation of these isotopes are modulated by magnetic fields in the solar wind. Other possible proxy records (Cini Castagnoli et al. 1990, Dermendjiev, Shopov and Buyukliev 1990) require further investigation. Figure 2 confirms that the variations in $^{14}$C and $^{10}$Be abundances are in good agreement over the last 1200 years (Raisbeck et al. 1990); the $^{14}$C record has been carried back for 9000 years (Stuiver and Braziunas 1988) and it shows that the envelope of solar activity is modulated aperiodically with a characteristic timescale of around 200 yr. In the Greenland ice cap the rate of deposition is so fast that annual layers can be identified and the 11-year cycle appears in the $^{10}$Be record. It is important that a full sequence of detailed $^{10}$Be abundances should be measured in the Milcent ice-core in order

Fig. 2. Variations in $^{10}$Be and $^{14}$C abundances over the past 1200 years (after Raisbeck et al. 1990). Note the lag between the two curves. The rates of production of these isotopes are anticorrelated with magnetic activity; the arrow indicates the Maunder minimum.
to determine whether phase is preserved through grand minima or not. The overall pattern of activity interrupted by grand minima is consistent with the range of $^{14}$C emission in other cool stars with ages similar to or greater than the sun's (Baliunas and Jastrow 1990). The statistics suggest that about one-third of these stars are inactive at any epoch, while their range of activity is comparable to that of the sun.

If the sun's magnetic field did not reverse it would be easy to explain its origin. The timescale for resistive decay of a stellar magnetic field is comparable with its lifetime on the main sequence, so there is no problem in regarding the fields in Ap stars as fossil remnants, gradually decaying. (By contrast, the earth's magnetic field, with a decay time of only $10^4$ yr, requires a dynamo to maintain it.) Some dynamical process is, however, required in order to explain the solar cycle with its short timescale. The only candidates are an oscillator (in which kinetic energy is transformed into magnetic energy without significant dissipation) or a dynamo (where ohmic dissipation is balanced by induction).

2. THE KINEMATIC DYNAMO PROBLEM

Although dynamo action seems to be common in nature it raises unusual mathematical difficulties, which have given dynamo theory its special flavor. These difficulties appear already in the linear or kinematic dynamo problem. The aim here is to solve the induction equation

$$\frac{\partial \mathbf{B}}{\partial t} = \nabla \times (\mathbf{u} \times \mathbf{B}) + \eta \nabla^2 \mathbf{B}$$

(2.1)

for some prescribed velocity $\mathbf{u}$ that allows the magnetic field $\mathbf{B}$ to grow exponentially despite the effects of the magnetic diffusivity $\eta$. The simplest configuration is axisymmetric: then we can split the field into poloidal and toroidal parts such that

$$\mathbf{B} = \mathbf{B}_P + \mathbf{B}_T, \quad \mathbf{B}_P \cdot \mathbf{\hat{\phi}} = 0, \quad \mathbf{B}_T = \mathbf{B}_\phi \mathbf{\hat{\phi}},$$

(2.2)

referred to cylindrical polar co-ordinates $(\theta, \phi, z)$. Since $\nabla \cdot \mathbf{B}_P = 0$, we can write $\mathbf{B}_P = \nabla \times (A \mathbf{\hat{\phi}})$ and integrate (2.1) to give

$$\frac{\partial A}{\partial t} = \mathbf{u} \times \mathbf{B} \cdot \mathbf{\hat{\phi}} - \mu_0 \eta j_\phi.$$  

(2.3)

It follows that the toroidal current $j_\phi$ cannot be maintained by induction at a neutral point, where $\mathbf{B}_P = 0$ (Cowling 1934, 1976, Hide and Palmer 1982). Cowling's theorem shows that the field must be non-axisymmetric. The problem then is to construct a velocity $\mathbf{u}$ that is sufficiently complicated to allow a magnetic instability to develop. Many examples of such velocity fields are known.

The large-scale magnetic field at the solar surface is predominantly toroidal. This is a natural consequence of differential rotation (the $\omega$-effect).
Consider, therefore, a differentially rotating star with a toroidal velocity \( U = s\Omega(s,z)\phi \). Then an axisymmetric magnetic field satisfies the equations
\[
\frac{\partial A}{\partial t} = \eta(\nabla^2 - \frac{1}{s^2})A, \\
\frac{\partial B_\phi}{\partial t} = s\mathbf{B}_\mathbf{P} \cdot \nabla \Omega + \eta(\nabla^2 - \frac{1}{s^2})B_\phi. \tag{2.4}
\]

Differential rotation generates toroidal flux from a poloidal field but there is no corresponding source term for poloidal flux. To produce a dynamo we introduce a small-scale turbulent velocity \( \mathbf{u}(r,t) \). Rather than follow all the convective eddies and small-scale magnetic structures it is convenient to study the evolution of the mean field, which is typically axisymmetric (Moffatt 1978, Parker 1979, Krause and Rädler 1980, Zel’dovich, Ruzmaikin and Sokoloff 1983). Thus we split the total field \( \mathbf{B}' \) into azimuthally averaged and fluctuating parts:
\[
\mathbf{B}' = \mathbf{B} + \mathbf{b}, \quad \langle \mathbf{B}' \rangle = \mathbf{B}, \quad \langle \mathbf{b} \rangle = 0. \tag{2.5}
\]

Then the mean field satisfies
\[
\frac{\partial \mathbf{B}}{\partial t} = \nabla \times (\mathbf{U} \times \mathbf{B}) + \nabla \times \mathbf{E} + \eta \nabla^2 \mathbf{B}, \quad \mathbf{E} = \langle \mathbf{u} \times \mathbf{b} \rangle. \tag{2.6}
\]

If there is a separation of spatial scales we can write
\[
\mathbf{E}_i = \alpha_{ij} B_j + \beta_{ijk} \frac{\partial B_k}{\partial x_j} + \ldots, \tag{2.7}
\]
and for pseudo-isotropic turbulence we have
\[
\alpha_{ij} = \alpha \delta_{ij}, \quad \beta_{ijk} = -\beta \epsilon_{ijk}, \quad \mathbf{E} = \alpha \mathbf{B} - \beta \nabla \times \mathbf{B}, \tag{2.8}
\]
so that
\[
\frac{\partial \mathbf{B}}{\partial t} = \nabla \times (\mathbf{U} \times \mathbf{B}) + \nabla \times (\alpha \mathbf{B}) + \eta \nabla^2 \mathbf{B}, \tag{2.9}
\]
where the total diffusivity \( \tilde{\eta} = \eta + \beta \). Hence we obtain the mean field dynamo equations
\[
\frac{\partial A}{\partial t} = \alpha B_\phi + \tilde{\eta}(\nabla^2 - \frac{1}{s^2})A, \\
\frac{\partial B_\phi}{\partial t} = s\mathbf{B}_\mathbf{P} \cdot \nabla \Omega + \frac{\phi}{\tilde{\eta}} \nabla \times (\alpha \mathbf{B}_\mathbf{P}) + \tilde{\eta}(\nabla^2 - \frac{1}{s^2})B_\phi. \tag{2.10}
\]
Comparing (2.10) with (2.4) we see that there is now a source-term (the \( \alpha \)-effect) for the poloidal field.

Mean field dynamos come in different varieties. In the absence of differential rotation both \( \mathbf{B}_\mathbf{P} \) and \( \mathbf{B}_\mathbf{T} \) can be maintained by the \( \alpha \)-effect, in \( \alpha^2 \)-dynamos. Adding differential rotation gives \( \alpha^2\omega \)-dynamos but if generation of toroidal flux is dominated by the \( \omega \)-effect we may neglect the \( \alpha \)-term in the equation for \( B_\phi \) to obtain an \( \alpha \omega \)-dynamo. First-order smoothing (which
corresponds to a separation of scales in time such that $|b| \ll |B|$ allows us to calculate the coefficients $\alpha$ and $\beta$. Typically we find that

$$\alpha = -\frac{1}{3} \tau_c \langle h \rangle, \quad \beta \approx ul \gg \eta,$$

(2.11)

where the helicity $h = u \cdot \nabla \times u$ and $l$ is the scale of the turbulent eddies. Thus dynamo action depends on a mean helicity, corresponding to cyclonic eddies produced by the action of a Coriolis force.

There are many examples of oscillatory $\alpha\omega$-dynamos in the literature. A typical eigenfunction of the linear problem – see, for example, Stix (1989, Fig. 8.38) – shows dynamo waves propagating towards the equator in accordance with the observed behavior of the solar magnetic field. The toroidal field is antisymmetric about the equator and there are two regions of opposite polarity in each hemisphere. So the time taken for toroidal flux of one polarity to migrate from the pole to the equator is approximately the full period of the magnetic cycle, i.e. twice the period of the activity cycle: Extended cycles appear as a characteristic feature of simple dynamos.

Mean field dynamo theory can be rigorously justified – but only for circumstances that do not obtain in the sun. So we should regard it as a reliable parametrized theory that captures the essential physics of a turbulent stellar dynamo. Indeed, Parker (1955) first justified the $\alpha$-effect by physical arguments; moreover, the different parameters ($\alpha, \beta$ and $\omega$, the rotational shear) can be related to properties of detailed models. In the dynamo equations (2.10) the crucial parameter is the dimensionless dynamo number, $D = \alpha \omega L^2/\eta^2$, where $\omega$ is a measure of $\nabla \Omega$ and $L$ is a characteristic length scale for the large-scale field. For turbulent convection we expect that $\alpha \approx \Omega l, \eta \approx ul$, from (2.11), while $\omega \approx \Omega/l$ so that

$$D \approx \Omega^2 \tau_c^2 \left( \frac{L}{l} \right)^2 = \left( \frac{L}{l} \right)^2 \sigma^2.$$

(2.12)

Hence the stability parameter $D$ depends on the inverse Rossby number $\sigma$ (Durney and Latour 1978) which, as we have seen, determines the observed level of activity in late-type stars.

3. LIMITATIONS OF DYNAMO THEORY

The above discussion clearly indicates that from a mathematical perspective, the existence of dynamos is no longer in question. There remains, however, the question of whether such dynamos are actually realized in nature.

From this perspective, the problem is thus not so much whether one can construct equations that describe a dynamo process, but rather whether the mathematical assumptions which underlie such dynamos are in fact physically justified. For example, it is obvious that the sun’s magnetic field cannot be regarded as simply passively advected by solar fluid motions – at least, not at the solar surface; this then calls into question the basic kinematic assumption which underlies the discussion in Section 2 above. This example illustrates the basic issue we will discuss in this section, namely the contrast between
classical solar dynamo theory, as discussed just above, and the actual physical circumstances in the solar interior. (The removal of the kinematic assumption will be discussed in detail in Section 5 below.)

The first issue revolves about the question of magnetic buoyancy, and the problem of where the magnetic dynamo process is actually located: It is well-known that magnetic fields are buoyant in a stratified medium, a physical effect which is generally not explicitly accounted for in the standard mean field dynamo models. Is this a difficulty? Piddington (1972) argued that magnetic buoyancy would lead to a progressive expulsion of magnetic flux from the solar convection zone, so that eventually even given that a dynamo were located in the convection zone proper all magnetic flux would leave the solar interior, and thus any extant dynamo would perforce shut off. This highly qualitative argument was quantified by Parker (1975), who showed that for mean field strengths $> 100$ G, the buoyancy rise rate in the solar convection zone may be sufficiently large that the $\alpha$-effect is suppressed; Parker argued that, in order to maintain the dynamo process, the dynamo region had to be placed sufficiently deeply in the solar convection zone that this suppression becomes irrelevant. (More recently, Parker has suggested that thermal shadows within the solar convection zone may act to reduce if not entirely eliminate the tendency of magnetic flux ropes to rise buoyantly in the convection zone.)

Now, it may be thought that this issue is readily dealt with by simply including magnetic buoyancy effects in the mean field dynamo models. Unfortunately, this is not the case. Early dynamo models, such as those of Leighton (1969), parametrized the buoyancy effect (indeed, buoyant eruption of magnetic fields was the essential nonlinear limiting mechanism for the dynamo process in Leighton’s models), while virtually all numerical simulations ignored the process altogether (see however below). Parametrizations of buoyancy are not really satisfactory for the present purpose because they may not properly describe the buoyancy process, especially in a convectively unstable fluid such as the solar convection zone. Indeed, the concerns about flux storage in the lower depths of the solar convection zone over the typical 11-year solar activity cycle, together with arguments about the large-scale coherent eruption of small spatial scale flux from the solar convection zone led Golub et al. (1981) to suggest that magnetic field generation was not located in the convection zone proper, but rather deeper, in the undershoot region separating the convectively-unstable layers from the underlying convectively stable, radiative interior.

This undershoot layer has the crucial property that it is stably stratified against convection. For this reason, magnetic buoyancy can be treated as an instability problem (where the basic equilibrium state is a slightly sub-adiabatic layer in which penetrative motions from the convection zone proper intrude). Acheson (1979), Schmitt and Rosner (1983), and Hughes (1985a,b) all investigated the linear instability properties of this layer, showing that multiply-diffusive instabilities might lead to the formation of flux bundles, reminiscent of discrete active region complexes. The nonlinear development of related instabilities has been followed in more recent calculations (Cattaneo and Hughes 1988, Cattaneo, Chiueh and Hughes 1990, Hughes 1991, 1992). How then would flux generation occur in this layer? Do the calculations for magnetic buoyancy in this layer give a reasonable account of the magnetic structures which ultimately form?
The difficulty just alluded to revolves around the question of whether the dynamo process has been properly located. At a more basic level, we may ask: Are the dynamo equations themselves properly justified? Consider, for example, the arguably most essential aspect of all dynamo models, turbulent diffusion. It has long been recognized that solar magnetic fields change on time scales that are entirely incompatible with ordinary "molecular" diffusion (cf. Cowling 1953). For this reason, appeal has long been made to turbulent diffusion, whose magnitude may be estimated from the statistical properties of the turbulent solar convection zone (cf. equation 2.11),

\[
\eta_t = \frac{\lambda^2}{2\tau}, \quad \text{(Leighton 1964)}
\]

\[
= 0.15(\delta v \cdot \delta v)^{1/2} \lambda, \quad \text{(Parker 1971)}
\]

\[
= \langle \delta v \cdot \delta v \rangle \tau / 3, \quad \text{(Roberts and Stix 1971)}
\]

\[
\approx 10^{13} \text{cm}^2 \text{s}^{-1},
\]

or roughly \(10^6\) times the ordinary molecular diffusivity; here \(\lambda\) is the typical scale of motions in the solar convection zone, \(\langle \delta v \cdot \delta v \rangle^{1/2}\) is the rms turbulent velocity, and \(\tau\) is the typical characteristic time scale of the turbulent motions. These estimates are clearly motivated by the expressions commonly derived for the eddy viscosity in a turbulent fluid (in Parker's words -- \(\mathbf{B}\) diffuses like smoke), but it is far from certain that they actually apply to the diffusion of a vector field such as the magnetic field.

Piddington (1972, 1973) argued vigorously against the viewpoint summarized by expressions (3.1). His fundamental objections rested on the notion that the small-scale magnetic fields in a turbulent flow may grow (due to stretching of field lines by turbulent eddies) faster than they can be dissipated. His qualitative argument viewed the Lorentz backreaction as the only ultimate limiter on field growth on small scales; hence, he envisaged that fluid motions on small scales would ultimately be damped, with the consequence that at these small scales, fluid motions corresponding to a superposition of Alfvén wave-like modes -- "pseudo-turbulence" -- would arise.

Was Piddington correct? To be more concrete, consider once again equation (2.1), and the physical interpretation of the two terms on the right-hand side of this equation: The first term, \(\nabla \times (\mathbf{u} \times \mathbf{B})\), leads to the stretching, bending, and twisting of magnetic field lines, while the second term, \(\eta \nabla^2 \mathbf{B}\), leads to diffusion, and hence is the term which allows topological changes in \(\mathbf{B}\). In the limit of \(\eta = 0\), magnetic field lines are frozen into the fluid; for \(\mathbf{u} = 0\), the field simply decays on a time scale \(\approx L^2 / \eta\). The efficacy of turbulent diffusion is thus examined by asking to what extent the \(\nabla \times (\mathbf{u} \times \mathbf{B})\) term contributes to an "effective" diffusivity of the form \(\eta \nabla^2 \mathbf{B}\).

To be more precise, one can ask: First, supposing that \(\mathbf{B}\) is always kinematic, can one define an effective diffusivity which acts like an eddy diffusivity? Second, under what conditions does \(\mathbf{B}\) remain kinematic in a turbulent flow? Third, what happens when \(\mathbf{B}\) is not kinematic? It is straightforward to answer these questions in two dimensions: For an incompressible flow, equation (2.1) is then identical to the evolution equation for a scalar field (such as temperature) in a turbulent fluid,

\[
\frac{\partial \theta}{\partial t} + \mathbf{u} \cdot \nabla \theta = \kappa \nabla^2 \theta,
\]
where the gradient of the scalar field, $\nabla \theta$, obeys the same equation as does $B$. If $\kappa = 0$, then it is straightforward to understand what happens: Consider a compact bubble of "$\theta$-stuff" in an incompressible turbulent fluid. As time progresses, the "$\theta$ stuff" is shredded, while the volume occupied by the "$\theta$ stuff" remains fixed (since equation 3.2 conserves "$\theta$ stuff"),

$$\int dV \theta = \text{constant}. \tag{3.3}$$

This implies that on average, the gradient of $\theta$ must increase without limit. In contrast, the mean of $\theta$ approaches a constant,

$$\langle \theta \rangle (\mathbf{r}) \rightarrow \tilde{\theta}, \tag{3.4}$$

where $\tilde{\theta}$ is the value of $\theta$ obtained if all of the initial "$\theta$ stuff" had been uniformly distributed over the volume in question; similarly, the gradient of $\langle \theta \rangle (\mathbf{r})$ vanishes,

$$\nabla \langle \theta \rangle (\mathbf{r}) \rightarrow 0. \tag{3.5}$$

Recalling the identification of $B$ with $\nabla \theta$ in two dimensions, we see that indeed

$$\langle B \rangle (\mathbf{r}) \rightarrow \bar{B} \tag{3.6}$$

($\bar{B}$ is the volume average of $B$), i.e. the magnetic field indeed diffuses like smoke.

Now, all of the preceding presumed that the field was kinematic, i.e. that the backreaction of the magnetic field on the fluid may be disregarded throughout the field shredding process. This assumption may not apply. To see why, note that we can compute the scale $\delta_t$ of the smallest structures formed in the turbulent fluid by assuming that on this smallest scale, advection is balanced by (molecular) diffusion, $u \ell \approx \eta / \delta_t$, or

$$\delta_t \approx \ell R_m^{1/2}. \tag{3.7}$$

where $R_m (\equiv u \ell / \eta \gg 1)$ is the magnetic Reynolds number. Structures on this size should be formed on time scales comparable to the typical turnover time scale, $\ell / u$; this then allows us to simply estimate the strength of the corresponding magnetic field fluctuation on that scale,

$$\langle B^2 \rangle \approx R_m \langle |B_0|^2 \rangle, \tag{3.8}$$

where $B_0$ is the initial large-scale field. If we define the equipartition field,

$$B_e^2 \equiv 4\pi p u^2, \tag{3.9}$$

we can use equation (3.8) to simply compute the initial field strength consistent with the assumption that the field reaches equipartition strength on scales comparable to $\delta_t$,

$$\max |B_0| = R_m^{-1/2} B_e. \tag{3.10}$$

Thus we have our sought-for result: Only for large-scale initial fields such that

$$|B_0| < R_m^{-1/2} B_e \tag{3.11}$$
will the system remain kinematic for all time (cf. Vainshtein and Rosner 1991). For typical astrophysical values of the magnetic Reynolds number (e.g. $10^{12}$ for the sun), this critical field strength is orders of magnitude less than the actual observed large-scale mean fields.

What happens when $|B_0|$ is larger than this critical initial value? Vainshtein and Rosner (1991) argued that in this case, the cascade of magnetic energy to smaller scales becomes disrupted, and turbulent diffusion – as generally understood – no longer functions. Confirmation, and detailed analysis, of this case has been recently obtained by Cattaneo and Vainshtein (1991,1992), who show via numerical simulations of randomly forced motion in an incompressible, two-dimensional magnetized fluid that this critical initial field strength indeed exists, and that if it is exceeded, the turbulent cascade of magnetic field energy is indeed interrupted. Cattaneo and Vainshtein show in particular that magnetic field growth occurs initially on a dynamical time scale over a wide range of initial magnetic field strengths. As long as the initial field is weaker than the critical field strength of equation (3.10), this growth is rapidly overwhelmed by a cascade of magnetic energy to small scales, where it is thermalized eventually on the resistive diffusion scale; the estimated diffusion rates in this regime correspond to eddy diffusivities computed from the statistics of the ambient turbulent flow. Once the initial field strength reaches the critical value, the initial period of rapid growth is instead supplanted by a period of apparent stasis, in which the energy in the large-scale magnetic field components changes only very gradually. This gradual decay continues to occur until the large-scale field has dropped sufficiently in strength to allow the cascade to small scales to resume; once this happens, the fields decay once more at the eddy diffusion time scale. What is striking is the extremely small value of the critical field strength at which this change in behavior occurs – for an equipartition field strength of 500 Gauss, and $R_m \approx 10^{12}$, the critical field strength is only $5 \times 10^{-4}$ Gauss!

This result strikes at the heart of standard mean field dynamo theory, for it argues that turbulent diffusion of magnetic fields may not operate for any interesting values of the large-scale magnetic field – mean field dynamos would then not really exist in nature. We return to this point in section 6 below. Here we only caution that the results we quoted strictly apply only in two dimensions – no comparable results for three dimensions exist at present. Moreover, it is not yet clear how far these results carry over to convectively driven flows, where there is a correlation between the motion and the magnetic field that may lead to qualitatively different behavior (Nordlund et al. 1992). Nevertheless, there are important implications for any model of a stellar dynamo (Cattaneo, Hughes and Weiss 1991).

4. OSCILLATORS

The difficulties in constructing magnetic dynamos for stars and planets prior to 1955 led to considerable efforts to go around these problems by proposing alternatives to dynamos. Perhaps the most important alternative is the magnetic oscillator, first considered in detail by Walén (1949). Because Walén’s ideas are revived from time to time, it seems useful to describe the essence of these models, and to work out one illustrative example. As will be clear
in the following, oscillator models do not have any advantages over magnetic
dynamos, but instead have, if anything, far more profound difficulties.

The basic idea of magnetic oscillator models is to give up on the idea of
magnetic flux generation, and to view the observed oscillating large-scale solar
poloidal and toroidal field systems as a consequence of an oscillation in the
differential rotation rate of the solar interior. Thus, oscillators by their very
nature view the observed fields as a temporarily amplified version of primordial
fields.

In their original form, oscillator models were severely flawed by virtue of the fact that models simple enough to be solvable did not have a discrete
spectrum, but rather a continuous spectrum of oscillations (cf. Layzer, Krook
and Menzel 1955, Plumpton and Ferraro 1955): Adjacent oscillating shells
were weakly, if at all, coupled, and hence these shells oscillated relatively
independently. One solution to this problem is to include interactions with
the star's rotation, which restore a discrete spectrum (Cowling 1976); another
is to identify a preferred shell, whose motion will dominate the production of
toroidal flux (cf. Layzer, Rosner and Doyle 1979); in that case, the oscillator
period will be dominated by the period of this preferred shell. Layzer et al.
identified the interface between the convection zone and the radiative
interior as the location of this preferred shell.

In order to place our discussion on a concrete basis, consider the following
simple model for an oscillator, namely a (very unrealistic) cylindrical star, in
which a layer of specified thickness is set into oscillation. We re-write equation
(2.1) together with the corresponding momentum equation:

\[
\rho_m \frac{\partial \mathbf{u}}{\partial t} = \frac{1}{4\pi} (\nabla \times \mathbf{B}) \times \mathbf{B} - \nabla p,
\]

\[
\frac{\partial \mathbf{B}}{\partial t} = \eta \nabla^2 \mathbf{B} + \text{curl}(\mathbf{u} \times \mathbf{B}),
\]

(4.1)

Now, for the sake of simplicity (but not realism), we consider a cylindrical
geometry, with an interior rotation rate given by the relation

\[
\mathbf{u} = u(R, t) \hat{\phi} = Q(R) \cdot \Omega(t) \cdot R \hat{\phi}
\]

(4.2)

Substituting this expression into (4.1), we obtain two equations for the
differential rotation:

\[
4\pi \rho_m \frac{\partial \Omega}{\partial t} = -k^2 \Omega(t)
\]

\[
f(R) \frac{d^2 Q}{dR^2} + g(R) \frac{dQ}{dR} + k^2 Q = 0
\]

(4.3)

where \( f(R) \) and \( g(R) \) are to be prescribed, and depend upon the functional
form of the magnetic field components. These equations govern the functional
behavior of the rotation rate in both time and space. The most trivial case is
obtained if \( B_R = B_0 R_0 / R \); in that case, the dispersion relation reads

\[
\omega^2 = k^2 / 4\pi \rho_m,
\]

(4.4)
with
\[ k^2 = \frac{1}{4} \left[ \frac{\pi^2}{\Delta} - 1 \right] \frac{B_0}{R_0}; \] (4.5)

here \( \Delta \) is the thickness of the layer in which the oscillator is located, in units of the solar radius. As an illustration, consider the following values for the physical parameters in the region of interest: \( B_0 = 1 \text{ G}, \Delta = 10^{-4}, \rho_m = 10 \text{ gm cm}^{-3}, R_0 = 5 \times 10^{10} \text{ cm} \) (i.e. the base of the convection zone for the sun). In that case, equation (4.4) gives a period for the oscillator of approximately 7.5 years; given our ignorance of the appropriate values for \( \Delta \) and \( B_0 \), this result is well within the realm of plausibility.

What then is the problem with the oscillator models? The most obvious difficulty is the matter of energetics: How is one to maintain the oscillations? No oscillator model to date has addressed this issue satisfactorily. Next, it is not trivial to identify the oscillating layer – here it was done by fiat; in the more general case, it is not known whether an appropriately oscillating layer would exist. Finally, there is the matter of observations. There is no evidence whatever in the surface observations to support the existence of interior oscillations of the kind discussed here: The torsional oscillations have a period of 11 years and not 22 years, and the poloidal field reverses at sunspot minimum. On the other hand, there are apparent variations in \( \Omega \) within the radiative zone (Goode and Dziembowski 1991) that have been interpreted as evidence for a 22 yr oscillation that might control the solar cycle (Goode and Gough 1992); what their relevance to the dynamo/oscillator problem is remains to be established.

5. NONLINEAR DYNAMOS

In kinematic dynamo models the magnetic energy can grow without limit but, in reality, the growth of the field is saturated owing to nonlinear effects. The Lorentz force becomes significant in the equation of motion and modifies the velocity. This process may be described either by solving the full partial differential equations in a numerical simulation or by constructing simplified models. In the spirit of mean-field dynamo theory we can isolate several effects of the Lorentz force. First, it may suppress helicity and so quench the \( \alpha \)-effect. Secondly, it may act so as to reduce or alter the differential rotation. Such a process can be represented by an equation of the form
\[ \frac{\partial v}{\partial t} = \frac{1}{\mu_0 \rho} \left[ (\nabla \times B_T) \times B_P \right] \nabla \phi + \nu (\nabla^2 - \frac{1}{s^2}) v, \] (5.1)

where \( v \) is the toroidal velocity perturbation and \( \nu \) is a (turbulent) viscous diffusivity. Note that the Lorentz force is quadratic in the magnetic field, so that \( v \) can have both a steady component and a component with twice the frequency of the dynamo itself. In general, we expect to find fluctuations in angular velocity with a period equal to half that of the magnetic cycle (Schüssler 1981, Yoshimura 1981, Kleoerin and Ruzmaikin 1984, 1991), corresponding to the torsional oscillations that have been observed on the sun (Howard and LaBonte 1980, Ulrich et al. 1987). In addition, there may be
enhanced losses through magnetic buoyancy, which allows toroidal flux tubes to leave the region where the poloidal field is regenerated.

Qualitative features of nonlinear dynamos can be established by constructing truncated models (low-order systems of ordinary differential equations) which can then be investigated in considerable detail. These toy systems exhibit generic patterns of behavior that are shared by solutions of the full partial differential equations. One-dimensional dynamo waves were first described by Parker (1955, 1979), who showed that there was an oscillatory (Hopf) bifurcation, leading to exponentially growing solutions of the kinematic problem, when an appropriately defined dynamo number $D = 1$ (see the review by Wilson in these Proceedings). Nonlinear extensions of this problem, with growth limited by various saturation processes, can be used to illustrate the development of complicated time-dependent behavior. Periodic nonlinear solutions can be constructed for $D > 1$; Weiss, Cattaneo and Jones (1984, and Jones, Weiss and Cattaneo 1985) studied a sixth-order model where the periodic solution lost stability in a Hopf bifurcation, giving rise to doubly-periodic solutions with trajectories lying on a torus in phase-space. A further Hopf bifurcation, followed by frequency-locking, led to chaotic oscillations for $D \approx 4$. These periodic solutions exhibited two time-scales (relics of the vanished torus): one was that of the magnetic cycle, the other corresponded to irregular modulation on a longer timescale, producing an envelope similar to that derived from the curves in Figure 2 and so showing that grand minima can appear naturally even in a very simple model. A similar procedure can be employed to investigate spatial structure. Solutions of the kinematic problem have either dipole or quadrupole symmetry, with eigenfunctions whose toroidal components are antisymmetric or symmetric about the equator, respectively. In the nonlinear domain these symmetries may be broken at secondary bifurcations that lead to the appearance of asymmetric oscillations (resembling observed behavior on the sun). In order to understand the associated bifurcation structure it is essential to follow both unstable and stable solution branches. This structure has been analyzed for a relatively low-order model (Jennings 1991, Jennings and Weiss 1991) in order to bring out generic features shared by any axisymmetric mean field dynamo. Such models also display another important feature of nonlinear systems: there may be several different stable solutions, as well as many unstable solutions, for the same parameter values.

The next stage is to construct mean field dynamos that include some arbitrary nonlinear quenching mechanism. The simplest models are one-dimensional, with $B$ dependent only on latitude and time (Leighton 1969, Stix 1970, Krause and Meinel 1988, Schmitt and Schüssler 1989, Belvedere, Pidatella and Proctor 1989, Schmalz and Stix 1991) and they too provide examples of symmetry-breaking and complicated time-dependence. There have also been many axisymmetric dynamo models, with both radial and latitudinal variation of the field. These models have been frequently described (e.g. Parker 1979, Stix 1989, Moss, Tuominen and Brandenburg 1990): when suitably tuned, they produce dynamo waves that migrate towards the equator, with butterfly diagrams like those obtained for active regions on the sun. By modifying the spatial variation of $\alpha$ or $\omega$ it is possible to generate simultaneous poleward and equatorward migration of activity (Stix 1987, D. Schmitt 1987, Belvedere, Proctor and Lanzafame 1991, Benevolenskaya, in
these Proceedings). Conversely, the observed poleward migration of fields at high latitudes can be used to infer the pattern of motion where the dynamo is located. Mean field dynamos also provide a means of exploring the effects of different quenching mechanisms (Noyes et al. 1984b) or the relative importance of the $\alpha$- and $\omega$-effects in an $\alpha^2\omega$ dynamo (cf. Jennings, in these Proceedings). In most nonlinear dynamo models the cycle frequency increases with increasing dynamo number, as inferred from observations.

A more ambitious approach is to solve the full partial differential equations, including the equation of motion, numerically without explicitly averaging over small-scale turbulent eddies. (For astrophysical applications this involves the assumption that turbulent diffusion can be represented by eddy diffusivities that are large enough to permit an accurate computation.) The first successful self-consistent dynamo model was computed by Gilman (1983). It was driven by convection in a Boussinesq fluid contained within a rotating spherical annulus and dynamo action appeared in a subcritical Hopf bifurcation, leading to dynamo waves that propagated towards the poles. Similar results were obtained by Glatzmaier (1985a) in the anelastic approximation. The importance of these calculations is as a demonstration that the dynamo process really works, even though details do not match observed features of the solar cycle. A crucial property of the models is that the form of the motion is determined by the Coriolis force, which leads to convection in cells elongated parallel to the rotation axis. Experiments on convection in a rotating system, carried out in space so as to obtain a zero- $g$ environment, confirm that banana cells appear when the rotation rate is sufficiently large (Hart et al. 1986a, b). In such a configuration the angular velocity tends to be constant on surfaces that are roughly cylindrical, whereas the rotational splitting of $p$-modes indicates that $\Omega$ depends only on latitude within the solar convection zone (Dziembowski, Goode and Libbrecht 1989, Rhodes et al. 1991). Thus the angular velocity profiles in the sun, where the inverse Rossby number $\sigma$ is of order unity, differ significantly from those in the numerical models (cf. Weiss 1989). Yet in a rapidly rotating star, with $\sigma \gg 1$, we may expect $\Omega$ to be constant on cylinders, so that the Gilman-Glatzmaier models assume much greater relevance.

Others have attempted to model turbulent dynamos in Cartesian geometry. The difficulty here is to distinguish between transient amplification of the magnetic field and true dynamo action (Cattaneo, Hughes and Weiss 1991). The recent results of Brandenburg et al. (1992, Nordlund et al. 1992) are very striking. They simulate compressible convection in a stratified rotating system, and show that magnetic fields can be maintained by dynamo action if the magnetic Reynolds number is sufficiently large. The resulting field is highly intermittent: the flux tubes wind round isolated sinking plumes with locally concentrated vorticity. There is no systematic cyclic behavior but these results do show that turbulent convection in a rotating star is bound to generate small-scale magnetic fields.

6. NEW DIRECTIONS

The above discussion might suggest to the casual reader that the clear direction of future dynamo calculations is toward more sophisticated versions
of the type of simulations carried out by Brandenburg and collaborators. To some extent we agree but one must also keep firmly in mind that numerical simulations cannot ever represent the physical conditions actually extant in the solar interior – the solar fluid is simply too inviscid. For these reasons, dynamo theory still has – and will continue to have – need for model calculations whose purpose is to explore physical circumstances that the simulations cannot hope to capture.

An example of calculations that are really beyond the ability of detailed numerical simulations is provided by models for “fast dynamos”. The study of fast dynamos has intrigued a whole generation of applied mathematicians (Childress 1992), and over the past 10 years has – more than any other development – vastly increased the community of dynamo theorists. What then are “fast dynamos”?

Consider once again the induction equation (2.1), in the kinematic limit (i.e. \( \mathbf{u} \) is specified). For simplicity, consider only time-independent flows. Then let us suppose that there are solutions to this equation of the form

\[
\mathbf{B}_i(r) = \text{Re} \left[ \mathbf{B}_i(r)e^{\gamma_i t} \right], \quad [\gamma_i] \text{ complex}
\]  

(6.1)

As before, we expect dynamo growth if \( \text{Re}(\gamma_i) > 0 \) for at least one \( i \); this growth will then be oscillatory or non-oscillatory if \( \text{Im}(\gamma_i) \neq 0 \) or \( =0 \), respectively. The crucial question is then how fast this growth occurs relative to the dynamical time of the system, typically taken to be the turn-over time \( t_o \equiv \ell_o/u_o \), where \( \ell_o \) and \( u_o \) are the typical length scale and velocity of the flow:

A slow dynamo has the property that

\[
\text{Re}(\gamma_i) \cdot t_o \rightarrow 0 \text{ as } R_m \rightarrow \infty;
\]

(6.2)

such dynamos generate fields in the diffusive regime.

A fast dynamo has the contrasting property that

\[
\text{Re}(\gamma_i) \cdot t_o \rightarrow \text{constant} \neq 0 \text{ as } R_m \rightarrow \infty;
\]

(6.3)

for such dynamos, diffusion is irrelevant.

Thus the challenge of fast dynamo theory is to compute the growth rate \( \text{Re}(\gamma_i) \) as a function of \( R_m \), and to show that it asymptotes to a constant value as the magnetic Reynolds number is increased without limit. The classic example of such a dynamo (which however has a number of peculiarities that make it irrelevant as a physically-meaningful model of naturally-occurring dynamos) is the “stretch-twist-fold” dynamo first studied by Vainshtein and Zel’dovich (1972, see also Zel’dovich and Ruzmaikin 1984, Zel’dovich et al. 1983), in this case, the flux doubling time scales as \( \ell_o/u_o \approx t_o \), and is independent of the diffusivity \( \eta \). Nevertheless, the fact that fast dynamos operate at very high Reynolds numbers by definition makes them extremely attractive candidates for astrophysical dynamos – but do such dynamos, characterized by smooth flows – really exist? The complete answer to this question is still outstanding, though certain idealized numerical models provide convincing evidence that the growth rate converges to a finite positive value as \( R_m \) is increased (Galloway and Proctor 1992). Perhaps more important from the astrophysical perspective is the question whether such dynamos exist if the
Lorentz force backreaction is taken into account; the answer to this question is entirely unknown.

A rather different approach which also addresses problems not well-addressed by numerical simulations concerns the backreaction of magnetic fields on small-scale fluid flows. By and large, the nonlinear effects we have considered so far typically affect only the large-scale motions in the systems. This focus is obvious in the context of the mean field dynamos for the simple reason that the corresponding equations are explicitly for the large-scale field; in the case of even sophisticated numerical simulations, the same occurs because these calculations cannot hope to model truly small-scale dynamics while at the same time modeling the field evolution on the scale of the entire system, viz. the Sun. In a recent paper, Vainshtein, Parker and Rosner (1992) address the backreaction of magnetic fields on small-scale motions, and show that this backreaction can in fact completely modify the nature of field amplification (such phenomena have already been discussed above in the context of the suppression of eddy diffusion by relatively very weak magnetic fields).

Vainshtein et al. make two fundamental points: First, they point out that in three dimensions, the suppression of turbulent diffusion discussed above cannot be total since a three-dimensional fluid has the freedom to arrange fluid motions such that the magnetic backreactions are minimized. In particular, the fluid motions can become two-dimensional (in planes locally orthogonal to strong magnetic field lines); such motions correspond to small-scale interchanges of field lines, which are not inhibited by the Lorentz force. Thus, the concern broached above about the absence of turbulent diffusion for even relatively weak magnetic fields is not as worrisome as might have appeared.

The second point made by Vainshtein et al. is that the very process of two-dimensional mixing may resolve a further difficulty of magnetic field backreaction: It is thought that the solar dynamo produces strong toroidal fields; such fields are likely to be sufficiently strong to suppress the $\alpha$-effect. The reason is that in the “standard” picture first discussed by Parker (1955), the $\alpha$-effect is visualized physically as a process in which a loop of rising toroidal flux is twisted by the Coriolis force so as to have a non-vanishing projection in the meridional plane; it is this twisting process that may be suppressed by sufficiently strong toroidal fields. The solution to this quandary offered by the two-dimensional mixing, consisting of mixing toroidal flux in meridional planes, is that these mixing motions can lead to reconnection of toroidal flux lines of opposite polarity, and thus to closed loops of toroidal flux. These closed loops, once they rise, are also subject to the Coriolis force, but now cannot react back since they are no longer anchored to the ambient toroidal flux system.

The insights gained by these studies of dynamo models should in principle be directly testable in the context of numerical simulations of simple model systems. This has already been attempted for fast dynamos (Galloway and Proctor 1992). In the case of the model calculations of Vainshtein et al., future directions are becoming clear: For example, it will be of considerable interest to look for the “two-dimensionalization” of turbulent flows at the small spatial scales at which magnetic fields first become dynamically important; and subsequently to study the effects on turbulent diffusion of magnetic fields which result from this behavior.
7. CONCLUSION

Any model of the solar dynamo rests on a few key observations. Hale (1908, 1913) discovered that the strong toroidal field emerging through sunspots was antisymmetric and alternated in direction. Forty years later, the Babcocks (1955) measured weaker fields and showed that the polar field reversed near sunspot maximum (H.D. Babcock 1959). Meanwhile, Parker (1955) had put forward the first theoretical model of an oscillatory dynamo. H.W. Babcock (1961) combined all these ideas into a remarkably impressive synthesis. He proposed that a shallow poloidal field, with dipole symmetry, was drawn out, amplified and twisted by differential rotation to produce isolated ropes of flux. From these flux ropes, loops (or stitches) rose to the surface and emerged as active regions, which gradually dispersed in such a way that the following polarity migrated polewards and reversed the dipole field (Babcock and Babcock 1955, Leighton 1969). Babcock’s phenomenological model has influenced all subsequent discussions of the solar cycle. After 30 years, however, it is clear that many features have to be updated.

A new consensus is gradually emerging and the picture that we outline would be accepted by many (though certainly not all) of those who are interested in this topic. First, the dynamo is now thought to be located near the base of the convective zone or, more likely, in a thin region of convective overshoot. The original arguments for such a shell dynamo were concerned with the competition between magnetic buoyancy and flux expulsion, the rotation rates of large-scale fields and the sense of the $\alpha$-effect (Parker 1979, Rosner and Vaiana 1980, Spiegel and Weiss 1980, Golub et al. 1981, Van Ballegooijen 1982, Schmitt, Rosner and Bohn 1984, Glatzmaier 1985b). More recently, helioseismic data have shown that strong gradients of angular velocity are only present near the base of the convection zone (Dziembowski et al. 1989, Rhodes et al. 1991). The transition from the latitudinal variation observed at the surface to solid body rotation takes place over a range of depth too small to be resolved (<0.1$R_\odot$) at or just below the base of the convection zone (Goode 1991, Goode et al. 1991). Within the magnetic layer there is a strong, predominantly azimuthal field that is more likely to vary smoothly than to be confined to isolated ropes. Crude estimates suggest a field strength of order $10^4$ G in a layer $10^4$ km thick (Galloway and Weiss 1981, Parker 1987b). Where the field is locally stronger (~ $10^5$ G) it becomes unstable to buoyancy-driven instabilities (e.g. Schmitt and Rosner 1983, Hughes and Proctor 1988, Hughes 1992 and references therein), giving rise to isolated flux tubes which rise to the surface and form active regions (Moreno-Insertis 1992, D’Silva, in these Proceedings).

The radial gradient of angular velocity (the tachocline) is an essential ingredient of the dynamo (Gilman, Morrow and DeLuca 1989, DeLuca and Gilman 1986, 1988). Moreover, since $\partial \Omega / \partial r$ changes sign at mid-latitudes, it is possible to generate dynamo waves that propagate in opposite directions at low and high latitudes (Belvedere et al. 1991, Benevolenskaya, in these Proceedings) as suggested by descriptions of the ‘extended solar cycle’. The tachocline itself drives meridional motions and generates two-dimensional turbulence (Spiegel and Zahn 1992). In this revised picture the poloidal field is reversed as a result of cyclonic turbulence (the $\alpha$-effect) within the magnetic layer, rather than by turbulent diffusion at the surface (as proposed
by Babcock and advocated by Sheeley in these Proceedings). This
distinction is supported by the different rotation rates measured for
large-scale and small-scale fields at the solar surface, which are
discussed by Stenflo elsewhere in these Proceedings.

Is there also a dynamo in the convection zone itself? The numerical
models of Nordlund et al. (1992) suggest that the combination of
turbulent convection and rotation can generate a disordered fibril field. Any such
dynamo must be coupled, however loosely, to the deep-seated dynamo that is
responsible for the relatively ordered behavior associated with the solar cycle.
The latter injects magnetic flux into the turbulent region, where it is shredded
and combined with locally generated flux. Disordered fields may then emerge
as ephemeral active regions and be responsible for X-ray bright points (Golub

This revised phenomenological model has yet to be supported by
detailed calculations. The best way ahead is first to establish the structure
of compressible convection beneath the observable surface of the sun; here
substantial progress is being made. Next the interaction between convection
and rotation, which produces the observed variation of $\Omega$ with radius and
latitude, has to be understood. Only then will it be possible to include
magnetic fields and to construct a convincing model of the solar cycle.

REFERENCES

Belvedere, G. 1990, in Inside the Sun, eds. G. Berthomieu and M. Cribier,
Spontaneous Formation of Space-Time Structures and Criticality, eds. T.
Brandenburg, A. and Tuominen, I. 1991, in The Sun and Cool Stars: Activity,
Magnetism, Dynamos, eds. I. Tuominen, D. Moss and G. Rüdiger,
Trans. R. Soc. Lond., A., 330, 481.


© Astronomical Society of the Pacific • Provided by the NASA Astrophysics Data System