

CHAOS IN SELF-EXCITING DYNAMOS AND THE MAIN GEOMAGNETIC FIELD

RAYMOND HIDE

Summary

Nonlinear feedback and coupling (F&C) in dynamical systems operating under fixed boundary conditions usually produce chaotic fluctuations, which are disorderly and unpredictable beyond a finite 'predictability horizon' ([Note] A), [(References) 1-5]). But in some circumstances nonlinear F&C *inhibit* chaos, producing *order* rather than *disorder*. Both types of behaviour may have to be invoked in the interpretation of long-term variations of the main geomagnetic field (MGF). This is generated by (mainly buoyancy-driven) magnetohydrodynamic (MHD) flow in the Earth's liquid metallic outer core, where the electrical conductivity is high enough (but not too high) for efficient MHD self-exciting dynamo action to take place. Amongst the nonlinear F&C agencies operating in the MHD 'geodynamo' are Lorentz forces, involving interactions between the electric currents generated by the geodynamo and concomitant magnetic fields. One generic nonlinear process in self-exciting dynamos is the redistribution of kinetic energy by such forces. When operating within a self-exciting Faraday-disk homopolar dynamo with its coil loaded with a nonlinear series motor, Lorentz forces can cause persistent large-amplitude chaotic fluctuations. But over a wide range of conditions they inhibit, rather than promote, persistent fluctuations, in some cases eliminating them altogether. If this 'nonlinear-quenching' process occurs in the MHD geodynamo it could account for the high degree of intermittency seen in the long-term behaviour of the MGF, as exhibited by the time-series of geomagnetic 'polarity reversals', the most striking features of which include intervals lasting as long as 3×10^7 years during which no polarity reversals appear in the palaeomagnetic record. Implied by this hypothesis, which could be tested with the aid of

valid numerical geodynamo models, is that eddies driven mainly by Lorentz forces play a crucial role in the attenuation of fluctuations. Also crucial, and in principle testable, are the roles played in the reversal process by modest changes in the lateral boundary conditions imposed on core motions, including those associated with very slow irregular convection in the highly-viscous overlying mantle, and also by the slow increase in size of the underlying solid inner core.

1. INTRODUCTION

When the MGF undergoes an occasional reversal in polarity – or, more frequently, a large-amplitude ‘excursion’ in the orientation of its magnetic axis which fails to achieve a full polarity reversal – it does so in a few thousand years. Determinations by palaeomagnetic workers of the ‘fossilized’ magnetization of sedimentary and igneous rocks also indicate that many hundreds of polarity reversals, and a larger number of excursions, may have occurred since the Earth came into existence some 4600 million years (Ma) ago – the most recent being the ‘Brunhes-Matyama’ reversal of *ca.* 0.8 Ma ago, when the MGF dropped in strength to less than about a fifth of its pre-transition and post-transition value [6, 7].

Intervals between reversals range in duration from *ca.* 0.25 Ma (‘sub-chrons’) to *ca.* 30 Ma (‘superchrons’), the average duration being *ca.* 1 Ma. During the past 400 Ma, covering the geological periods studied most intensively by palaeomagnetic workers, there have been two superchron intervals – the Permian superchron from *ca.* 290 to *ca.* 260 Ma ago, when a magnetic compass would have pointed roughly south, and the Cretaceous superchron from *ca.* 110 to *ca.* 80 Ma ago, when the polarity was the same as it is today – and there may have been many such intervals at earlier times.

The detailed interpretation of polarity superchrons and other features of the long-term behaviour of the MGF in terms of basic processes occurring at great depths within the Earth ranks as a major problem in geophysics. Indeed, the reversal time-series may hold valuable and singular clues to the structure, dynamics and evolution of the Earth’s metallic core and overlying mantle. The MGF is a manifestation of more or less irregular motions at speeds of a fraction of a millimetre per second (kilometres per year) in the liquid metallic outer core, as evinced by detailed observations of the MGF over the past few centuries [B]. Few now disagree that the MGF must be due to electric currents generated and maintained in the outer core

by a MHD self-exciting dynamo process involving inductive interactions between the magnetic field and core motions – motional induction (rather than chemical or thermoelectric effects) being the only quantitatively-viable agency for providing the electromotive forces needed to maintain the electric currents against ohmic resistance [7-12].

The near alignment of the Earth's magnetic axis with its rotation axis – a property exploited in the navigational use of the magnetic compass [B] – is most probably due to the dominant influence of gyroscopic (Coriolis) forces on core motions [13]. Less obviously, but importantly, such forces would also render the spatial and temporal characteristics of the MGF sensitive not only to the presence of the solid electrically-conducting inner core but also to modest lateral variations in the (thermal, mechanical and electromagnetic) boundary conditions imposed on motions in the fluid outer core by the overlying highly-viscous mantle [14,9]. As a working hypothesis the latter suggestion has been taken seriously by geophysicists for the past few decades – from the time when dynamical models of the Earth's mantle in which convective flow occurs at speeds with which continents drift apart (centimetres per year) and extends throughout the whole depth of the mantle, down to the core-mantle boundary (CMB), became generally acceptable [15,16].

Mantle convection varies on geological timescales and concomitant variations in the boundary conditions imposed by the mantle on motions in the underlying outer core would influence the degree of intermittency exhibited by the observed timeseries of polarity reversals. Some degree of intermittency would also be associated with any intrinsic chaotic behaviour of an essentially nonlinear geodynamo operating under *fixed* boundary conditions. The unexpected discovery of 'nonlinear quenching' of chaos in a physically-realistic Faraday-disk self-exciting dynamo loaded with a nonlinear series motor ([17], see also Section 3 below) indicated a useful line of research towards the physical interpretation of the intermittency seen in the reversals time-series, in terms of the influence of time-varying boundary conditions on chaos in the geodynamo [9].

Some of this geophysical contribution to a Pontifical Academy of Sciences symposium on 'Predictability in Science: Accuracy and Limitations' is based on background material prepared originally for informal seminars on the dynamics and MHD of spinning fluids. The References and Notes given below indicate sources of diagrams and tables used as visual aids (but which for reasons of space cannot be included here) and of useful technical bibliographies.

2. SELF-EXCITING DYNAMOS AND COSMICAL MAGNETIC FIELDS

The first to outline the idea of the self-exciting dynamo process in a fluid was J. Larmor when, in 1919, he argued that the magnetic fields of sunspots are produced and maintained by motional induction, involving thermal convection in the outer layers of the Sun. The dynamo process is now widely invoked in the interpretation of cosmical magnetic fields, i.e. those of galaxies, stars and planets, including the MGF supposedly produced by the 'MHD geodynamo' operating in the Earth's liquid metallic outer core [9, 10, 12, 15, 16, 19-22].

Such dynamos are governed by the four-dimensional (space and time) nonlinear partial differential equations (PDEs) of MHD. These express the laws of electromagnetism, mechanics and thermodynamics, to be solved under realistic boundary conditions [10-12]. Dynamo theory aims to understand the complex details of the processes involved. During the past half-century, since the publication by G.E. Backus and A. Herzenberg of the first mathematical 'existence proofs' based on the (pre-Maxwellian) equations of electromagnetism [8-12; C], dynamo theory has developed rapidly as an active area of applied mathematics. In common with other areas of geophysical and astrophysical fluid dynamics, dynamo theory now benefits from the availability of powerful computers, capable of tackling, with growing success, the full set of equations [10, 12]. But the subject suffers not only from paucity of observations but also, in comparison with other areas of fluid dynamics, from lack of guidance from crucial laboratory experiments. Interesting laboratory work has certainly been undertaken [10], but with the electrically-conducting fluids available for use on the small scale of the terrestrial laboratory it is not possible to carry out wide-ranging investigations of the MHD processes thought to be involved in self-exciting dynamos.

The most extensive mathematical studies to date are those of 'kinematic' dynamo models, in investigations of which, for reasons of mathematical expediency, the governing equations are simplified by specifying *ab initio* the fluid velocity field in the equations of electromagnetism [C; 8, 10, 11], thereby removing nonlinearities from the equations and obviating the need to solve them simultaneously with the equations of mechanics and thermodynamics. Such studies can be important, especially in the elucidation of spatial structure. But they shed little light on temporal behaviour, in the investigation of which it is necessary to simplify the equations in other ways, as in some of the successful 'mean-field' dynamo models, where some spatial details are parameterized rather than represented explicitly [9, 10, 12].

Further simplifications are needed when interest centres on the essentially nonlinear processes that underlie the complex large-amplitude fluctuations seen in observational data, such as the time-series of polarity reversals and excursions of the MGF. Recourse then has to be made to analyses of what mathematicians term 'low-dimensional' models (and physicists term 'toy' models). These are governed by nonlinear ordinary differential equations (ODEs) (rather than PDEs), requiring only modest computing facilities for their investigation. But great care has to be taken when formulating a model to ensure that it is not oversimplified to the point of being physically unrealistic and mathematically misleading.

The most extensively-studied low-dimensional self-exciting dynamo models are based on the Faraday-disk homopolar dynamo. (For references to an extensive literature starting nearly half a century ago with pioneering studies by E.C. Bullard, T. Rikitake, W.V.R. Malkus, H.K. Moffatt and others, see [10, 18, 23]). Some models are physically unrealistic, usually because mechanical friction is neglected thereby rendering the governing ODEs 'structurally unstable'. But detailed analyses of the 'structurally-stable' ODEs governing physically-realistic models are of interest not only in their own right but also because they offer insights capable of guiding research into the much more complex MHD systems [17].

3. PHYSICALLY-REALISTIC SELF-EXCITING DYNAMO MODELS

Self-exciting dynamos are electromechanical engineering devices or naturally-occurring MHD fluid systems that are capable of converting mechanical energy into magnetic energy without the aid of permanent magnets. They differ widely in their details but they all share the following essential characteristics [18; 23; C]:

(a) motional induction – as represented in the PDEs governing MHD dynamos by the nonlinear term $\mathbf{u} \times \mathbf{B}$ (where \mathbf{u} is the (Eulerian) flow velocity at a general point \mathbf{P} and \mathbf{B} is the magnetic field, see equations (C1 & C5)) – is responsible for converting mechanical energy into magnetic energy, which starts with the amplification of any infinitesimally-weak adventitious background magnetic field;

(b) motional induction must overcome ohmic losses for amplification to occur, implying that the electrical resistance of the system must be sufficiently low – i.e. a sufficiently *high* magnetic Reynolds number $R = UL\mu\sigma$ in MHD dynamos (see equation (C7), also [7-10]), where U is a characteristic

flow speed, L a characteristic length, σ the electrical conductivity of the fluid and μ its magnetic permeability;

(c) but the electrical resistance must not be so low that the magnetic field is unable to diffuse beyond the dynamo region, which sets an *upper* limit on R in MHD dynamos;

(d) Lorentz forces – as represented in governing equations of MHD dynamos by the nonlinear term $\mathbf{j} \times \mathbf{B}$, where \mathbf{j} is the electric current density at \mathbf{P} – redistribute kinetic energy within the system, thereby retarding buoyancy-driven eddies and accelerating motions in other parts of the eddy spectrum;

(e) mechanical friction – viscosity in MHD dynamos – no matter how weak is never negligible;

(f) internal F&C – as represented by the terms $\mathbf{u} \times \mathbf{B}$ and $\mathbf{j} \times \mathbf{B}$ in MHD dynamos – give rise to behaviour characteristic of nonlinear systems, including sensitivity to initial conditions, non-uniqueness, chaotic large amplitude fluctuations, hysteresis, nonlinear stability, etc.

A simple physically-realistic low-dimensional model (that takes mechanical friction into account and includes a crucial circuit element which enables Lorentz forces to redistribute kinetic energy) comprises a Faraday disk-and-coil arrangement with a series electric motor loading the coil [17, 18, 23]. The disk – to the axle of which the stationary coil is connected by a sliding contact and to the rim of which the series motor loading the coil is connected by another sliding contact – is driven into rotation with (dimensionless) angular speed $y(t)$ by a steady applied couple proportional to the dimensionless parameter a (say, see equation (3.4)), which is inversely proportional to the moment of inertia of the disk. Here the dimensionless time t is measured in units of the ratio of the self-inductance of the coil to the total electrical resistance of the dynamo circuit – the corresponding timescale in the case of an MHD dynamo being $\mu\sigma L^2$, which is several thousand years for the geodynamo. Retarding the motion of the disk are a frictional couple $-ky(t)$ and a Lorentz couple $-ax(t)w(t)$, where $x(t)$ is the main electric current generated by the dynamo and $w(t)$ is the magnetic flux intersecting the disk. In the absence of Lorentz forces, when friction alone retards the motion of the disk, y has the steady value a/k .

The armature of the series motor is driven into rotation with angular speed $z(t)$ (relative to the stationary ambient magnetic field within the motor) by a Lorentz couple $x(t)f(x(t))$ produced by the dynamo current $x(t)$, and it is retarded by a frictional couple $-lz(t)$ (see equation (3.5)). Here

$$f(x(t)) = 1 - e + esx(t), \quad 0 \leq e \leq 1, \quad (3.1)$$

where the value of the crucial parameter e depends on the design of the motor. The second and first terms on the right hand side of this equation are in the ratio $esx(t)$ to $(1-e)$, so the parameter e is a measure of the non-linearity of the electromechanical characteristics of the motor, the linear case being when $e = 0$. The nonlinear contribution to f , proportional to $esx(t)$, is produced by diverting the dynamo current through the stationary field windings of the motor (s being a measure of the mutual inductance between the armature and the field windings). This nonlinear 'internal' contribution to f is complemented by a linear contribution, proportional to $(1-e)$, provided by 'outside' sources of the ambient magnetic field in which the armature rotates. In general these sources comprise (i) currents generated by 'secondary' dynamos with $e=1$, and (ii) permanent magnets, but the latter must be excluded from consideration when dealing with systems of interest as low-dimensional models of natural MHD dynamos [23].

The nonlinear autonomous set of ODEs satisfied by the time-dependent variables (x, y, w, z) is the following [18]:

$$dx/dt + mdw/dt = (1 + m) [-x + yw - bz f(x)], \quad (3.2)$$

$$ndw/dt = (x - w), \quad (3.3)$$

$$dy/dt = a(1 - xw) - ky, \quad (3.4)$$

$$dz/dt = xf(x) - lz, \quad (3.5)$$

Equations (3.2 & 3.3) express Kirchhoff's laws applied respectively to the dynamo current, $x(t)$, flowing in the main circuit and the induced azimuthal 'eddy' current (proportional to $w(t)$) flowing in the disk. Equations (3.4 & 3.5) express angular momentum considerations applied respectively to the motion of the disk and the motion of the armature of the motor.

The essentially non-negative dimensionless control parameters (a, k, n, m) specify the electromechanical characteristics of the disk, while (b, l, s, e) specify those of the motor. The parameter n is inversely proportional to the electrical resistance of the disk to the flow of the azimuthal eddy current; m is proportional to the square of the mutual inductance between the disk and coil and inversely proportional to the difference between the product of the self inductances of the disk and coil and the square of their mutual inductance; and b is proportional to the self inductance of the armature of the motor and inversely proportional to its moment of inertia. The F&C terms in these governing equations are $(1+m)(yw - bz f(x))$ in (3.2), $-axw$ in (3.4), and $xf(x)$ in (3.5).

Instability analyses of the equations combined with numerical integrations using both digital and analogue computers show that when (i) a/k is

large enough for dynamo action to occur at all, and (ii) the parameter e is not too close to unity, over wide ranges of conditions – as specified by the other control parameters – their solutions exhibit chaotic behaviour, in which the dynamo current $x(t)$ undergoes large amplitude fluctuations reminiscent of polarity excursions and reversals of the MGF reversals [17, 18, 10]. Remarkably, however, chaotic behaviour is rare when e is close to unity, with fluctuations disappearing altogether when $e=1$, irrespective of the values of the other control parameters! It is this result, as we shall see in Section 4, that offers a possible basis for interpreting the long-term behaviour of the MGF, with its highly intermittent time-series of polarity reversals.

The original discovery of ‘nonlinear-quenching’ of persistent fluctuations (after the decay of initial transients) was made during an investigation of cases when $n=0$ [17]. The discovery emerged from the unexpected finding that although ‘Hopf bifurcations’ can occur when $0 \leq e < 1$, they disappear altogether when $e=1$, irrespective of the values of the other control parameters. The phenomenon may turn out to be fairly general, for it was subsequently found in solutions of other autonomous sets of nonlinear ODEs [24], obtained by taking sets (e.g. Lorenz, Rössler) that are known to possess persistent chaotic solutions and modifying their F&C terms [24], along lines suggested by the original work reported in [17].

We note that for every solution (x, y, w, z) of equations (3.1-3.5) there is a corresponding solution, $(-x, y, -w, z)$, with the same motions, (y, z) , but oppositely-directed currents. This ‘magnetic symmetry’ – of interest when considering ‘reversals’ of the MGF (see Section 4 below and [C]) – is obvious by inspection when $e=1$. But its demonstration (not presented here) for other (lower) values of e requires detailed considerations of the currents in the secondary dynamo to which the primary dynamo has to be coupled in order to achieve (without using permanent magnets) values of e that are less than unity, thereby stimulating chaotic large-amplitude fluctuations in $x(t)$.

4. LONG-TERM VARIATIONS IN THE MAIN GEOMAGNETIC FIELD

Very slow intermittent convection taking place within the Earth’s highly-viscous mantle that overlies the liquid metallic outer core must give rise to fluctuations, on geological timescales, in the lateral variations of the physical and chemical conditions prevailing at the core-mantle boundary (CMB) [14, 15, 9, 25]. Owing to Coriolis forces, these cause marked changes in both spatial and temporal characteristics of core motions, and conse-

quently in the magnetic fields they produce. So it is not unreasonable to suppose that each polarity superchron could be associated with a long quiescent period, when convection in the lower mantle (but not necessarily in the upper mantle) is comparatively feeble [18].

Because the CMB would then be relatively undisturbed, the corresponding state of the geodynamo (on this particular hypothesis) would be highly stable owing to processes analogous to the nonlinear quenching mechanism found in disk dynamos. What corresponds within the core to the presence of a series motor in the disk dynamo with a Lorentz torque proportional to the square of the electric current (the case when $e=1$, see equation (3.1)) is that part of the spectrum of core motions containing eddies driven mainly by Lorentz forces, rather than by buoyancy forces. However, during comparatively intensive phases of intermittent convection in the lower mantle conditions at the CMB become disturbed – with buoyancy forces associated with increased lateral temperature gradients in the lower mantle now contributing to the driving of core motions and distortions in the shape and other conditions prevailing at the CMB also influencing the pattern of motions. Concomitant distortions and other changes of the magnetic field within the core would stimulate changes in the geodynamo, possibly placing it within regimes corresponding to those found in a single-disk dynamo when the parameter e is no longer close to unity, which favour frequent reversals and excursions.

A simple model capable of such behaviour comprises a ‘primary’ single-disk dynamo subject to *steady* forcing and interacting with a ‘secondary’ single-disk dynamo subject to *slowly varying* irregular forcing. The secondary dynamo interacts with the primary dynamo solely by modulating the ambient magnetic field in which the armature of the series motor of the primary dynamo rotates, thereby modulating the value of e in the primary dynamo and moving the system slowly between the quenched and chaotic regimes, in response to the modulated forcing of the secondary dynamo.

Typical time-series of the dynamo current would include ‘superchrons’ – when x is steady and either positive or negative depending on initial conditions – occurring during intervals when the influence of the secondary dynamo on the primary dynamo is so weak that e is close to unity in the primary dynamo. By contrast, highly time-dependent behaviour found between ‘superchrons’ occurs during phases when the current in the secondary dynamo is strong enough to reduce the value of e in the primary dynamo to values significantly less than unity.

Noteworthy features of typical time-series of $x(t)$ include fluctuations that are more frequent than reversals, less pronounced in amplitude, and

exhibit systematic monotonic build up in amplitude before each reversal occurs. They are reminiscent of the excursions seen in the geomagnetic record which, according to the ideas presented here, should be less pronounced during superchrons than at other times.

5. CONCLUDING REMARKS

Modern research on the MGF includes improved analyses and applications of geomagnetic data derived from ground based and orbiting satellite observations. In rock magnetism, time-series of polarity reversals and excursions continue to be extended and refined, and in theoretical work, applied mathematicians interested in MHD continue to investigate and apply the equations governing self-exciting dynamos. Numerical modelling of dynamos [10, 12] has strengthened collaboration between observational workers and theoreticians, but the MGF is a complex phenomenon and much remains to be elucidated. The main ideas outlined in this paper were put forward in the hope that they will stimulate crucial observational and theoretical studies by those equipped to carry them out.

REFERENCES

1. Lighthill, J., 'The recently recognized failure of predictability in Newtonian mechanics', *Proc. R. Soc. Lond.* A407, 35-50 (1986).
2. Lorenz, E.N., *The essence of chaos*, London: UCL Press (1993).
3. Bjerknes, V., 'Das Problem der Wettervorhersage, betrachtet vom Standpunkt der Mechanik und Physik', *Meteorol. Zeitschr.* 21, 1-7 (1904).
4. Richardson, L.F., *Weather prediction by numerical process*, Cambridge University Press (1922).
5. Lorenz, E.N., 'Deterministic non-periodic flow', *J. Atmos. Sci.* 20, 130-141 (1963).
6. Gubbins, D., Kent, D. & Laj, C. (eds.), 'Geomagnetic polarity reversals and long term secular variation', *Phil. Trans. R. Soc. Lond.* A358, 889-1223 (2000); Opdyke, N.D. & Channell, J.E.T., *Magnetic stratigraphy*, San Diego: Academic Press (1996); Jacobs, J.A., *Reversals of the Earth's magnetic field*, Cambridge University Press (1994).
7. Merrill, R.T., McElhinny, M.W. & McFadden, P.L., *The magnetic field of the Earth*, Academic Press (1996).

8. Backus, G., Parker, R. & Constable, C., *Foundations of geomagnetism*, Cambridge University Press (1996).
9. Gubbins, D. & Herrero-Bervera, E. (eds.), *Encyclopedia of geomagnetism and paleomagnetism*, Springer Geosciences (2007).
10. Dormy, E. & Soward, A.M. (eds.), *Mathematical aspects of natural dynamos*, Grenoble Science Publishers (2007).
11. Friedlander, S. & Serre, D. (eds.), volume 2 of *Handbook of mathematical fluid dynamics*, Elsevier Science BV (2003); Ferriz-Mas, A. & Nuñez, M. (eds.), *Advances in nonlinear dynamos*, London: Taylor & Francis (2003); Moffatt, H.K., *Magnetic field generation in electrically-conducting fluids*, Cambridge University Press (1978).
12. Glatzmaier, G.A. & Roberts, P.H., 'Simulating the geodynamo', *Contemp. Phys.* 38, 269-288 (1997); Zhang, K.K. & Jones, C.A., 'The effect of hyperviscosity on geodynamo models', *Geophys. Res. Lett.*, 24, 2869-2972 (1997); LeMouél, J.-L., Allègre, C.J. & Narteau, C., 'Multiple scale dynamo', *Proc. Nat. Acad. Sci. USA* 94, 5510-5514 (1997); Busse, F.H., 'Homogeneous dynamos in planetary cores and in the laboratory', *Annu. Rev. Fluid Mech.* 32, 383-408 (2000); Dormy, E., Valet, J.-P. & Courtillot, V., 'Numerical models of the geodynamo and observational constraints', *Geochemistry, geophysics, geosystems*, Paper number 2000GC 000062, ISSN 1525-2027, 42 pages (2000); Narteau, C. & LeMouél, J.-L., 'Transient evolution regimes in a multiscale dynamo model: timescales of the reversal mechanism', *J. Geophys. Res.* 110, B01, doi: 10.29/2004J B002, 983, (2005).
13. Elsasser, W.M., 'The origin of the Earth's magnetic field', *Phys. Rev.* 55, 489-498 (1939).
14. Hide, R., 'Motions of the Earth's core and mantle and variations of the main geomagnetic field', *Science* 157, 55-56 (1967); 'Interaction between the Earth's liquid core and solid mantle', *Nature* 222, 1055-1056 (1969); 'On the Earth's core-mantle interface', *Quart. J.R. Meteorol. Soc.* 96, 579-590 (1970); 'On the role of rotation in the generation of magnetic fields by fluid motions', *Phil. Trans. R. Soc. Lond.* A306, 223-234 (1982).
15. Jones, C.A., Soward, A.M. & Zhang, K.K. (eds.), *Earth's core and lower mantle*, London: Gordon & Breach (2003).
16. Melchior, P.J. *The physics of the Earth's core*, Oxford: Pergamon Press (1986).
17. Hide, R., 'Nonlinear quenching of current fluctuations in a self-exciting homopolar dynamo', *Nonlinear Processes in Geophysics* 4, 201-205

- (1997); 'Structural instability of the Rikitake disk dynamo', *Geophys. Res. Letters* 22, 1057-1059 (1995).
18. Hide, R., 'Generic nonlinear processes in self-exciting dynamos and the long-term behaviour of the main geomagnetic field, including polarity superchrons', *Phil. Trans. R. Soc. Lond.* A358, 943-955 (2000).
 19. Proctor, M.R.E. & Gilbert, A.D. (eds.), *Lectures on solar and planetary dynamos*, Cambridge University Press (1994).
 20. Soward, A.M., Jones, C.A., Hughes, D.W. & Weiss, N.O. (eds.), *Fluid dynamics and dynamos in astrophysics and geophysics*, London: Taylor & Francis (2004).
 21. Thompson, M.J. & Christensen-Dalsgaard, J. (eds.), *Stellar astrophysical fluid dynamics*, Cambridge University Press (2003).
 22. Parker, E.N., *Cosmical magnetic fields: their origin and their activity*, Oxford: Clarendon Press (1979); Krause, F. (ed.), *The cosmic dynamo*, Dordrecht: Kluwer (1993); Mestel, L., *Stellar magnetism*, Oxford: Clarendon Press (2003); Schrijver, C.J. & Title, A.M., *Solar and stellar magnetic activity*, Cambridge University Press (2000).
 23. Hide, R., 'The nonlinear differential equations governing a hierarchy of self-exciting coupled Faraday-disk homopolar dynamos', *Phys. Earth Planet. Interiors* 103, 281-291 (1997).
 24. Hide, R., McSharry, P.E., Finlay, C.C. & Peskett, G.D., 'Quenching Lorenzian chaos', *Int. J. Bifurcation and Chaos* 14, 2875-2884 (2004).
 25. Gunis, M., Wysession, M.E., Knittel, E. & Buffett, B.A., (eds.), *The core-mantle region*, Washington DC: American Geophysical Union (1998).

NOTES

A. PREDICTABILITY HORIZONS

...notwithstanding the continuing success of applications of Newtonian dynamics in many fields of science and engineering...modern theories of dynamical systems have...clearly demonstrated that *the equations of Newtonian dynamics do not necessarily exhibit the 'predictability' property*. Indeed, ...recent researches have shown that in wide classes of...simple systems satisfying the equations *predictability is impossible beyond a definite time horizon* (J. Lighthill, President of the International Union of Theoretical and Applied Mechanics [1]).

The idea of a 'predictability horizon' is important [1,2]. Nearly fifty years ago meteorologists, encouraged by advances in computing and satellite technology, started planning their ambitious 'Global Atmospheric Research Programme' (GARP), which was formally announced in 1963 (in a speech by the then President of the United States, J.F. Kennedy). The planners hoped that the predictability horizon for useful numerical weather forecasts based on greatly improved data sets and advanced computational methods could eventually be extended from a day or so to more than a week. Computer technology would render feasible methods based on the governing PDEs of hydrodynamics along lines first discussed decades earlier by V. Bjerknes, a Norwegian meteorologist (and later a member of the Pontifical Academy of Sciences), and L.F. Richardson, a British meteorologist [3,4]. From their experience with weather forecasting and, more recently, climate forecasting, modern meteorologists appreciate – at least as well as other geophysical scientists engaged in research on large-scale natural phenomena – the difficulties encountered when translating understanding of basic dynamical processes into practical schemes for predicting future behaviour. Indeed, it was no accident that one of the major advances in what in 1973 became known as 'chaos theory' was the publication ten years earlier, in a leading meteorological journal, of a paper [5] entitled 'Deterministic non-periodic flow' by E.N. Lorenz, who had been engaged with other leading meteorologists in the assessment of likely improvements in 'predictability horizons' to be expected from GARP. The Earth's atmosphere is by far the most intensively-studied 'geophysical fluid', and many ideas about predictability stem from research connected with the very difficult practical activity of forecasting the atmosphere's future behaviour.

B. GEOMAGNETIC SECULAR VARIATION AND CORE MOTIONS

Geomagnetism is a major branch of modern geophysics. It emerged as a science after centuries of navigational use by seafarers of the magnetic compass, the invention of which has been compared in its historical importance to that of gunpowder and printing. The magnetic compass is not a perfect instrument, for its magnetized needle deviates in alignment from true North (based on the Earth's rotation axis) at a typically small but significant 'declination' angle, usually denoted by D . According to the 1980 global chart showing isolines ('isogonals') of D , the 'agonic' line (where $D=0$) then passed through North America and South America and also

through Australia, Malaysia and the Western Pacific; D was generally negative in the longitudes of Europe, Africa and Asia, positive over most of the Pacific Ocean, and typically less than 20° in magnitude over most of the Tropics and mid-latitude regions.

A key development came with the sixteenth-century discovery in Europe that a magnetized needle allowed to swing freely in a *vertical* as well as a horizontal plane would point below the horizon at the so-called 'dip' or 'inclination' angle I . In 1980 I was roughly $+55^\circ$ in Rome, $+65^\circ$ at the higher latitudes of London and Paris, and negative over most of the geographic southern hemisphere. By definition I is $+90^\circ$ at the 'magnetic north pole' (then, as now, located in northern Canada), -90° at the 'magnetic south pole' (located in Antarctica), and is everywhere zero on the 'magnetic equator'. This closed line in 1980 was confined to the geographic latitude belt extending from *ca.* $+15^\circ$ to *ca.* -15° and cut the geographic equator at two points, one in the Atlantic Ocean and the other in the Pacific Ocean.

Inspired by the sixteenth-century discovery of the magnetic dip, physician W. Gilbert of Colchester undertook systematic experiments on the properties of lodestone and other magnetic materials. His wide-ranging findings were reported in 1600 in *De Magnete* (full title *De Magnete, Magneticisque Corporibus, et de Magno Magnete Tellure (Concerning Magnetism, Magnetic Bodies, and the Great Magnet Earth)*), a treatise said to have influenced Galileo and Kepler and which one leading historian of science declared to be the first truly scientific text book (see Gilbert, W., *De Magnete*, Gilbert Club revised English translation, London: Chiswick Press (1900)). Gilbert's conclusion that the agency responsible for aligning the compass needle must reside not in the heavens but within the Earth itself quickly gained acceptance. But the nature of the agency and its exact location remained mysterious until the twentieth century, when geophysicists inferred from seismological and other evidence that the Earth has a metallic (mainly iron) core with a solid inner part and liquid outer part, and eventually became persuaded, mainly on general quantitative grounds, that the MGF must be a manifestation of ordinary electric currents within the core maintained by electromotive forces due to motion-induced induction involving flow in the outer core [7-12].

Research on this 'self-exciting MHD geodynamo' benefited from twentieth-century discoveries in terrestrial palaeomagnetism and planetary science. In palaeomagnetism systematic investigations of fossilized magnetization of sedimentary and igneous rocks revealed (i) that the MGF must have existed since soon after the Earth was formed some 4600 MY ago, and

(ii) (as we have seen in section 1 above) that at irregular intervals ranging in duration from *ca* 0.25MY to *ca* 30 MY there have been many polarity changes, each taking no more than a few thousand years to be accomplished [6, 7]. In planetary science ground-based radio-astronomical observations and magnetic data collected near the major planets by space-craft equipped with magnetometers revealed that Jupiter and Saturn possess magnetic fields aligned nearly (anti-) parallel with their respective rotation axes, and that Uranus and Neptune possess magnetic fields aligned at substantial angles – more than 40° – to their rotation axes [10, 22].

Hints that the MGF is not a steady phenomenon were already appearing by the time of Gilbert's death in 1603. Accurate measurements showing that D in London decreased more or less steadily from its value of $+11.25^\circ$ in 1571 to $+4.05^\circ$ in 1634 provided the first evidence of what later became known as the 'geomagnetic secular variation' (GSV), a spatially and temporally complex global phenomenon characterized by timescales of decades and centuries. These modest but unexpected temporal changes were announced in 1635 by the professor of astronomy at Gresham College, H. Gellibrand, in a treatise entitled *A Discourse Mathematical on the Variation of the Magneticall Needle*. For practical navigational purposes they implied that charts of D would have to be revised from time to time and carry predictions indicating future changes. A recent Ordnance Survey map of London indicates that in 2003 the local value of D was -3.5° and its time rate of change was $+0.15^\circ$ per year.

Amongst those who became interested in the GSV were E. Halley and K.F. Gauss, both famous in astronomy for impressive predictions based on Newton's theory of orbital motions in the Sun's gravitational field – Halley of the return in 1758 of the comet that bears his name, and Gauss of the orbit of Ceres, the first asteroid to be discovered (in 1801), by applying his now widely-used 'method of least squares' to observations covering a geocentric arc of no more than three degrees. Their efforts in geomagnetism greatly furthered the systematic acquisition and analysis of data, in Gauss's case including measurements of the intensity at a network of stations, thereby laying foundations for modern research on the MGF.

The practice of presenting geomagnetic data in the form of 'contour maps' was invented for the purpose by Halley and goes back to the publication in 1701 of his chart showing isolines of constant D for the Atlantic Ocean. The practice of representing the MGF mathematically, as the sum of contributions made by hypothetical multipoles (dipole, quadrupole, octupole, etc.) located at the centre of the Earth, goes back to the publica-

tion in 1838 of Gauss's book *Allgemeine Theorie des Geomagnetismus*, in which the magnetic field is expressed as the gradient of a potential function expanded as an infinite series of spherical harmonics [C].

According to the data thus analyzed [9], in the spherical harmonic expansion of the MGF the dominant term corresponds to the hypothetical centred dipole aligned with the Earth's rotation axis, with an amplitude exceeding those of the equatorial components of the dipole and of all other (i.e. 'non-dipolar') terms in the expansion by at least a factor of five. But it is the 'non-dipolar' part of the MGF that undergoes the most rapid secular changes. On typical magnetic maps, lines of equal annual change of D or any other element ('isopors') form sets of oval curves surrounding points at which the changes are most rapid (isoporic foci). A typical set of isopors covers an area of continental size and is separated from neighbouring sets by regions over which changes are comparatively small.

A substantial fraction (about a half) of the GSV can be accounted for in terms of a general westward drift of the non-dipole component of MGF at about 0.18 degrees of longitude per year. The westward drift of the MGF (first noted by Halley from the movement in longitude of the points where the magnetic equator intersects the geographic equator) is a rough measure of typical speeds of fluid motions in the Earth's liquid outer core, somewhat less a millimetre per second (several kilometres per year). There is a general (but not of course detailed) resemblance between global magnetic maps and global maps contouring pressure or other meteorological elements in the Earth's atmosphere, where typical wind speeds are several metres per second. So a century of detailed geomagnetic observations is roughly equivalent, from a dynamical point of view, to no more than a few days of meteorological observations, a useful but hardly excessive data set for detailed scientific studies.

Owing to the twentieth-century decline in the use of the magnetic compass associated with the introduction of new navigational aids such as the gyrocompass and, more recently, the Global Positioning System (GPS) based on signals from a swarm of orbiting artificial satellites, the practical need for determinations and predictions of changes of the MGF has virtually disappeared. But the geomagnetic observations obtained originally for navigational purposes are now exploited by geophysicists undertaking research on the dynamics and structure of the Earth's deep interior. This continuing development has for the past twenty years been facilitated by the International Union of Geodesy and Geophysics under a successful special programme with the acronym SEDI (Study of the Earth's Deep Interior) [6].

The irregular fluctuations of modest amplitude seen in the MGF on GSV timescales, decades and centuries, largely reflect the continual rearrangement (rather than creation or destruction) of magnetic lines of force by the mainly horizontal non-steady motions near the top of the core, in accordance with Alfvén's 'frozen flux' theorem [8; 16; C] (see also Jackson, A., Constable, C.G., Walker, M.R. & Parker, R.L., 'Models of the Earth's magnetic field incorporating flux and radial vorticity constraints', *Geophys. Journ. Inter.*, (submitted); Roberts, P.H. & Glatzmaier, G.A., 'A test of the frozen-flux approximation using a new dynamo model', *Phil. Trans. R. Soc. Lond.* A358, 1109-1121 (2000)). These motions should include a general class of geophysically and astrophysically important 'magnetostrophic' oscillations associated with restoring forces equivalent to the combined effects of Lorentz forces and Coriolis forces when these forces oppose each other. (See Braginsky, S.I., 'Magnetic waves in the Earth's core', *Geomag. & Aeron.* 7, 851-859 (1967); Hide, R., 'Free hydromagnetic oscillations of the Earth's core and the theory of the geomagnetic secular variation', *Phil. Trans. R. Soc. Lond.* A259, 614-647 (1966); also [10, 11, 12, 16]). Any creation or destruction of magnetic field lines occurs on the longer timescales characteristic of processes involving ohmic diffusion and dissipation. Consistent with this picture are polarity reversals and excursions of the MGF, each taking thousands of years [6, 7, 8].

C. INFERENCES FROM EQUATIONS OF ELECTROMAGNETISM

We assume here a simple spherically-symmetric model of the Earth's interior comprising an electrically conducting 'core' of radius 3485 km surrounded by an electrically-insulating 'mantle' extending out to the surface of the model Earth, radius 6371 km. Underlying the liquid outer part of the conducting core is the solid 'inner core' of radius 1220 km. Observations of the MGF are confined to regions at and near the Earth's surface, but with the aid of the equations of electromagnetism it is possible to make useful inferences about the magnetic field within the Earth, especially at the core-mantle boundary (CMB).

Supposing that the MGF is produced entirely by electric currents within the core we take the magnetic permeability μ to be everywhere equal to that of free space and the electrical conductivity σ to be non-zero and constant throughout the core and zero throughout the mantle. The concepts and ideas developed by nineteenth-century physicists from their discover-

ies in electromagnetism can be applied to an electrically-conducting fluid such as the Earth's liquid outer core in motion at time t with Eulerian flow velocity $\mathbf{u}(\mathbf{r}, t)$ at general point P at position \mathbf{r} in the chosen reference frame. They give the following nonlinear partial differential equation relating \mathbf{u} and $\mathbf{B}(\mathbf{r}, t)$, the magnetic field, at P:

$$\partial\mathbf{B}/\partial t - \text{curl}(\mathbf{u}\times\mathbf{B}) = (\mu\sigma)^{-1} (\text{divgrad})\mathbf{B}. \quad (\text{C1})$$

This is obtained by eliminating \mathbf{E} and \mathbf{j} from the equations expressing the (pre-Maxwellian) laws of Gauss, Ampère, Faraday and Ohm, respectively:

$$\text{div}\mathbf{B} = 0, \quad \text{curl}\mathbf{B} = \mu\mathbf{j}, \quad (\text{C2, C3})$$

$$\partial\mathbf{B}/\partial t + \text{curl}\mathbf{E} = 0, \quad \mathbf{j} = \sigma(\mathbf{E} + \mathbf{u}\times\mathbf{B}), \quad (\text{C4, C5})$$

where \mathbf{E} is the electric field at P and \mathbf{j} the electric current density.

A general result of interest in connection with the interpretation of polarity reversals is that for every solution (\mathbf{u}, \mathbf{B}) of these equations there is a corresponding solution $(\mathbf{u}, -\mathbf{B})$ if the boundary conditions are independent of the sign of \mathbf{B} . This 'magnetic symmetry' is also a property of the full MHD equations, for the Lorentz force (per unit volume), $\mathbf{j}\times\mathbf{B}$, in the equations of mechanics does not change when \mathbf{B} changes sign.

According to equations (C2) & (C3), \mathbf{B} may be split into two parts, a 'toroidal' part for which the radial component, B_r , is everywhere equal to zero, and which cannot exist in an insulator such as the mantle, and a 'poloidal' part for which B_r is generally non-zero and can exist anywhere. The toroidal part of \mathbf{B} (which is associated with the *poloidal* part of \mathbf{j}) within the conducting core must vanish at the CMB, but its average strength throughout the core is likely to be higher than that of the poloidal part (associated with the *toroidal* part of \mathbf{j}), implying that much and probably most of the Earth's magnetic energy $((2\mu)^{-1} B^2$ per unit volume) is associated with field lines that are confined to the core and cannot therefore be observed directly at the Earth's surface [9].

The poloidal part of \mathbf{B} possesses field lines that cross the CMB, with a structure which can, in principle, be inferred throughout the insulating 'mantle' from determinations of the MGF at the Earth's surface. The extrapolation procedure makes use of (i) the general *solenoidal* character of \mathbf{B} (i.e. $\text{div}\mathbf{B}=0$ everywhere owing to the absence of magnetic monopoles, see equation (C2)) and (ii) its *irrotational* character (i.e. $\text{curl}\mathbf{B}=0$) in insulating regions, where $\mathbf{j}=0$ (see equations (C3) & (C5)). Then it is possible to

express \mathbf{B} as the gradient of a scalar potential, $-V$ (say), satisfying Laplace's equation $\text{div grad}V = 0$.

From the solenoidal character of \mathbf{B} alone it is possible to infer general topological features of the pattern formed by intersections of the field lines of \mathbf{B} with a general closed surface such as a spherical surface S concentric with the centre of our model Earth. Thus, on any S there must always be one or more closed C-lines (null-flux lines) – defined as the loci of points where the normal component of \mathbf{B} , namely B_r , vanishes, which separate regions where $B_r > 0$ on S from regions where $B_r < 0$. Moreover, the pattern must include one or more pairs of D-points where only the normal component of \mathbf{B} is non-zero. In some but not all cases, depending on the behaviour of the non-radial components of \mathbf{B} on C-lines and in the vicinity of D-points (features which are related through the 'hairy ball' theorem of topology), there may also be 'touch points' T (say) on C-lines, where the non-radial components of \mathbf{B} are tangential to the C-line. And there may be neutral points, N, where all components of \mathbf{B} are zero.

Consistent with what would be expected of a field pattern at levels lying well above the source of the field, in 1990 the MGF at the surface of the Earth typically exhibited just one C-line, the magnetic equator, one pair of D-points, the north and south magnetic poles, and no T(touch)-points. The extrapolated field increased in complexity (and strength) with depth and according to one study the corresponding field pattern at the CMB exhibited 4 C-lines 'nested' with the main magnetic equator and 3 'non-nested' C-lines, 13 pairs of D-points additional to the main pair of magnetic poles, and 6 pairs of T-points (Hide, R., Barraclough, D.R. & McMillan, S., 'Topological characteristics of magnetic and other solenoidal vector fields', *Annales Geophysicae* 15 (Supp. 1), page C122, (1997)). The extent to which some of these features are characteristic of magnetic fields produced by self-exciting dynamo action is probably worthy of further study, for we know [8,9,10] that such action cannot maintain a field possessing an axis of symmetry, which would have neither touch points nor non-nested C-lines. When in due course geodynamo computer modellers undertake detailed studies of the long-term behaviour of the MGF, as a useful preliminary they will undoubtedly investigate whether topological features change in characteristic ways during polarity reversals and excursions.

Denote by $N(S; t)$ the number of intersections of field lines of $\mathbf{B}(\mathbf{r}, t)$ with S at time t , as given by the surface integral of $|B_r|$ over S (the corresponding surface integral of B_r being zero in virtue of equation (C2)). Determinations of N at the bottom of the mantle based on extrapolations of the surface

MGF using the best geomagnetic data available, including those obtained in 1980 and 2000 from orbiting magnetometers on the Magsat and Oersted artificial satellites respectively, provide convincing evidence that the GSV largely involves the continual re-distribution of field lines (rather than their creation and destruction) by motions in the upper reaches of the core [9; B]. This result is understandable for, according to equation (C1), on timescales much less than $\mu\sigma L^2$ (where L is a characteristic length scale), which is several thousand years when L is comparable with the core radius, the core behaves like a perfect conductor, for which it is impossible to change the total linkage of magnetic field lines, as measured by N . Indeed, a novel method for determining the radius of the electrically-conducting core from magnetic observations alone – by finding that level where N is independent of t – gives values close (to within about 2%) of the accepted value based on the more accurate methods of seismology (see [8, 9], also Hide, R., ‘How to locate the electrically-conducting fluid core of a planet from external magnetic observations’, *Nature* 271, 640-641 (1978)).

The magnetic field in a perfect conductor satisfies Alfvén’s ‘frozen magnetic flux’ theorem

$$\partial\mathbf{B}/\partial t - \text{curl}(\mathbf{u}\times\mathbf{B}) = 0, \quad (\text{C6})$$

to which equation (1) tends when the ‘magnetic Reynolds number’,

$$R = UL\mu\sigma, \quad (\text{C7})$$

is very large, U being a characteristic flow speed. On the timescales over which this equation holds within the core the structure of \mathbf{B} at the CMB should be such that each C-line, as it moves and suffers distortion under the influence of the horizontal flow just below the CMB, retains both its separate identity and its relationship with dip poles and touch points. Any touch points present move with that flow, but the determination of the flow at other points requires additional information [8, 9]. This comes from the equations of motion expressing the laws of mechanics, which show that just below the CMB (where the toroidal part of \mathbf{B} is much weaker than at deeper levels within the core), large-scale flow is mainly ‘geostrophic’, with Coriolis forces in balance with horizontal pressure gradients (as are large-scale flows in the atmosphere and oceans) (LeMouél, J.-L., Gire, C. & Maden, T., ‘Motions at core surface in the geostrophic approximation’, *Phys. Earth Planet. Interiors* 39, 270-287 (1985)).

Equations (C1) & (C6) lead to one useful general prediction, which we note here in passing. When R is large, Alfvén's theorem (equation (C6)) holds *nearly* everywhere within the fluid. But acceptable mathematical solutions must be such that there exist elsewhere within the fluid localized regions where \mathbf{B} changes on length scales that are so much less than L that the right-hand side of equation (C1) is *not* negligible. Otherwise it would be impossible to satisfy all the necessary boundary conditions, for the order of equation (C6) is lower than that equation (C1). (Reasoning along similar lines, it follows from the equations of hydrodynamics that necessary concomitants of geostrophic flow in rapidly rotating fluids are detached shear layers, a finding which has been amply confirmed by laboratory experiments and by the discovery of frontal systems in atmospheres and oceans (see Hide, R., 'Experiments with rotating fluids; Presidential Address', *Quart. J. R. Meteorol. Soc.* 103, 1-28 (1977)).

We conclude this appendix with inferences from electromagnetism concerning the generation of the MGF by the subtle self-exciting dynamo process, which occurs on timescales for which the right-hand side of equation (C1) cannot be neglected and involves the creation (and destruction) of field lines. The value of R must be high enough for efficient amplification of the strength of the field within the core by motional induction (as represented by the term $\text{curl}(\mathbf{u} \times \mathbf{B})$), which has to overcome ohmic decay (as represented by $(\mu\sigma)^{-1}(\text{divgrad})\mathbf{B}$). But R must not be so high that the diffusion of the field from the core into the mantle is unduly inhibited, for the flux linkage of a perfect conductor cannot be changed.

In my own work I have found it useful to define dynamo action in terms of N at the CMB (rather than in terms of the magnetic energy or the equivalent magnetic moment of the system), requiring only that this quantity should not tend to zero as t tends to infinity (Hide, R., 'The magnetic flux linkage of a moving medium: a theorem and geophysical applications', *J. Geophys. Res.* 86, 11,681-11,687 (1981)). With this definition it is possible to express criteria for dynamo action in terms of the structure of \mathbf{B} just below the CMB. It is also possible to show that irrespective of the compressibility of the fluid no steady or fluctuating magnetic field can be maintained by motional induction if the field possesses an axis of symmetry (R. Hide & T.N. Palmer, 'Generalization of Cowling's theorem', *Geophys. Astrophys. Fluid Dyn.* 19, 301-309, (1982)). This is an extension of a 'non-existence' theorem due to T.G. Cowling [10], published in 1934 as an attempted rebuttal of J. Larmor's original suggestion [22] that sunspot magnetic fields are produced and maintained by dynamo action [10]. Not until the late 1950s did

the first *existence* theorems (for classes of magnetic fields that do *not* possess an axis of symmetry), by Backus and Herzenberg [8-10], make their appearance in the literature. According to these theorems, there must exist (for values of R that are neither too small nor too large) non-decaying configurations of \mathbf{u} and \mathbf{B} that satisfy equation (C1). Finding and investigating configurations that are also consistent with the laws of mechanics and thermodynamics is the principal aim of dynamo theory [10].