Astron. Astrophys. 318, L55–L58 (1997)

ASTRONOMY AND ASTROPHYSICS

Letter to the Editor

The rôle of absolute instability in the solar dynamo

S.M. Tobias[1,2], **M.R.E. Proctor**[1], **and E. Knobloch**[2]

[1] Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge CB3 9EW, UK
[2] JILA, University of Colorado, Boulder, CO 80309-0440, USA

Received 3 September 1996 / Accepted 19 December 1996

Abstract. The presence of the equator and poles has a profound effect on the critical dynamo number D_f for the onset of dynamo waves and their subsequent nonlinear evolution. The instability threshold is associated with the onset of absolute instability, which is necessary for the existence of a globally unstable mode. With increasing D the waves become very rapidly nonlinear and the region of activity extends to higher and higher latitudes. The wave frequency is selected in the polar regions where the mode amplitude is small. Secondary absolute instabilities lead to the coexistence of waves with different wavenumbers, frequencies and amplitudes at different latitudes.

Key words: MHD – Sun: magnetic fields – dynamo

1. Introduction

The most promising explanation for the origin of solar magnetic activity is that strong toroidal magnetic field is generated by a magnetohydrodynamic dynamo acting in a thin layer at the base of the solar convection zone (see Weiss 1994). Many models (*e.g.* Kitchatinov $et\ al$ 1994) employ the mean field electrodynamic formulation of the dynamo problem in which the toroidal field is generated by the action of a strong shear and cyclonic turbulence is responsible for the regeneration of the poloidal field. The models usually involve the numerical integration of partial differential equations in two or three dimensions and only a limited survey of the parameters is possible. For this reason it is difficult to ascertain the rôle of the various parametrizations in the model in determining the nonlinear behaviour of the dynamo. In this paper we present a simple model of the solar dynamo that explains the consequences of including realistic boundaries at the pole and equator.

In Parker's original conception (Parker 1955), the instability leading to dynamo waves was essentially local in nature, so that the critical dynamo number for the onset of instability and the associated critical wave speed could be calculated by assuming a solution in the form of a uniform periodic wave. One might think that if the pole-equator distance were large compared with the dynamo wavelength such a local analysis would give useful information. However, as noted already by Parker (1971), the presence of boundaries at the pole and equator results in behaviour that has no counterpart in the case of the uniform wave. Indeed, Worledge $et\ al$ (1996) (hereinafter WKTP) have shown that the onset of dynamo action for a shell dynamo model (as described below) in a bounded region takes the form of a confined, time-dependent "wall" mode rather than a global wavelike solution, and that, as a consequence of the boundaries, the onset of instability is delayed by an order one amount, *regardless* of the aspect ratio L of the shell. Here we present nonlinear simulations of the same model. We consider an $\alpha\Omega$ dynamo operating in the deep convection zone or overshoot layer. By averaging over the radial direction we regard the magnetic field as a function only of time and x, the distance from the equator. For simplicity a cartesian geometry will be assumed, and in order to focus on the rôle of the boundaries both the α-effect and the differential rotation will be supposed to be independent of x. (In recent work on the effects of spatial variation in these quantities [Kuzanyan & Sokoloff 1995] the boundary conditions are of secondary importance.) In the usual nondimensional notation, the governing equations are (*cf.* Proctor & Spiegel 1991)

$$\frac{\partial A}{\partial t} = \frac{DB}{1+B^2} + A'' - A \ , \tag{1}$$

$$\frac{\partial B}{\partial t} = A' + B'' - B \ . \tag{2}$$

Here A is the potential for the poloidal magnetic field, B is the toroidal field, and primes indicate differentiation with respect to x. We take the equator to be at $x = 0$ and the pole at $x = L$. We nondimensionalise x with the (small) depth of the overshoot layer, so that we expect L to be substantially larger than unity. The only parameter in the model, apart from L, is the dynamo number D, proportional to the product of the α-effect and the shear. The nonlinearity in the equation for A parametrizes the effect of the large-scale magnetic field in suppressing the α-effect (Tao $et\ al$ 1993) and is of standard "α-quenching" type (see $e.g.$ Stix 1972); we have not investigated the effects of other types of nonlinearity in detail. We adopt boundary conditions at $x = 0, L$ that mimic those at the equator and pole; specifically, we put $A(L,t) = B(L,t) = 0$, while at the equator we choose conditions appropriate to dipole parity of the field, namely $A'(0,t) = B(0,t) = 0$. We have experimented with a variety of boundary conditions, and it appears that our results are insensitive to the conditions adopted, provided L is not too small.

The paper is organised as follows: in §2 we review the linear results of WKTP and then describe the transitions that occur in the nonlinear regime as D is increased. The results are explained in §3, and the paper concludes in §4 with a discussion of the importance of the results for the solar dynamo.

2. Results

The results presented here are all for positive dynamo numbers so that the waves propagate from pole to equator, as observed at least at low to moderate solar latitudes. In an unbounded domain we can look for marginally stable solutions of (1,2) (with the nonlinearity switched off) which take the form of spatially uniform waves $e^{i\omega t+ikx}$. Such solutions are present at $D = 2(1+k^2)^2/k$ and have frequency $\omega = 1+k^2$. It follows that the critical dynamo number (obtained by minimizing D with respect to k) $D_c = 32/3\sqrt{3} \approx 6.16$, critical wavenumber $k_c = 1/\sqrt{3}$ and frequency $\omega_c = 4/3$. The quantity D_c represents the threshold for *convective* instability.

In contrast, in a *finite* domain the (linear) onset of dynamo action is delayed to the value of the dynamo number $D_f(L) = D_a + C_1 L^{-2} + o(L^{-2})$, where $D_a = \sqrt{2}D_c$. The constant C_1 is independent of the boundary conditions applied at $x = 0, L$. D_a turns out to mark the onset of the *absolute* instability of the trivial solution in an unbounded domain and is obtained from the requirement that the dispersion relation $(i\omega + k^2 + 1)^2 - ikD = 0$ has a double root in the complex k-plane (*cf.* Huerre & Monkewitz 1990). The eigenfrequency $\omega_f(L) = \sqrt{3} + C_2 L^{-2} + o(L^{-2})$, where C_2 is also universal. The associated eigenfunction takes the form of a wall mode attached to the left hand boundary (equator), with lateral extent of order unity. As D is increased from D_f there is a very small range ($D - D_f = O(L^{-5})$) in which the nonlinear solution at large times resembles this eigenfunction.

In the nonlinear regime, simulations show that for periodic boundary conditions the waves persist, becoming less sinusoidal in form as they grow in amplitude, but suffering no secondary bifurcations if their wavelength is sufficiently close to the optimal value $2\pi\sqrt{3}$. We now contrast this with the nonlinear behaviour in a finite domain.

When $D - D_f$ rises to values of order L^{-2} the finite amplitude solution changes its spatial form, although its frequency scarcely changes from ω_f. As shown in Fig. 1 for two different domain lengths, the disturbance spreads to fill part of the domain, with the region of disturbance terminated by a front that resembles, for small amplitude, the linear eigenfunction. The position $x = L_{front}$ of the front appears to be well predicted by the relation $D = D_a + C_1(L - L_{front})^{-2}$. Behind the front the solution takes the form of a uniform, finite amplitude wavetrain, with a wavenumber close to the value $\sqrt{(\sqrt{3}-1)}$, larger than k_c, that would be predicted from the linear dispersion relation with $\omega = \sqrt{3}$. The selection of this wavenumber appears to be independent of the initial conditions. These solutions can be interpreted in the dynamo context as implying that for small amplitudes the zone of dynamo action is close to the equator, but that the region of activity expands very quickly to higher latitudes as D becomes larger.

As D is increased further, so that $D - D_f = O(1)$, the front collides with the right hand boundary (pole), and begins to change its character. The frequency of the solution, which is still strictly periodic, starts to increase and at a critical dynamo number slightly less than 12 there is a secondary oscillatory instability that manifests itself first at the equator. Further increase in D leads to the appearance of a "second front" in the interior of the computational domain, separating two apparently periodic and uniform wavetrains with different wavenumbers, frequencies and amplitudes. The resulting solution may be quasiperiodic or periodic. For yet larger values of D the high latitude part of the wavetrain becomes chaotic (see Fig. 2) with the location of the intervening front undergoing chaotic oscillations. Within the chaotic regime one can find intervals of periodic solutions.

3. Interpretation of the results

As discussed by WKTP for $D_c < D < D_f$ the system (1,2) describes dynamo waves that undergo transient amplification but eventually decay. This is because the waves necessarily travel equatorward. Consequently any wave reflected from the equator must be evanescent, and the equator thus plays the role of an absorbing boundary. Only with the onset of the global mode at D_f can the driving mechanism supply sufficient energy to overcome the dissipation in the equatorial regions. At $D = D_f$ this mode is localized near the equator, with the dissipation at the equator just balanced by spatially amplifying waves coming in from higher latitudes. With increasing D the spatial growth rate of the incoming waves increases and the global mode consequently expands to higher latitudes, attaining finite amplitude at lower latitudes. The structure of the front separating the active latitudes from the inactive ones is determined by a balance between the rate of advection of the wave energy towards the equator (which depends on the group velocity of the waves) and their poleward diffusion due to the (turbulent) magnetic diffusivity. In all cases, the polar regions serve to select the frequency of the waves, and this (linear) frequency then determines the wavelength (and amplitude) of the nonlinear waves at lower latitudes via equations (1,2). The nature of the boundary conditions at $x = L$ is immaterial, provided only that they exclude incoming wave flux. The system behaves essentially like an amplifier, amplifying small perturbations at high latitudes into observable waves at lower latitudes. When $D - D_f = O(L^{-2})$ the amplified frequency is $\omega = \omega_f + O(L^{-2})$ and consequently the polar regions set the frequency of the observable waves, serving as a *pacemaker* for the solar cycle. In this regime we expect the timing of the solar cycle to be largely independent of its amplitude. In contrast, when $D_c < D < D_f$ our model is very sensitive to stochastic perturbations at high latitudes (*cf.* Deissler 1985, 1989). In this regime (not illustrated) our dynamo picture bears some similarity to that invoked by Hoyng (1990), although in our model the perturbations continue to amplify well into the nonlinear regime.

When $D - D_f$ exceeds $O(L^{-2})$ the global mode extends (nearly) to the pole and the frequency of the waves begins to increase rapidly with D, and now depends on the details of the boundary conditions at the pole. This change in the frequency typically triggers a secondary instability which leads to adjacent regions in latitude with waves of different amplitude, wavenumber and frequency, separated by a relatively sharp "shock" front. An important property of such shocks is that the phase of the wave is *not* conserved across them. The physics behind the appearance of these shocks is similar to that for the primary instability: the shocks form when the group velocity of the perturbations of the basic wavetrain vanishes, *i.e.*, when the dispersion relation for these perturbations contains a double root in the complex plane, signalling the appearance of a *secondary* absolute instability (*cf.* Deissler 1989, Brevdo and Bridges 1996). Detailed substantiation of this conclusion can be found in Tobias $et\ al$ (1996). The new frequency introduced by the secondary instability appears to be responsible for the selection of the wavenumber downstream from the "shock". Our calculations indicate that the selected frequencies and the corresponding wavenumbers are uniquely defined by this process.

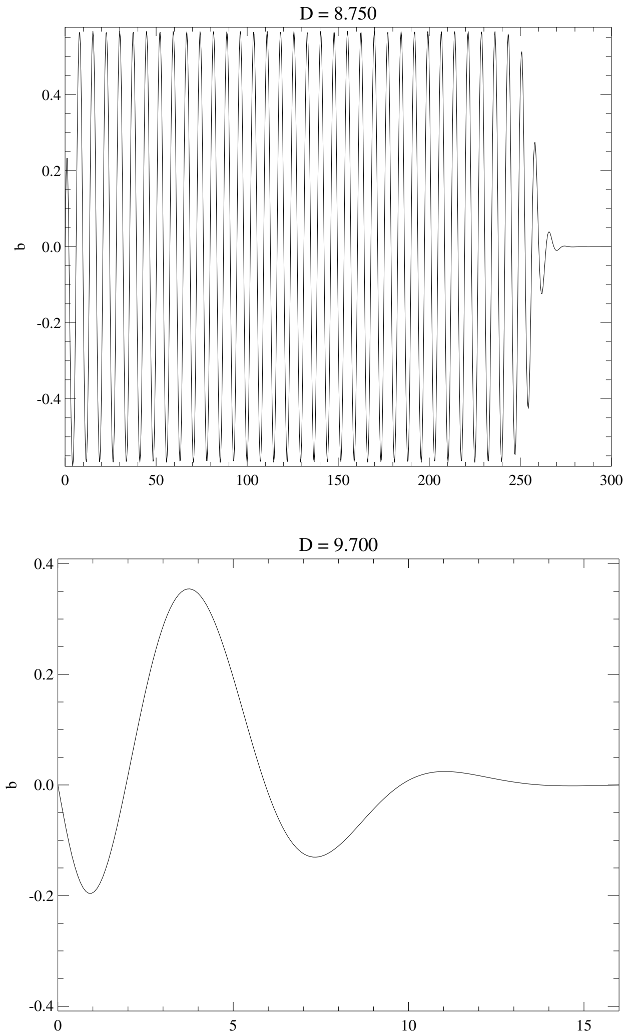

Fig. 1. Filling transition: Toroidal field $B(x)$ for (a) $L = 300$, $D = 8.750$, (b) $L = 16$, $D = 9.700$. As the dynamo number increases above $D_f \approx 8.7$ ($L = 300$), $D_f \approx 8.9$ ($L = 16$), the region of dynamo action expands towards higher latitudes. The solutions consist of a uniform wavetrain near the equator ($x = 0$) bounded by a front that resembles the linear wall-mode solution. The wavelength is determined by the frequency of infinitesimal waves near the right boundary (pole).

4. Discussion

In this paper we have shown that imposing boundary (or regularity) conditions in a dynamo model where the α-effect vanishes at the equator has a profound effect on the dynamo instability. The unidirectional nature of the waves and the absence of a source of wave flux from the pole together imply that the critical dynamo number for the instability will be substantially larger than that predicted by local calculations, *regardless* of the aspect ratio of the domain. The presence of the equator is in turn responsible for the shock structure of the resulting nonlinear wavetrains. Thus our model predicts a natural tendency to form active and inactive latitudes, with a relatively sharp front separating them. The polar regions select the frequency of the waves (either linearly or nonlinearly), and this frequency is in turn responsible for the selection of the wavenumber and amplitude of the waves observed behind the front. This mechanism may account for the temporal stability of the solar cycle. In addition, we have uncovered the possibility that secondary fronts separating two wavetrains with different properties may form, and have shown that the formation of such fronts is associated with the onset of absolute instability of the wavetrain, in the same way that the initial instability in a finite domain is associated with the absolute instability (as opposed to convective instability) of the trivial state $A = B = 0$.

Although the solar cycle visible at the photospheric level is more akin to that shown in fig. 1b than fig. 1a, we believe that if the seat of the dynamo is a large aspect ratio layer like the convective overshoot region the large aspect ratio results shown here could well be relevant to the solar dynamo. This is because the weaker fields created at high latitudes may never become buoyant enough to emerge at the photosphere.

Existing numerical simulations of mean field dynamos in spherical shells with a latitude dependent α do not focus on the latitudinal structure and wavelength of the waves, and as a result do not reveal unambiguously the tendency to form such

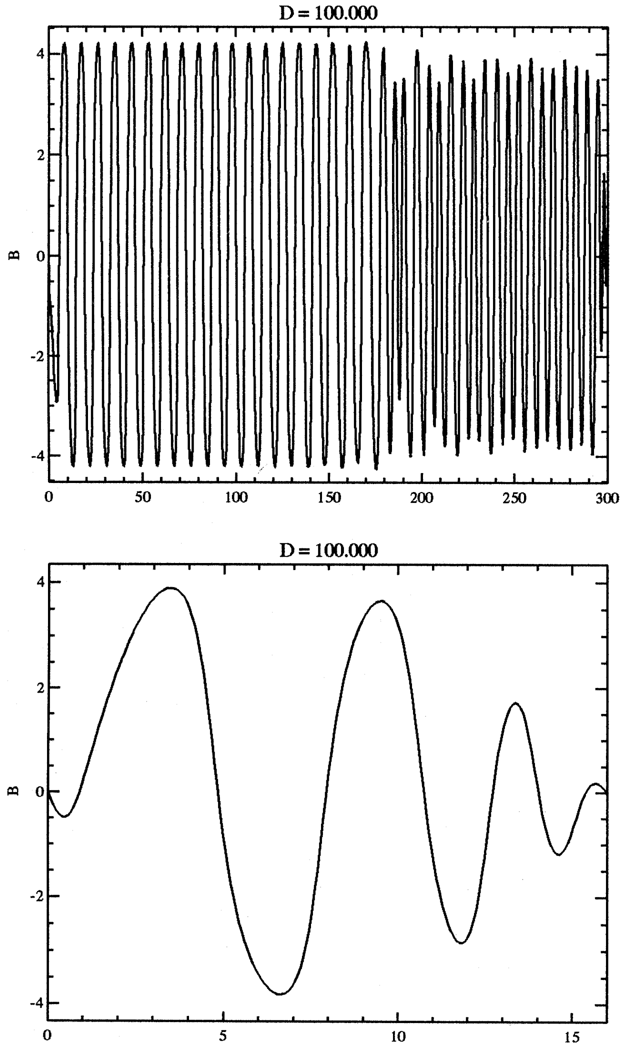

Fig. 2. Chaotic solution with two wavelengths: Toroidal field $B(x)$ for (a) L = 300, D = 100, (b) L = 16, D = 100. The solutions now take the form of two wavetrains of different amplitudes, wavenumbers (and frequencies), separated by a (chaotically oscillating) front.

active latitudes. In such calculations there is a natural tendency to attribute any nonuniformities in the waves to either the effect of spherical geometry or to nonuniformities in the generation mechanism. We have shown here that the dynamo process has a natural proclivity to form spatially inhomogeneous waves, even when the generation mechanism is spatially uniform. Indeed, we have found similar phenomena with appropriate nonlinearities even with spatially nonuniform αs, provided only that the dynamo number is sufficiently supercritical. It should be stressed that these results are very general and so should also apply to the generation of magnetic fields in other astrophysical bodies such as accretion discs.

Acknowledgements. S.M.T. is grateful to Trinity College, Cambridge University, for a Research Fellowship. This work was completed while E.K. was a Visiting Fellow at JILA, University of Colorado, Boulder. We would like to thank Derek Brownjohn for his assistance with the figures.

References

Brevdo L. & Bridges T. J. 1996, Phil. Trans. R. Soc. London A 354, 1027

Deissler, R. J., 1985, J. Stat. Phys. 40, 371

Deissler, R. J., 1989, J. Stat. Phys. 54, 1459

Hoyng, P., 1990, in: Solar Photosphere: Structure, Convection, and Magnetic Fields, ed. J. O. Stenflo, IAU, p. 359

Huerre, P. & Monkewitz, P. A., 1990, Ann. Rev. Fluid Mech. 22, 151

Kitchatinov L. L., Rüdiger G. & Küker M., 1994, A&A 292, 125

Kuzanyan K.M. & Sokoloff D.D., 1995, GAFD 81, 113

Parker E.N., 1955, ApJ 122, 293

Parker E.N., 1971, ApJ 164, 491

Proctor M. R. E. & Spiegel, E. A., 1991, in: The Sun and Cool Stars: activity, magnetism, dynamos, eds. I. Tuominen, D. Moss & G. Rüdiger, Springer-Verlag, Berlin, p. 117

Stix M., 1972, A&A 20, 9

Tao, L. L., Cattaneo, F. & Vainshtein, S. I., 1993, in: Theory of Solar and Planetary Dynamos, eds. M. R. E. Proctor, P. C. Matthews & A. M. Rucklidge, Cambridge University Press, Cambridge, p. 303

Tobias S. M., Proctor, M. R. E. & Knobloch, E., 1996, preprint

Weiss N. O., 1994, in: Lectures on solar and planetary dynamos, eds. M. R. E. Proctor & A. D. Gilbert, Cambridge University Press, Cambridge, p. 59

Worledge, D., Knobloch, E., Tobias, S. & Proctor, M., 1996 Proc. R. Soc. London A, in press