GEODYNAMO SIMULATIONS—HOW REALISTIC ARE THEY?

Gary A. Glatzmaier
Department of Earth Sciences, University of California, Santa Cruz, California 95064; e-mail: glatz@es.ucsc.edu

Key Words  core convection, geomagnetic field, computer models

Abstract  The past seven years have seen significant advances in computational simulations of convection and magnetic field generation in the Earth's core. Although dynamically self-consistent models of the geodynamo have simulated magnetic fields that appear in some ways quite similar to the geomagnetic field, none are able to run in an Earth-like parameter regime because of the considerable spatial resolution that is required. Here we discuss some of the subtle compromises that have been made in current models and propose a grand challenge for the future, requiring significant improvements in numerical methods and spatial resolution.

1. INTRODUCTION

A dynamically self-consistent convective dynamo model is a system of equations that, when solved, simulates the maintenance of magnetic field against diffusion by convection of an electrically conducting fluid. Here we are interested in convection and magnetic field generation in the Earth's core, i.e., the geodynamo. The system of equations represents conservation of mass, momentum, energy, and magnetic flux, an equation of state, and the magnetic induction equation. A computer solves this system of equations, one numerical time step at a time for typically millions of time steps, to provide the time evolution of the three components of the fluid velocity, the three components of the magnetic field, and the thermodynamic perturbations in a three-dimensional (3D), rotating, spherical shell of electrically conducting fluid. The numerical solutions are self-consistent in the sense that all of these variables are part of the 3D time-dependent solution and that convection and magnetic field generation are driven only by the heat flux through the boundaries. The induced magnetic field, for example, is free to decay away if the structure and amplitude of the fluid flow are not adequate to maintain it.

Although the term geodynamo model is used here to refer to spherical dynamo models currently being run as simple analogues of the geodynamo, no one has yet been able to simulate a really “Earth-like” geodynamo because of the huge computing resources that are required. However, considerable progress has been
made during the past decade with self-consistent models run in more accessible parameter regimes.

Geodynamo simulations are conducted to investigate the structure and time dependence of convection and magnetic field generation in the Earth’s core. Because so little can be observed, other than the surface field (today’s field in detail and the paleomagnetic field in much less detail) and what can be inferred from seismic measurements, geodynamo models are used as much to predict what has not been observed as they are used to explain what has. The underlying assumption is that when a computer simulation generates a magnetic field that, measured at the model’s surface, looks qualitatively similar to the Earth’s surface field in terms of structure, intensity, and time dependence, then it is possible that the 3D flows and fields inside the model core could be qualitatively similar to those in the Earth’s core. Of course, the more geophysically realistic the model and the greater the numerical resolution, the greater our confidence that we are not getting the right answer for the wrong reason.

Much has already been written on this subject in recent review articles. For example, Hollerbach (1996), Glatzmaier & Roberts (1997), Fearn (1998), Busse (2000), and Roberts & Glatzmaier (2000a) review the theory of the geodynamo, describe the physics that is represented in geodynamo models, and illustrate results from various models. Dormy et al. (2000) review the current models in more detail and compare their results to geomagnetic and paleomagnetic observations. Zhang & Schubert (2000) emphasize still other aspects of dynamo modeling. Detailed statistical comparisons with magnetic observations have also been made by Bloxham (2000a), Coe et al. (2000), Kono et al. (2000), and McMillan et al. (2001).

Instead of reviewing this research again, I touch on some important issues in geodynamo modeling that seldom appear in print. There is not enough room to go into great detail on any of these issues, so I just briefly describe them with the hope of providing those readers who have not developed such models some insight and appreciation for the choices and challenges faced by those who have. After a brief history of 3D spherical dynamo models, I discuss issues regarding the treatment of turbulent diffusion, hyperdiffusion, and boundary conditions in these models. Then I describe some of the choices that need to be made when designing a numerical method to solve the system of equations and a numerical benchmark to validate the codes. In the last section I discuss a proposed grand challenge to someday achieve an “Earth-like” geodynamo simulation.

2. BRIEF HISTORY

The first 3D convective dynamo model (Gilman & Miller 1981) was developed to study the solar dynamo. This pioneering work two decades ago modeled an electrically conducting fluid with a constant background density using the Boussinesq approximation. The first dynamo model that accounted for the large density
stratification appropriate for the solar convection zone (Glatzmaier 1984) employed the anelastic equations of motion instead of the Boussinesq.

Global 3D dynamo models designed to study the Earth's core were developed a decade later (Glatzmaier & Roberts 1995a, Kageyama & Sato 1997, Kuang & Bloxham 1997, Kida et al. 1997, Busse et al. 1998, Christensen et al. 1998, Sakuraba & Kono 1999, Katayama et al. 1999, Hollerbach 2000, Christensen et al. 2001). Simpler and more affordable models that are not globally 3D or not fully time dependent have also provided valuable insights (Cuong & Busse 1981, Zhang & Busse 1988, Jones et al. 1995, Kageyama et al. 1995, Sarson et al. 1997); however, they are accurate only for weakly driven convection. The Glatzmaier-Roberts dynamo model, a modified version of the 1984 solar dynamo model, was first used to produce a Boussinesq simulation (Glatzmaier & Roberts 1995a,b) and later anelastic simulations (Glatzmaier & Roberts 1996a,b). The Kageyama-Sato model (Kageyama & Sato 1997, Kageyama et al. 1999) is similar in some ways to the original solar dynamo models; for example, the fluid is modeled as a perfect gas instead of a liquid metal. It also solves the fully compressible equations, which requires temporal resolution of sound waves. Currently all other models make the Boussinesq approximation. Most dynamo models have simulated several magnetic diffusion times (hundreds of thousands of years) and have produced self-excited magnetic fields that, when observed at the surface of the model Earth, have an intensity, an axial dipole dominated structure, and a westward drift of the nondipolar structure that are all qualitatively similar to the Earth’s.

For example, a snapshot of the patterns of flow, field, and temperature from one of the solutions produced by Christensen et al. (1999) is illustrated in Figure 1. Similar field structure is seen in Figure 2 from a simulation by Kuang & Bloxham (1997). A snapshot (Figure 3) of the 3D structure of the magnetic field simulated by Glatzmaier & Roberts (1995a) shows the field stretched around the “tangent cylinder” (an imaginary cylinder tangent to the inner core equator) by the shear in the zonal flow.

A higher resolution simulation (Roberts & Glatzmaier 2000a) is illustrated in Figure 4. It shows a snapshot of the radial component of the simulated field both at the surface of the model Earth and at the core-mantle boundary (CMB). For comparison, the geomagnetic field is also shown at the surface and projected down to the CMB, assuming an insulating mantle. It is plotted only out to spherical harmonic degree 12 because the higher degrees of the measured surface field are “contaminated” by crustal field. The simulated field, when also plotted only out to degree 12, looks qualitatively similar, in both structure and intensity, to the degree-12 filtered geomagnetic field. However, when the full resolution is plotted the pattern of radial field at the CMB is significantly more structured, with many intense small-scale field concentrations. It is unrealistic to assume that the energy spectrum of the Earth’s core-generated field happens to drop off dramatically at roughly the same spherical harmonic degree that crustal magnetism becomes comparable to the core field at the Earth’s surface. It is much more likely
that the structure of the geomagnetic field at the CMB is instead more like the simulated field shown in Figure 4 than the degree-12 pattern of the measured geomagnetic field. In fact, because even this simulation is too diffusive, it is likely that the Earth’s CMB field has much finer structure and, possibly, intense small-scale “core spots” not unlike sun spots on the solar surface. Any analysis based on the degree-12 limited pattern of the geomagnetic field at the CMB is capturing the effects of the large global-scale fields but missing the intense small-scale effects.

Some models (e.g., Glatzmaier & Roberts 1995b, Glatzmaier et al. 1999, Kageyama et al. 1999) occasionally produce spontaneous reversals of the magnetic dipole polarity between long, relatively stable, nonreversing periods on timescales similar to that seen in the paleomagnetic record. The Glatzmaier-Roberts simulations also produce incomplete reversals much more frequently than full reversals. As described in Glatzmaier & Roberts (1995b), reversals fail to complete in 3D computer simulations when the field reverses only outside the tangent cylinder and then quickly reverses back to the original polarity because the convective timescale of the magnetic instability is usually much shorter than the time required for the original field to diffuse out of the solid inner core. The stability of the magnetic dipole that an electrically conducting inner core provides, relative to that provided by an insulating inner core, was first demonstrated by Hollerbach & Jones (1993) with a mean-field (2D) calculation. It showed that a conducting inner core can prevent a very time-dependent magnetic field from reversing its polarity. These 2D and 3D simulations provided the basis for an explanation (Gubbins 1999, Glatzmaier et al. 1999) of 14 geomagnetic excursions that were reported (Lund et al. 1998) to have occurred since the last complete geomagnetic reversal 780,000 years ago.

The models that have finite conducting inner cores also allow them to freely rotate according to the net time-dependent torque on them. As a result, the inner core oscillates relative to the model’s surface. In some simulations there is also an average eastward drift, a “super rotation” (e.g., Glatzmaier & Roberts 1995a, Sakuraba & Kono 1999, Christensen et al. 2001). This can be explained (e.g., Glatzmaier & Roberts 1996b, Aurnou & Olson 2000) as a thermal wind in the fluid outer core magnetically coupled to the solid inner core. Several seismic studies (e.g., Song & Richards 1996) subsequently supported this prediction, although estimates of the amplitude of the inner core super rotation have decreased by an order of magnitude as more sophisticated analyses of this difficult measurement have been made. The model-predicted amplitude also decreases as greater spatial resolution is used and gravitational coupling between the mantle and inner core is included (Buffett & Glatzmaier 2000). Because the model-predicted period of oscillation is of order 75 years, it will be a long time before seismic observations are able to tell us what the mean rotation rate is; some models see no net rotational drift of the inner core (Kuang & Bloxham 1999).

Other dynamo studies investigate the effects of the size of the solid inner core (Sakuraba & Kono 1999, Bloxham 2000b, Morrison & Fearn 2000, Roberts & Glatzmaier 2001), a stably stratified layer at the top of the core (Glatzmaier & Roberts 1997), a nonuniformly prescribed heat flux through the CMB (Sarson
et al. 1997, Glatzmaier et al. 1999, Bloxham 2000a,b), different velocity boundary conditions (Kuang 1999, Christensen et al. 1999), and parameter space (Olson et al. 1999, Christensen et al. 1999, Grote et al. 2000a).

Outside the core, the simulated fields produced by the various models appear similar to the geomagnetic field. Inside the model core, however, the field morphologies can be quite different. The form of the equations (e.g., Boussinesq or anelastic) and the numerical method (e.g., spectral or finite difference) differ among the models. However, there are three major reasons for the differences in the solutions: different treatments of the inner core, different boundary conditions, and, most significantly, different specified parameters.

3. TURBULENT DIFFUSION

3.1 Grand Challenge

Although the simulated fields outside the core look quite similar to the geomagnetic field, the models all have severe deficiencies. For example, whereas viscous, thermal, and magnetic diffusion are represented in all models of the geodynamo, only the magnetic diffusivity (the inverse of the product of magnetic permeability and electrical conductivity) can be set to the appropriate value for the Earth’s core (2 m²/s). Every geodynamo model completely neglects the actual viscous, thermal, and compositional diffusivities of the fluid because these are orders of magnitude smaller (e.g., de Wijs et al. 1998, Braginsky & Roberts 1995). However, because they are so small there must be significant energy in small-scale turbulent eddies. That is, the Reynolds number (the ratio of the viscous diffusion time to the convective time) is roughly 10⁸. This turbulence, which 3D global dynamo models have not yet been able to resolve, is much more efficient in mixing momentum, entropy, and chemical constituents than the molecular processes so it cannot be neglected. All models of the geodynamo crudely approximate this turbulent transport as a simple, isotropic, homogeneous, diffusive process modeled after molecular diffusion but with an enhanced “turbulent diffusivity.” For example, the convergence of the turbulent transport of momentum is modeled as a constant scalar viscous diffusivity multiplied by the Laplacian of the resolved fluid velocity. Similar “large eddy simulations” have been conducted for decades in global circulation models of the Earth’s atmosphere and ocean, which certainly cannot use the small molecular diffusion coefficients of air or water. The existence of turbulent boundary layers in the atmosphere and ocean (e.g., Hunkins 1966) that are orders of magnitude deeper than a laminar molecular Ekman boundary depth would be provides physical justification for employing turbulent diffusivities in the Earth’s fluid core.

However, turbulent transport in the Earth’s core is likely not isotropic, homogeneous, and time independent (Braginsky & Meytlis 1990). Instead, it depends on the local magnetic field, gravity, and rotation vectors. As long as a significant portion of this turbulence is not resolved in the model simulations, the turbulent diffusion of momentum and entropy should be represented by nonlinear, time and
spatially dependent, tensor diffusivities. However, so far all geodynamo models have employed scalar diffusivities, constant in space and time. This would not be so bad if the turbulent viscosity invoked in all current geodynamo models did not need to be, for computational reasons, orders of magnitude greater than the Earth’s magnetic diffusivity.

Certainly in the Earth’s core there is a continuous spectrum of kinetic energy versus length scale, and what is labeled subgrid scale (unresolved) turbulence in a simulation depends on the spatial resolution of that simulation. Modelers plan to increase spatial resolution as greater computing resources become available, to resolve more of this spectrum while reducing the turbulent subgrid scale diffusivities.

A goal that no geodynamo simulation has yet to achieve would be to solve the 3D magnetohydrodynamic equations with the rotation rate, dimensions, density, and heat flux of the Earth’s core and with all diffusivities set to the Earth’s magnetic diffusivity, 2 m²/s. It is possible that, when this proposed grand challenge is achieved, the spatial and temporal resolution required for such a calculation will be sufficient to capture enough of the turbulence spectrum that the isotropic, homogeneous treatment of the remaining unresolved turbulence may no longer be a serious problem.

It is traditional to use a set of nondimensional numbers to describe the parameter space in which a given numerical problem is run. The Rayleigh number measures buoyancy driving (by the imposed thermal boundary conditions, gravity, and thermal expansion coefficient) relative to thermal and viscous diffusion, which inhibit convection. The Prandtl number, $P_r = \nu/\kappa$, is the ratio of the effects of viscous to thermal diffusion, or in other words, the ratio of a characteristic thermal diffusion time to a viscous diffusion time. The ratio of the effects of thermal to magnetic diffusion is called the Roberts number, $q = \kappa/\eta$. The Ekman number measures the effects of viscosity relative to rotation, $E_k = \nu/(2\Omega D^2)$, and the magnetic Ekman number, $E_m = \eta/(2\Omega D^2)$, is the ratio of magnetic diffusion to rotational effects. Here, $\nu$, $\kappa$, $\eta$, $\Omega$, and $D$ are the turbulent viscous and thermal diffusivities, the actual magnetic diffusivity, the mean angular rotation rate of the Earth, and the depth of the fluid outer core, respectively. For the grand challenge mentioned above, we have $P_r = q = 1$ and $E_k = E_m = 10^{-9}$; the Rayleigh number would be orders of magnitude larger than the critical value needed for the onset of convection.

### 3.2 Compromises for Current Models

The main computational challenge is resolving the small-scale flow structures that occur with this small turbulent Ekman number. The depth of the viscous boundary layer, for example, is roughly $E_k^{1/2}D$. A Chebyshev grid in radius, employed in several models, smoothly concentrates radial grid levels near the boundaries, providing greater spatial resolution where it is needed. Because the number of grid levels within a given small distance from the boundary increases as the square of the total number of Chebyshev levels, the total number needed to resolve the Ekman boundary layer is proportional to $E_k^{-1/4}$. For example, to decrease the Ekman
number from $10^{-5}$ to $10^{-9}$, while maintaining the same number of grid levels within the boundary layer, requires only a tenfold increase in the total number of Chebyshev levels. However, corresponding increases in the horizontal resolution (e.g., number of spherical harmonics) will also be required, along with a decrease in the size of the numerical time step.

Current dynamo models have not been run at high enough resolution to reach the goal proposed in the previous section, so compromises had to be made. For comparison’s sake, let us assume $\tilde{\omega}$ and $D$ are Earth values: $7.29 \times 10^{-5}$ s$^{-1}$ and 2260 km, respectively. The smallest $E_k$ that current models are able to reach is approximately $10^{-5}$, so the viscosity, $\nu$, in these models is approximately $10^4$ times larger than the goal of 2 m$^2$/s. That is, the turbulent viscous diffusion time in all models of the geodynamo is at least $10^4$ times shorter than the Earth’s actual magnetic diffusion time.

Two different approaches have been adopted to model the geodynamo in light of this severe limitation. One is to set $P_r = q = 1$, which is desirable because then the characteristic timescales for viscous, thermal, and magnetic diffusion are the same. However, they are all $10^4$ times too short relative to the rotation period and both Ekman numbers are $10^4$ times too large (e.g., Kuang & Bloxham 1997, Christensen et al. 1998). Another way of interpreting this choice of parameters is to choose the magnetic diffusion time as the measure of the real time instead of the Earth’s rotation period. That is, consider all the diffusivities actually set to 2 m$^2$/s, but set the rotation period to 27 years ($10^4$ days). Because 27 years is still short compared with the time required for the global-scale structure of the magnetic field to diffuse significantly [e.g., the dipole diffusion time, $r_{cmb}/(\pi^2 \eta) = 20,000$ years], one could argue that Ekman numbers set to $10^{-5}$ are asymptotically small enough. However, because the traditional definitions of $E_k$ and $E_m$ are based on the viscous and magnetic diffusion times across the depth ($D$) of the outer core, this argument only holds for the large global scales of the motion and field. Yet another way of interpreting this choice (Kuang 1999) is to say that the parameters are set according to the goal above but each time step applies only $10^{-4}$ of the net force to determine the acceleration and Coriolis force on the fluid and inner core.

The alternative approach is to argue that because the two largest forces controlling the dynamics in the Earth’s outer core are the Coriolis and Lorentz, it is important that their respective timescales, i.e., the rotation period and the magnetic diffusion time, have the correct ratio in the simulations; that is, $E_m = 10^{-9}$. This is doable, along with $q = 1$, but at the price of having the Prandtl number orders of magnitude too large (e.g., Glatzmaier & Roberts 1996a). That is, viscosity is still too large but, as in the first approach, one prescribes a superadiabatic heat flux out of the core large enough to drive convective flows against the enhanced viscosity.

So the question is, Is it better to have all diffusion timescales be equal but too short relative to the Earth’s rotation period, or to have the correct magnetic and turbulent thermal diffusion timescales and only the viscous diffusion timescale be too short? Another way of asking this is, Is it better to artificially damp the acceleration and Coriolis forces or to explicitly increase the viscosity? Both avoid
small spatial and temporal scales, but neither is acceptable. It is unlikely that either produces results that closely approximate those that would be obtained with the grand challenge parameters proposed above.

Until that goal is achieved, we can make rough estimates of the relative amplitudes of the two largest forces, the Lorentz and Coriolis forces, by computing from the simulations the magnetic coupling parameter

\[ \frac{B^2}{2\Omega \mu_0 \rho D V} \]

This is the ratio of the Elsasser and magnetic Reynolds numbers. Here, \( \rho \) is the average mass density and \( \mu \) is the magnetic permeability. \( V \) and \( B \) are maximum amplitudes of the fluid velocity and magnetic field, respectively, and therefore this nondimensional number, like the Elsasser and magnetic Reynolds numbers, is a measure of the solution, not a description of the problem. For example, in the Glatzmaier-Roberts simulation illustrated in Figure 4, the maximum fluid velocity is typically \( 5 \times 10^{-3} \) m/s and the maximum field intensity is typically 40 mT, giving a magnetic coupling parameter of 0.08. The actual balance depends on the local amplitudes and directions of the flow (relative to \( \Omega \)) and field (relative to its curl). In addition, \( D \) in the above expression is an overestimation of the characteristic length scale of the field structure inside the core. However, the magnetic coupling parameter provides a more direct estimate of the relative strengths of the two major forces in this problem than does the Elsasser number, \( \frac{B^2}{2\Omega \mu_0 \rho \eta} \), which involves the much longer magnetic diffusion time instead of the convective timescale. For the above example the Elsasser number is approximately 440. A more accurate evaluation would require spatial and temporal averages of the actual Coriolis and Lorentz forces (Kuang 1999).

### 3.3 Hyperdiffusion

In addition to these compromises, some simulations also invoke hyperdiffusion in order to reach somewhat smaller values of the effective \( E_k \) for the large scales. That is, the amplitude of the viscous coefficient is a function of the velocity scale; larger scales experience less viscosity than do the smaller velocity scales. The high viscosity on the smallest resolved scales is needed to dissipate the energy that turbulently cascades down to that truncation level and, in the real world, would continue to cascade to much smaller scales before molecular viscosity became important. Without this high viscosity in the model at the smallest resolved scales, energy would unrealistically build up there and quickly produce a numerical instability.

Similar techniques have long been employed in other fluid dynamical communities; for example, in atmospheric dynamics. The physical justification is that the unresolved turbulent eddies should interact much more effectively with the smallest resolved eddies (because they are of similar size) than with the largest resolved eddies. Preliminary spectral analyses of the nonlinear transfers of kinetic energy in a dynamo simulation support this concept (Buffett & Bloxham 2001).
However, for the current hyperdiffusive dynamo simulations, the wave number dependence of the diffusion coefficients is specified arbitrarily and independent of time and space. When hyperdiffusion is used (e.g., Glatzmaier & Roberts 1995a, Kuang & Bloxham 1997, Sakuraba & Kono 1999), the effective Ekman number for the smallest resolved scales (i.e., at the truncation level) is comparable to the (scale-independent) Ekman number specified in most nonhyperdiffusive dynamo simulations. For example, the model results illustrated in Figure 4 have a viscosity proportional to the square of the spherical harmonic degree, with an effective Ekman number of $2 \times 10^{-6}$ for the largest scales and $2 \times 10^{-5}$ for the smallest (degree 95) scales.

Zhang & Jones (1997) have tested the effects of hyperdiffusion on linear convective stability calculations and Grote et al. (2000b) on fully nonlinear dynamo simulations. Both studies, not surprisingly, show differences between the hyperdiffusive and nonhyperdiffusive cases they tested. This, of course, is the reason for employing hyperdiffusion; one wishes to improve the representation of the large scales by reducing the viscosity on them without having to increase spatial resolution.

Figure 5 schematically illustrates this concept. Three hypothetical cases are displayed, each showing the specified viscosity and the resulting volume-averaged kinetic energy as functions of the spherical harmonic degree. Case $a$ is a low resolution nonhyperdiffusive case, which severely damps all the resolved length scales (spherical harmonic degrees). Case $b$ (also nonhyperdiffusive) uses a smaller viscosity and therefore produces a more realistic energy spectrum, but it requires much greater spatial resolution (more spherical harmonics) to maintain numerical stability. That is, kinetic energy needs to cascade to smaller scales before the Laplacian of velocity is large enough to compensate for the smaller viscous diffusivity. In an attempt to produce more accurate large scales without increasing resolution, one applies the hyperviscosity of case $c$; of course the smallest resolved scales are still poorly represented.

The implied assumption in all these comparative studies is that the simple Laplacian operator on velocity multiplied by a constant scalar diffusion coefficient (i.e., nonhyperdiffusion) is the most “correct” representation; however, there is nothing “sacred” about the Laplacian operator when parameterizing the effects of the nonlinear transport by turbulent eddies. It is the most familiar and one of the simplest approximations, but it does not, for example, account for maintenance of the resolved scales by the nonlinear interactions between the unresolved eddies. The hyperdiffusive algorithms used in the geodynamo community are not tailored to account for this either. Instead they are quite crude and really only designed to compensate numerically for lack of sufficient spatial resolution. Also, the additional viscous heating resulting from hyperviscosity unrealistically affects the energy balance (Roberts & Glatzmaier 2000a). To improve this situation, one could formulate a more physically based, nonlinear, tensor algorithm based on a series of direct numerical simulations of the small-scale turbulence in local regions under various conditions (Matsushima et al. 1999). Alternatively, a second-order closure
Figure 5  Schematic plots of three hypothetical dynamo simulations, each showing the specified viscosity and the resulting volume-averaged kinetic energy as functions of the spherical harmonic degree. Case \( a \) is a low resolution nonhyperdiffusive case. Case \( b \) is also nonhyperdiffusive but uses a smaller viscosity and higher resolution. Case \( c \) is an example of hyperdiffusion at the low resolution.

method could provide estimates of the mean nonlinear effects of the unresolved turbulence. Until dynamo models can be run with sufficient spatial resolution that the effects of the remaining unresolved turbulence are small enough to be adequately modeled as isotropic diffusion, more sophisticated turbulence models are needed.

4. BOUNDARY CONDITIONS

4.1 Thermal and Compositional Boundary Conditions

Most models of the geodynamo impose constant temperature or constant radial heat flux on the inner core boundary (ICB) and core-mantle boundary (CMB). The anelastic, thermal-compositional model of Glatzmaier & Roberts (1996a), on the other hand, is driven only by the prescribed heat flux at the CMB. This diffusive heat flux is the sum of thermal conduction down the adiabatic temperature gradient and a smaller, turbulent (subgrid scale) diffusive heat flux driven by the superadiabatic temperature gradient (i.e., entropy gradient). At the ICB, the entropy, composition, and radial fluxes of heat and light constituent are part of the solution, varying in time and location on the ICB. However, the local fluxes of light constituent and latent
heat are constrained to be proportional to the local cooling and growth rates of the solid inner core, with the constants of proportionality based on the best estimates of the material properties of the core. All of the latent heat released in the process of inner core growth is assumed to be deposited in the fluid core because the turbulent diffusion and convection there are much more efficient than molecular diffusion and possible solid state convection are in the solid inner core. This treatment of the thermal boundary conditions at the ICB also provides a self-consistent, time- and space-dependent record of the growth of the inner core.

The actual radial heat flux at the Earth’s CMB is presumably controlled by the amplitude and pattern of the temperature in the mantle just above the CMB. Temperature perturbations relative to the mean in the fluid core at a given radius are typically no more than 0.001 K, whereas the temperature in the much more viscous mantle varies by hundreds of K. Therefore, where the lower mantle is cooler, the temperature drop across the CMB, and the heat flux, will be greater. The effects of prescribed heterogeneous CMB heat flux boundary conditions have been tested in dynamo simulations. However, because the convective timescale in the mantle is a million times longer than it is in the fluid core, these boundary conditions are treated as being time independent. Several studies investigate how different imposed patterns of CMB heat flux affect the secular variation of the field (Sarson et al. 1997; Glatzmaier et al. 1999; Bloxham 2000a,b; Coe et al. 2000; McMillan et al. 2001) and the frequency of dipole reversals (Glatzmaier et al. 1999). However, one could argue that until an Ekman number of $10^{-9}$ is achieved with a highly supercritical Rayleigh number, it will be uncertain how robust the conclusions from these studies are.

### 4.2 Velocity Boundary Conditions

Because the effective mass flow rates through the ICB due to the growth of the inner core and through the CMB due to penetration of iron into the mantle are relatively small and poorly constrained, all geodynamo models have imposed impermeable velocity conditions at the inner and outer boundaries of the fluid core. The question is what additional velocity conditions should be imposed at these boundaries. The two extreme choices are nonslip (all three components of the flow vanish at the boundaries) and viscously stress free (the radial gradient of the angular velocity vanishes). The former is the physical choice, but because we are forced, for numerical reasons, to use a large turbulent viscosity, it has been argued the latter choice is more appropriate (Kuang & Bloxham 1999). That is, simulated flows may be more realistic if viscous torque were forced to vanish at the boundaries instead of being too large there. If a laminar flow existed along the boundaries, an Ekman boundary layer, based on molecular viscosity, would be only a few centimeters deep; however, such a layer is no more physical in the Earth’s fluid core than in the Earth’s atmosphere or ocean. (For analysis of boundary layer instabilities with magnetic fields, see Desjardins et al. 1999.) Instead, the Earth’s fluid core must have turbulent boundary layers, not unlike the planetary boundary layer at the base.
of the atmosphere or the mixed layer at the top of the ocean. Therefore, the real argument against nonslip boundary conditions in current geodynamo models is that they produce viscous boundary layers approximately two orders of magnitude deeper than a reasonable turbulent layer should be because the Ekman number is at least four orders of magnitude larger than a reasonable turbulent Ekman number.

The argument for nonslip boundary conditions (Roberts & Glatzmaier 2000a) is one of consistency. Because the transport of momentum between fluid parcels within the core by the unresolved turbulence is modeled as diffusion with an enhanced turbulent viscous diffusivity, why not also model the transport of momentum between a fluid parcel and a solid boundary by this same process? This is how the superadiabatic diffusive heat flux at the ICB and CMB is treated in all geodynamo models. That is, because diffusive heat flux in the models is represented as the product of a turbulent thermal diffusivity and the superadiabatic temperature gradient, it represents turbulent diffusion of thermal energy, both within the fluid and to the boundaries. In addition, nonslip boundaries, unlike stress-free boundaries, provide a source of helical flow (Ekman pumping), which helps maintain the magnetic field. Of course, these would be stronger arguments if current models were able to use a more reasonable turbulent Ekman number, like $10^{-9}$.

A comparison of two dynamo simulations with hyperdiffusion, one with nonslip and the other with stress-free boundary conditions, shows significant differences (Kuang 1999). However, a similar comparison, but without hyperdiffusion, shows little difference when the generated magnetic field is strong (Christensen et al. 1999). Until sufficient spatial resolution is affordable to reach much lower Ekman numbers and higher Rayleigh numbers, stress-free boundary conditions may be a reasonable choice, especially when one does not want the effects of the magnetic torque on the inner core or mantle (e.g., Buffett & Glatzmaier 2000) to be altered by exaggerated viscous torques.

Instead of choosing one of the two extreme velocity boundary conditions, one could impose some linear combination of the nonslip and stress-free conditions. One could also attempt to either parameterize or resolve the effects of boundary topography (for topographic and gravitational coupling), the axisymmetric centrifugal bulge (for luni-solar precession of the mantle), or time-dependent non-axisymmetric tidal deformations in the boundaries. However, none of these effects can be studied realistically until the models are able to run with a significantly smaller Ekman number and larger Rayleigh number.

### 4.3 Magnetic Boundary Conditions

It is customary to approximate the Earth’s mantle as a perfect insulator because the electrical conductivity of the mantle is orders of magnitude less than that of the iron-rich core. Therefore the magnetic field at the CMB must match a potential (current-free) field outside the core. The external potential field is a time-dependent part of the solution and is known everywhere beyond the CMB by knowing the radial component of the field and its radial derivative at the CMB. One of the
advantages of expanding in spherical harmonics is that this boundary condition is easily expressed in terms of the spherical harmonic coefficients of the radial field (e.g., Glatzmaier 1984).

The models by Gilman & Miller (1981) and Kageyama & Sato (1997), which do not employ spherical harmonic expansions, force the magnetic field to be purely radial at the outer boundary. It is not obvious how significant this simplification is, but it does mean that part of the field is maintained by whatever time-dependent distribution of electric currents external to the model is needed to maintain this boundary condition.

Typically, one wishes to allow for magnetic coupling between the fluid core and the lowermost mantle, which is assumed to have a somewhat greater concentration of iron. This is implemented in at least two different ways. Glatzmaier & Roberts (1995a, 1996a) use a shallow conducting layer approximation that results in an additional linear term in the magnetic boundary condition, proportional to the total conductance of the shallow layer. Alternatively, Kuang & Bloxham (1997, 1999) have extended their computational domain for the magnetic field into the lower mantle, allowing only magnetic diffusion there.

The conditions at the ICB depend on the assumed electrical conductivity of the solid inner core. Assuming it is an insulator results in a simple (spherical harmonic) boundary condition similar to that at the CMB. Assuming it is a perfect conductor (as was done in the original solar dynamo models) also results in a simple boundary condition. However, the realistic assumption is that the inner core has finite conductivity, which is usually taken to be equal to that of the fluid core, although it probably should be somewhat higher. This requires extending the computational domain for the magnetic field to include the entire inner core. All three components of the magnetic field and the tangential component of the electric field are continuous across the ICB. This is simple to do with spherical harmonic expansions, especially if nonslip velocity boundary conditions are imposed. Tangentially stress-free velocity boundary conditions require nonlinear magnetic boundary conditions at the ICB because the tangential velocity is discontinuous across the boundary.

Symmetry conditions also exist on the magnetic field at the origin (radius = 0) when a spherical coordinate system is used. These are typically accommodated by defining new magnetic field variables in terms of powers of radius that depend on the degree of the spherical harmonic coefficient (Hollerbach & Jones 1993, Glatzmaier & Roberts 1995a, Hollerbach 2000).

5. NUMERICAL METHODS

5.1 Spherical Harmonic Expansions

The models by Gilman & Miller (1981) and Kageyama & Sato (1997) employ finite difference methods; all other 3D spherical models of the geodynamo currently use spherical harmonic expansions. Glatzmaier (1984), Tilgner (1999), and
Hollerbach (2000) describe a spherical-harmonic poloidal-toroidal decomposition method with Chebyshev collocation in radius. The poloidal-toroidal decomposition takes advantage of the divergence-free mass flux (or velocity for the Boussinesq approximation) and magnetic flux and reduces six functions, the three components of the flow and field, to four, the poloidal and toroidal parts of the flow and field. This ensures that conservation of mass and magnetic flux are exactly satisfied everywhere at all times. These poloidal and toroidal scalars, together with the thermodynamic variables, are then expanded in Chebyshev polynomials to resolve their radial structure and in spherical harmonics to resolve their horizontal structure. The finite set of spherical harmonics actually used can be chosen in many ways; the two most popular are the triangular and rhomboidal truncations.

The nonlinear terms are computed each time step by transforming functions from spectral space to grid space, where the nonlinear products are computed and then transformed back to spectral space where the time step is performed. Transforms in longitude and radius are done with fast Fourier transforms, which make this method feasible at current resolutions. The transforms in colatitude are done via Gaussian quadratures, making this method costly at very high resolutions. The time integration is performed in spectral space using a Courant-limited numerical time step; this means that it needs to be less than the time it takes fluid to flow from one grid point another. When solving the fully compressible equations (Kageyama & Sato 1997) instead of the anelastic or Boussinesq equations, the Courant condition is based on the sound speed instead of the fluid velocity. A “pole problem” usually exists if the model uses spherical coordinates without spherical harmonic expansions because the longitudinal grid spacing becomes vanishingly small as the geographic poles are approached; typically some special filtering or pole exclusion is used in such models (Gilman & Miller 1981, Kageyama & Sato 1997). One of the advantages of the spherical harmonic method is that the horizontal Courant condition is based on an average horizontal resolution (i.e., the highest degree of the spherical harmonics), which is significantly less restrictive. A Courant condition in radius also exists, but the radial velocity becomes vanishingly small near the impermeable boundaries where the Chebyshev mesh is finest. However, the radial Alfvén velocity (i.e., radial magnetic field) near the boundaries does not vanish and so, for large field intensities, poses a more severe constraint on the time step.

Instead of using a Chebyshev method in radius, some groups have chosen to use a finite difference method, which is easier to implement and requires only local communication on parallel computers. The nonuniform finite difference mesh, however, is usually chosen to be the same Chebyshev grid because this provides much greater radial resolution in the boundary layers where it is needed. The drawback is that two to four times more nodes are required to obtain numerical accuracy comparable to that provided by the spectral Chebyshev collocation method (Orszag 1971, Christensen et al. 2001).

The system of equations is usually integrated in time with a combination of an explicit method, based on the known values of variables at the present and
previous time steps, and an implicit method, based on values and spatial derivatives of the variables at the new step. The implicit part provides numerical accuracy and stability and allows much larger time steps but requires the solution of a set of matrix equations each time step. However, if one treats the Coriolis and nonlinear terms explicitly the remaining linear terms decouple in spherical harmonic degree and order, so each spherical harmonic coefficient can be updated separately with a relatively small matrix operation.

For small Ekman-number convection, the two largest forces in the momentum equation are usually the Coriolis and Lorentz, which approximately cancel locally. The first is linear and the second is nonlinear. Therefore, some groups have developed computer codes that also treat the Coriolis forces implicitly to obtain more accurate solutions and to be able to take larger numerical time steps. But this causes a major complication for spherical harmonic methods compared with the explicit treatment of the Coriolis terms because, for each longitudinal wave number, the spherical harmonic coefficients of the even-degree poloidal functions and the odd-degree toroidal functions are coupled and vice versa. That is, instead of solving a large number of small matrices each time step, now a smaller number of much larger, complex, block-banded matrices need to be solved. This requires much more computer memory for direct matrix inversions or more computer time for a sophisticated iterative scheme. Hollerbach (2000) describes in detail the implicit treatment of the Coriolis forces for the spherical-harmonic-Chebyshev method that is fairly similar to that developed for several other geodynamo models (e.g., Glatzmaier & Roberts 1995a, Kuang & Bloxham 1997).

The system of equations that is solved to conserve mass and momentum is sixth order in radius. However, choices need to be made regarding exactly which equations are used. Equations are needed to update the poloidal and toroidal functions each time step. A poloidal equation can be obtained from the radial component of the momentum equation (MEQ) or from the radial component of the double curl of the MEQ. The pressure perturbation is present in the former; it is not present in the latter for the Boussinesq equations or for the anelastic equations when using the reduced pressure formulation suggested in Braginsky & Roberts (1995). A poloidal equation can also be obtained from the divergence of the MEQ or from just the horizontal part of the divergence of the MEQ; both of these contain the pressure. The former has a higher order radial derivative of the nonlinear advection terms compared with the latter, whereas the latter has a higher order radial derivative in the linear viscous term compared with the former. The toroidal equation is obtained from the radial component of the curl of the MEQ. Most Boussinesq models use only the radial component of the double curl and the radial component of the single curl of the MEQ and do not solve for the pressure. Another choice would be to solve the radial component, the divergence, and the radial component of the curl of the MEQ. Although this latter choice means solving three instead of two equations, the three have only second order radial derivatives of the linear terms, whereas the former has fourth order radial derivatives. To achieve comparable accuracy, one would need more radial resolution for the former. Another disadvantage of the
former, when Chebyshev collocation with a fast Fourier transform is used, is that this fourth order equation is not satisfied on two radial grid levels at each boundary in order to impose the boundary conditions (Tilgner 1999).

The other equations in this system provide time derivatives for the entropy perturbation, the light constituent perturbation, and the magnetic field. These are usually solved first during each time step (using the nonlinear terms computed from the variables at the present and previous time steps) to get updated values of the buoyancy and Lorentz forces, which are then used to solve the MEQ for the velocity (and pressure).

5.2 Code Parallelization

Once a code is written and debugged for a single processor, it needs to be modified to run on many processors simultaneously, i.e., in parallel. Currently, this is the only way to get the large memory and performance required to run with high resolution. The simplest way to do this is to use a shared memory machine, which typically means a relatively small number of processors with a compiler that does most of the parallelism for the user.

However, greater computing power is provided by massively parallel systems for which the user must program all communication among the processors. Dividing a problem among hundreds (or thousands) of processors with a communication scheme that minimizes the time processors wait for others can be more of an art than a science. There are many different ways to parallelize any particular code, but only a small number of these are efficient. It can take considerable time and effort, especially for spectral methods, which require global communication (Glatzmaier & Clune 2000). That is, each processor sends data to and receives data from all the others every time step. The “nearest neighbor” communication of local methods will, at some resolution, make them preferable, even though they require many more nodes than spectral methods to achieve a given accuracy.

5.3 Code Validation

After a numerical method and parallelization scheme are formulated and coded, a dynamo model should be validated before it is put into production. The equations and numerical techniques are so complicated that subtle numerical errors could easily exist without being realized if the “solutions” are not compared with several other codes written by others. The first 3D, spherical dynamo benchmark was recently conducted by Christensen et al. (2001). Fairly “mild” parameters were chosen ($E_k = 10^{-3}$, $P_i = 1$, $q = 5$, and a slightly supercritical Rayleigh number) to get Boussinesq solutions that are steady in frames of reference drifting at constant rates relative to the boundaries. Three cases were tested: rotating but nonmagnetic convection, a dynamo with an insulating and stationary inner core, and a dynamo with a conducting and freely rotating inner core.

The exercise produced good agreement among the six different dynamo models that were compared. Several different spatial resolutions were also tested to check
the convergence rate of spherical harmonics and to compare the convergence rates of Chebyshev spectral and finite difference methods. When these codes begin running at much larger Rayleigh numbers and smaller Ekman numbers, we will need more sophisticated benchmarks that compare time-averaged quantities of the time-dependent solutions.

6. SPATIAL RESOLUTION

Once a model is validated at low spatial resolution, it should be exercised at much greater resolution to study more realistic, highly supercritical, time-dependent parameter regimes. This is being done in the solar physics community, where recent high resolution solar convection simulations (Miesch et al. 2000) are showing how numerically resolved turbulent convection, which occurs when the turbulent diffusivities are set low enough, is transporting momentum and energy within the convection zone in a manner different from that seen in the original laminar solar dynamo simulations conducted two decades ago (Gilman & Miller 1981, Glatzmaier 1984). The result is much more time-dependent small-scale convection that maintains a differential rotation profile more in line with what is inferred from helioseismology.

It is important to realize that although considerable progress has been made over the past decade in our understanding of the geodynamo, we are a long way from a realistic simulation. As computing resources continue to improve, spatial resolution can be increased and turbulent viscosity (the Ekman number) can be decreased. For example, the simulation illustrated in Figure 4 used spherical harmonic degrees up to 95 and orders up to 47 (i.e., over 2000 complex spherical harmonics) and 65 Chebyshev radial levels. Much higher spatial resolution, without hyperdiffusion, has been achieved for a short simulation (Roberts & Glatzmaier 2000b) using over 14,000 complex spherical harmonics. The Ekman number, magnetic Ekman number, Prandtl number, and Roberts number were $4 \times 10^{-6}$, $10^{-8}$, 150, and 3, respectively; the resulting broad spectra of kinetic and magnetic energies illustrate the tendency toward smaller-scale turbulence-dominated dynamics.

Even higher spatial resolution is needed to reach the parameter values of the proposed grand challenge. For this, the characteristic turbulent Ekman boundary layer depth, below the CMB and above the ICB, would be $(10^{-9})^{1/2}D = 70$ m, which would need to be resolved with perhaps seven Chebyshev radial levels (more if a finite difference method were employed). This would require the variables to be expanded in Chebyshev polynomials up to degree 2000 in the outer fluid core. Boundary layer instabilities and internal shear layers on the tangent cylinder will require correspondingly high horizontal spherical harmonic resolution.

This higher spatial resolution will result in Courant conditions that constrain the numerical time step to be much smaller in order to maintain numerical stability. Therefore, not only will each numerical step require much more computational
time and memory, but many more will be needed to span a given amount of simulated time.

Forcing symmetries in longitude by using only longitudinal wave numbers, \( m \), divisible by \( n \) to get an \( n \)-fold longitudinal symmetry, or in latitude by using only even or odd latitudinal wave numbers \( (l - m) \), to get a symmetry with respect to the equatorial plane, allows one to achieve significantly greater horizontal resolution at an affordable computational cost. However, these are not fully 3D global simulations, and they preclude certain flow structures, like flow across the geographic poles if any longitudinal symmetry is imposed or flow through the equatorial plane if symmetry with respect to the equator is imposed. Precluding the latter usually causes more problems than the former.

Although spectral methods have so far proven to be more efficient than local methods for achieving a given accuracy, an efficient fast Legendre transform is needed at these high resolutions to favorably compete with finite-element and finite-volume methods, which have the added advantage of requiring only local communication on parallel computer architectures.

7. CONCLUDING REMARKS

The past seven years have seen the development of several 3D numerical models of the geodynamo that are dynamically self-consistent. Numerical methods have improved and computing power has increased by many orders of magnitude since the first dynamo models were developed two decades ago (to study the sun). Recent geodynamo simulations have provided new insights and predictions for convection and magnetic field generation in the Earth’s core. However, we still have a long way to go before we can be confident that these results are robust—they may not be.

The grand challenge proposed here is to run a 3D geodynamo simulation with viscous, thermal, and compositional turbulent diffusivities set to 2 m²/s (the actual magnetic diffusivity of the Earth’s core) and setting all other model specifications to their Earth values. Reaching this goal will require a significant increase in numerical resolution on a massively parallel computer with improved numerical methods. If spectral methods are going to continue to be competitive, an efficient fast Legendre transform will be needed. Methods other than spectral that have high accuracy but require only local communication among computer processors need to be tested. In any case, the computational costs will be staggering by today’s standards.

Because of this, some people feel that this goal will be impossible to achieve within the next decade or two, and they therefore argue for simpler models that, although lacking self-consistency, can easily be run on today’s computers. But the Earth is not simple and an improved understanding of it will not be achieved with simple, poorly resolved models. Significant breakthroughs that will fundamentally improve our understanding of the geodynamo will come from young people who are inspired to work on what others proclaim is impossible.
ACKNOWLEDGMENTS

I would like to thank my colleague, Paul Roberts, for his encouragement and collaboration on dynamo simulations over the years and my student, Martha Evonuk, for many constructive suggestions regarding this manuscript. I also thank Ulrich Christensen for Figure 1 and Weijia Kuang for Figure 2. My research is supported by the Institute of Geophysics and Planetary Physics, the University of California Research Partnership Initiatives program, the NSF Geophysics program, and the NASA Planetary Atmospheres program. Computing resources were provided by a grant from the NSF MRI program and allocations from the following supercomputing centers: NPACI, NCSA, GSFC, LANL, and PSC.

Visit the Annual Reviews home page at www.annualreviews.org

LITERATURE CITED

Aurnou JM, Olson PL. 2000. Control of inner core rotation by electromagnetic, gravitational and mechanical torques. Phys. Earth Planet. Inter. 117:111–21


Figure 1  Snapshot of a dynamo simulation with $E_k = 1.5 \times 10^{-4}$, $P_r = 1$, $q = 2$, and a Rayleigh number 8.4 times critical. Red colors indicate positive values and blue colors indicate negative values. (a) The radial component of the magnetic field at the outer boundary; (b) the radial component of the fluid velocity at 80% of the outer boundary; (c) temperature perturbation in the equatorial plane; (d) longitudinally averaged field (left) and flow (right): toroidal components are in color and the poloidal components are represented as field and stream lines, respectively. (Courtesy of U. Christensen)
Figure 2  Snapshot of the radial component of the magnetic field at the core-mantle boundary (in an equal area projection of the entire CMB surface, with red outward and blue inward directed) and the longitudinally averaged field throughout the core (with toroidal field contours on the left and poloidal field lines of force on the right). (Courtesy of W. Kuang)
Figure 3  Snapshot of the 3D magnetic field structure for a Glatzmaier-Roberts simulation, illustrated with a set of magnetic lines of force, which are blue where directed inward and gold where directed outward. The axis of rotation is vertical and centered in the image. The lines are drawn out to two Earth radii. The field is a smooth, dipole-dominate, potential field outside the core. (From Glatzmaier & Clune 2000)
Radial Component of the Magnetic Field

At surface
Geomagnetic field (1980) up to degree 12

At core-mantle boundary
G-R simulation plotted up to degree 12
G-R simulation up to degree 95

Figure 4  Radial component of the magnetic field (red for outward directed and blue for inward) plotted at the Earth’s surface and at the core-mantle boundary. The surface fields are multiplied by 10 to obtain comparable color contrast. The Earth’s field is plotted out to spherical harmonic degree 12. For comparison, a snapshot from a Glatzmaier-Roberts simulation is plotted out to degree 12 and out to degree 95. (From Roberts & Glatzmaier 2000a)