Magnetoconvection Patterns in Rotating Convection Zones

Paul H. Roberts

Center for Earth and Planetary Interiors
Institute of Geophysics and Planetary Physics
University of California, Los Angeles, California, 90024

Abstract: In addition to the well-known granulation and supergranulation of the solar convection zone (the "SCZ"), the presence of so-called "giant cells" has been postulated. These are supposed span the entire thickness of the SCZ and to stretch from pole to pole in a sequence of elongated cells like a "cartridge belt" or a bunch of "bananas" strung uniformly round the Sun. Conclusive evidence for the existence of such giant cells is still lacking, despite strenuous observational efforts to find them. After analyses of sunspot motion, Ribes and others believe that convective motions near the solar surface occurs in a pattern that is the antithesis of the cartridge belt: a system of "toroidal" or "doughnut" cells, girdling the Sun in a sequence that extends from one pole to the other. Galloway, Jones and Roberts have recently tried to meet the resulting theoretical challenge, with the mixed success reported in this paper.

A significant parameter of rotating magnetoconvection is the Elsasser number, $\Lambda = \sigma B^2/2\Omega\rho$, where $\mathbf{B}$ is the magnetic field (predominantly zonal in the Sun), Ω is the angular velocity, ρ is the density and σ is the (turbulent) electrical conductivity. Intuitively one expects that, when $\Lambda \ll 1$, Coriolis forces will dominate Lorentz forces and, because of the Proudman–Taylor theorem, cells will be elongated in the direction of Ω and will form the cartridge belt pattern; when $\Lambda \gg 1$, one anticipates that the magnetic field will be decisive, and that the convection cells will align themselves with $\mathbf{B}$ in the toroidal pattern.

Jones, Roberts and Galloway (1990) have sought the preferred pattern of convection in an apparently simple system intended to mimic regions near the solar equator: a plane horizontal layer of compressible conducting fluid rotates about a horizontal axis perpendicular to an ambient uniform horizontal magnetic field. They determined the onset of convection, and demonstrated that the convection is necessarily time-dependent, a result that holds for all Λ, including even the non-magnetic case $\Lambda = 0$. Apart from this interesting difference, the model behaves in much the same way as the corresponding incompressible (Boussinesq) model, for which convection at onset is direct (time independent). In particular, even for large variations in the local value of Λ across the layer, the convection pattern

is the same at all depths, and shows no preference for the toroidal pattern. To simulate the large superadiabatic gradient in the upper convection zone, they also examined a two-zone model in which the Rayleigh number, R, in the upper zone is ten times that in the lower zone. Two kinds of convective modes were discovered, one highly concentrated in the upper zone, one predominant in the lower zone. The latter resemble banana cells; the former the doughnuts.

Jones and Roberts (1990) studied the one zone model (modified by the addition of vertical side walls parallel to the applied field) at finite convective amplitudes, and showed that the mean Lorentz force introduced by the convection creates a large "geostrophic" flow parallel to $\mathbf{B}$, this flow being a function of height only. In certain circumstances (stress–free side walls) an increase in R leads to symmetry-breaking in the Boussinesq model. One mode is concentrated near the surface and moves as a wave in one direction; the other is concentrated near the base of the convection zone and moves in the opposite direction. Jones and Roberts studied the unfolding of this symmetry-breaking through compressibility. The relationship of this research to the Ribes hypothesis is discussed in this article.

1. Background

The question of what kind of large-scale convection occurs in the Sun is of considerable interest not only in understanding how the Sun creates its magnetic field but also in providing an explanation of the internal rotation of the Sun determined by solar seismology. Although there have been advances recently on both observational and theoretical fronts, complete elucidation is not imminent. The present contribution, the result of joint investigations by Galloway, Jones and Roberts, will (it is felt) add a little insight to the theory. In what follows, the paper of Jones, Roberts and Galloway (1990) will be called "Paper 1"; its sequel, Jones and Roberts (1990), will be "Paper 2". We will not repeat the detailed analysis and results of these papers. We shall merely focus on a few salient points that, it is hoped, will be of interest.

The existence of two scales of motion on the Sun has long been known: the granulation and the supergranulation. A third scale of motion was postulated in 1968 by Simon and Weiss: "giant cells" that span the whole thickness of the solar convection zone ("SCZ"). Their hypothesis has generally been supported by the results of numerical simulations of convection in compressible atmospheres, even ones that extend over several scale-heights; it has usually been found that convection is dominated by cells that extend across the entire depth of the convecting layer.

When a convective layer rotates rapidly, slow motions within it try to accommodate themselves to the Proudman–Taylor theorem, which states that the steady infinitesimal motion of an inviscid fluid is two–dimensional with respect to the direction of the angular velocity of rotation, $\mathbf{\Omega}$. The pattern of convection therefore tends to be in rolls parallel to the axis of rotation. This led to the speculation that large-scale convection occurs in the SCZ in a sequence of rolls, each of which extends from pole to pole, and which together form a cartridge belt pattern round

the solar equator [Figure 1(a)]. While the conditions of the Proudman–Taylor theorem are not precisely satisfied, there have been a number of theoretical calculations (e.g. Roberts, 1968; Busse, 1970) and laboratory experiments (e.g. Carrigan and Busse, 1983; Hart *et al.*, 1986) that have confirmed the dominant role of the Proudman-Taylor constraint. Furthermore, numerical experiments, explicitly designed to simulate the SCZ by a spherical fluid shell, have exhibited [for both incompressible (Boussinesq) and compressible shells] clear indications of a cartridge belt pattern at equatorial latitudes. Associated with this pattern there is, at finite convective amplitudes, a zonal motion that is approximately "geostrophic", i.e. approximately constant on cylinders parallel to the rotation axis. The existence of this motion is in contradiction with recent results from solar seismology on the internal rotation of the Sun.

The strong theoretical and experimental expectation that cartridge belt convection exists in the Sun led observers to make a determined attempt to establish, by analysing the Sun's surface motion, that such giant cells actually are present. This search proved to be largely fruitless, and the situation became even more mysterious when Ribes *et al.* (1985) claimed that, superimposed on the granulation and supergranulation seen on the Sun's surface, there exists a pattern of toroidal cells, each of which encircles the Sun, and which together form a sequence that extends from one pole to the other [Figure 1(b)].
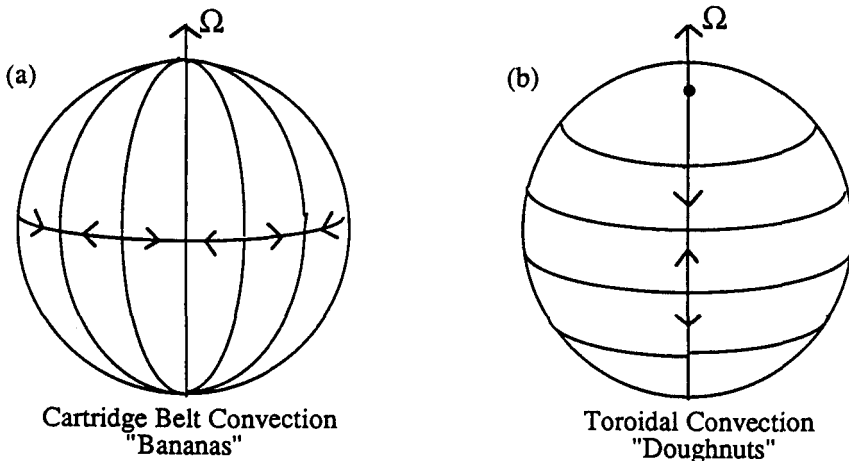

(a) Cartridge Belt Convection
"Bananas"

(b) Toroidal Convection
"Doughnuts"

Fig. 1. Types of convection in a sphere.

The idea that initiated the following program of research is very simple. Magnetohydrodynamics ("MHD", for short) boasts its own Proudman–Taylor theorem, which states that steady infinitesimal motions in a perfectly conducting fluid are two-dimensional with respect to the direction of the magnetic field, **B**. When a layer of electrically conducting fluid lies in a magnetic field, slow motions within

it try to accommodate themselves to the MHD theorem by forming a pattern of convection in rolls parallel to the magnetic field. Since **B** is predominatly toroidal in the Sun, there would, in the absence of rotation, be a strong tendency for large–scale convection in the SCZ to occur in a sequence of tori of the type envisaged by Ribes *et al.* In short, there are two extremes: if B is "large" and Ω is "small", we expect doughnuts threaded onto the solar axis but, if B is "small" and Ω is "large", we expect bananas girdling the equator. And remarkably a compressible layer like the SCZ can, at one and the same time, possess both tendencies. This may be understood in the following way: The correct dimensionless parameter that quantifies the adjectives "small" and "large" above is the Elsasser number,

$$\Lambda = \frac{\sigma B^2}{2\Omega\rho},\tag{1.1}$$

where ρ is density, and σ is the magnetic conductivity, which in the Sun is effectively the turbulent conductivity. [In what follows, we shall often use in place of σ the "magnetic diffusivity", $\eta = 1/\mu\sigma$, where μ is the permeability of free space.] When one forms a dimensionless number like Λ, he usually assigns it a single value by inserting "representative" magnitudes of B, σ and ρ. In astrophysical reality, the effective σ depends on position, being large in stably stratified regions where molecular transport coefficients are appropriate, and small in turbulent regions such as the SCZ where the magnetic field is, like the fluid motions, on a very small length scale and therefore ohmically dissipates electromagnetic energy at a greatly enhanced rate. Even more pronounced than the variation of η within the SCZ, is the variation through seven orders of magnitude of ρ. This is so great that, if we temporarily regard Λ as a function of position that depends on the local values of B, σ and ρ, we may well believe that Λ is small at the bottom of the SCZ but is large at its top. In such a case one must expect that the convection "wants" to obey the Proudman-Taylor theorem at the bottom of the SCZ, but wants to encourage toroidal motions at the top! The main objective of the two papers was to explore this idea.

Initially it was unclear how a linear stability analysis could either support, or cast doubt on, our idea. For, in such a theory, all perturbations away from the basic conduction solution are proportional to

$$\exp[im\phi],\tag{1.2}$$

where ϕ is longitude; bananas require $m \neq 0$ while doughnuts are axisymmetric: $m = 0$. But these m are global. One cannot envisage $m \neq 0$ at the bottom of the SCZ and $m = 0$ at the top. So an analysis of the linear problem cannot validate or invalidate our idea. Nevertheless, it is hardly a good theoretical practice to go straight to nonlinear convection before one has understood the marginal situation!

2. On making life bearable

Nearly everyone approaches MHD with that sinking feeling, "Too many parameters!" Every force and every diffusivity seems to have its own dimensionless measure, and soon they create an environmental miasma. Adding rotation makes matters worse, and including compressibility makes them much worse. With this in mind, we resolved to investigate as simple a situation as possible, consistent with retaining the key physical effects identified above. In this section we shall describe briefly the four main simplifications that we shall at various times employ.

First and foremost (simplification 1), we decided to eliminate all geometrical effects by investigating convection in a plane polytropic layer, $z_0 d \leq z \leq (1 + z_0)d$, in which the conduction solution (in which the fluid velocity, $\mathbf{V}$, is zero) is

$$T_0 = \beta_0 z, \qquad p_0 = P\left(\frac{z}{d}\right)^{m+1} = \mathcal{R}\rho T, \qquad \rho_0 = \frac{P(m+1)}{gd}\left(\frac{z}{d}\right)^m, \qquad (2.1)$$

where p is gas pressure, T is temperature, $\mathbf{g} = g\mathbf{1}_z$ is the effective gravitational acceleration, assumed constant ($\mathbf{1}_q$ denotes the unit vector in the direction of q increasing), $\mathcal{R}$ is the gas constant, and m is the polytropic index; β_0 is the temperature gradient, which is not to be confused with the more significant "superadiabatic" temperature gradient β:

$$\beta_0 = \frac{g}{\mathcal{R}(m+1)}, \qquad \beta = \beta_0 - \frac{g}{c_p} = \beta_0\left(\frac{m+1}{\gamma} - m\right). \qquad (2.2)$$

Here γ is the ratio of specific heats (assumed constant), and $c_p = \gamma\mathcal{R}/(\gamma - 1)$ is the specific heat at constant pressure. By varying z_0 one can vary the effects of compressibility from the case of the complete polytrope ($z_0 = 0$) to the incompressible – or "Boussinesq" – limit ($z_0 \to \infty$). We concentrated most on the cases $z_0 = 0.1$ and $z_0 \to \infty$. [By supposing that the effective gravitational acceleration, g, is constant, rather than the actual acceleration due to gravity, we have excluded complications arising from the centrifugal force.]

The extra suffix $_0$ on $T, p, \rho, \ldots$ serves to distinguish values in the conduction state from the corresponding values

$$T = T_0 + T', \qquad p = p_0 + p', \qquad \rho = \rho_0 + \rho', \qquad \ldots, \qquad (2.3)$$

of these variables when convection occurs. In the linear convective stability problem, $\mathbf{V}, T', p', \rho', \ldots$ are infinitesimal quantities whose squares and products may be neglected, and which may therefore be divided into non-interacting Fourier modes proportional to [in place of (1.2)]

$$\exp\left[ik\frac{x}{d} + il\frac{y}{d} + (\sigma + i\omega)\frac{\eta t}{d^2}\right], \qquad (2.4)$$

where k and l are the horizontal wavenumbers, and $\sigma + i\omega$ is the growth rate; all these constants are real and dimensionless. The use of σ both for growth rate

and conductivity should not cause confusion since we shall usually replace the conductivity by the equivalent diffusivity η.

We decided to simulate the region near the equator in the spherical problem, and took therefore both $\mathbf{B}$ and $\mathbf{\Omega}$ horizontal and perpendicular to each other, so that we have

$$\mathbf{\Omega} = \Omega\mathbf{1}_y, \qquad \mathbf{B}_0 = B_0\mathbf{1}_x, \qquad \mathbf{V}_0 = 0. \qquad (2.5)$$

The doughnut cells then correspond to $k = 0$ and the banana cells to $l = 0$. See Figure 2 on which the correspondence between coordinates is also shown.

Simplification 2 is the "anelastic approximation", i.e. the linearization of the thermodynamic relations about a reference state, a procedure that can be defended if $\beta \ll \beta_0$. The technical burden is eased, e.g. the Coriolis force, $2\rho\mathbf{\Omega}\times\mathbf{V}$ is replaced by the simpler $2\rho_0\mathbf{\Omega}\times\mathbf{V}$. Most important is the fact that the mass conservation equation is reduced to

$$\nabla\cdot(\rho_0\mathbf{V}) = 0. \qquad (2.6)$$

As a result, sound waves are "filtered out" of the system; we need only worry about longer time scales.

Simplification 3 is also to do with timescales. If $B = 0$, inertial-gravity waves can propagate in the layer, and we shall suppose not only that $\beta/\beta_0 \ll 1$ (see above) but also that the buoyancy frequency is much less than rotation frequency:

$$N \equiv \left(\frac{g\beta}{d\beta_0}\right)^{1/2} \ll 2\Omega. \qquad (2.7)$$

When $\Omega = 0$, the layer can transmit Alfvén waves of frequency V_A/d, where $V_A = B_0/\sqrt{(\mu\rho)}$ is the Alfvén velocity. We shall suppose that

$$V_A/d \ll 2\Omega. \qquad (2.8)$$

This leaves the inertial wave at much the same frequency $[\mathrm{O}(\Omega)]$ as in the case $B = 0$, but slows the Alfvén wave down to frequencies of order $V_A^2/2\Omega d^2$. These are the timescales of interest in the thermal convection problem, and the fast, inertial waves are filtered out by omitting the inertial acceleration, $D\mathbf{V}/Dt$, in the equation of motion in comparison with the Coriolis acceleration, $2\mathbf{\Omega}\times\mathbf{V}$; here D/Dt is the motional derivative.

Simplification 4 concerns the diffusivities. We shall suppose that

$$E \equiv \frac{\nu_0}{2\Omega d^2} \ll 1, \qquad \frac{\kappa_0}{2\Omega d^2} \ll 1, \qquad \frac{\eta}{2\Omega d^2} \ll 1. \qquad (2.9)$$

Here $\nu_0 = \lambda/\rho_m$ and $\kappa_0 = K/\rho_m c_p$ are the kinematic viscosity and thermal diffusivity at the center of the layer; since the dynamic viscosity λ and thermal conductivity K are constants, ν and κ are inversely proportional to ρ_0. Because of (2.7) and (2.8), inequalities (2.9) do not imply that Λ and the relevant Rayleigh number, R, are large, where
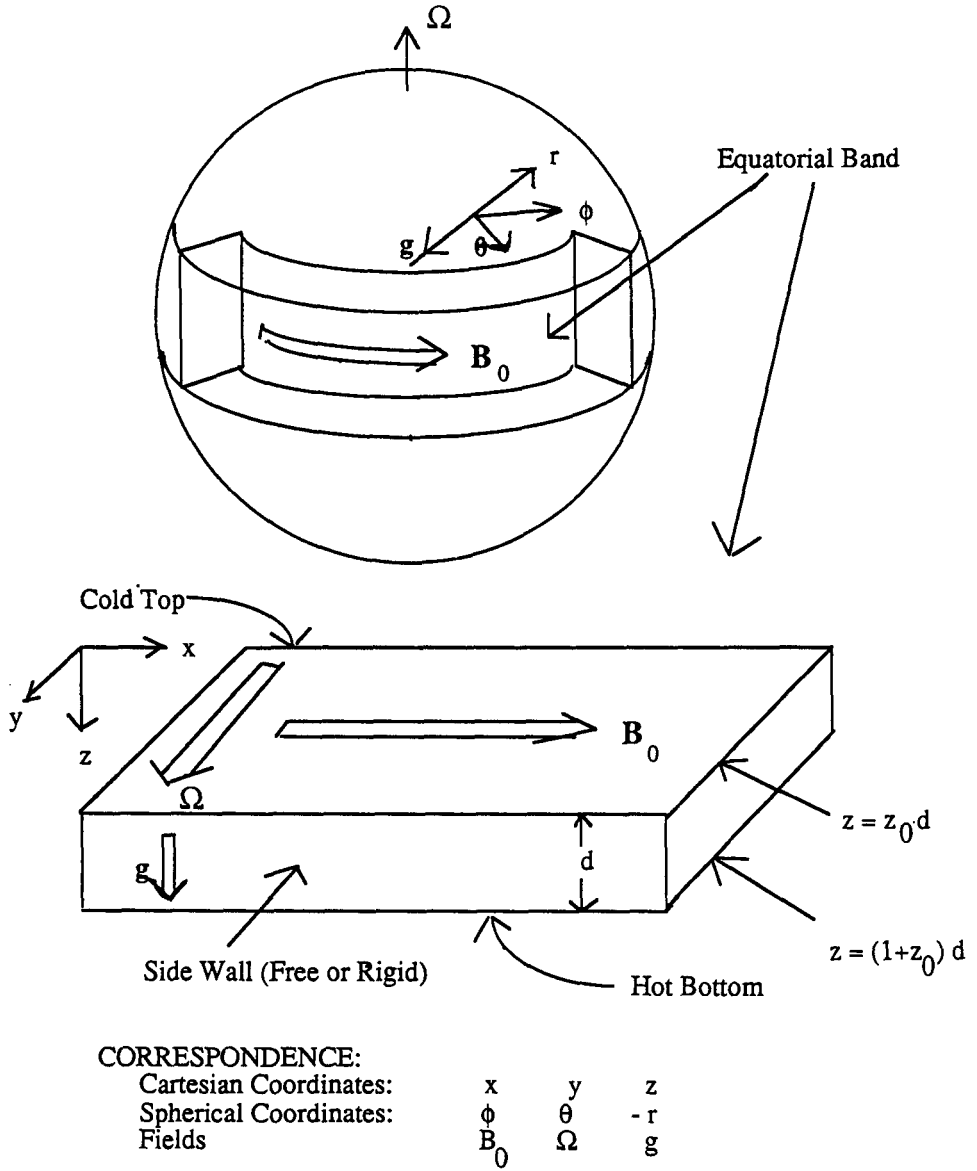
Fig. 2. Relationship between spherical and plane geometries

$$\Lambda = \frac{B^2}{2\Omega\mu\rho\eta}, \qquad R = \frac{g\beta d}{2\Omega\beta_0\kappa_0}. \qquad (2.10)$$

For we may write Λ and R as $O(1)$ ratios of small quantities:

$$\Lambda = \frac{(V_A/2\Omega d)^2}{\eta/2\Omega d^2}, \qquad R = \frac{(N/2\Omega)^2}{\kappa_0/2\Omega d^2}. \qquad (2.11)$$

The main effect of (2.9) is to eliminate viscous forces, except in boundary layers. Not all unpleasantness is removed by this simplification. The polytropic state is a conduction solution only if the thermal conductivity, K, is constant, but the thermal diffusivity, $\kappa = K/\rho c_p$, enters more naturally into the convection theory. By simpification 2, this may be replaced by $\kappa_0 = K/\rho_0 c_p$, a known function of z. This kind of nuisance is unavoidable, but is largely irrelevant to the main topic.

We conclude this section with a summary of the basic equations that result from our simplifications; we write them in dimensionless form, and replace z by

$$\zeta = z/(z_0 + \tfrac{1}{2})d. \qquad (2.12)$$

We then have

$$\nabla\cdot(\zeta^m \mathbf{V}) = 0, \qquad (2.13)$$

$$\zeta^m \mathbf{1}_y \times \mathbf{V} = -\frac{(z_0 + \tfrac{1}{2})}{m+1}\nabla p + \rho\mathbf{1}_z + \Lambda(\nabla\times\mathbf{B})\times(\mathbf{1}_x + \mathbf{B}) + E[\nabla^2\mathbf{V} + \tfrac{1}{3}\nabla(\nabla\cdot\mathbf{V})], \qquad (2.14)$$

$$\frac{\partial\mathbf{B}}{\partial t} = \nabla\times[\mathbf{V}\times(\mathbf{1}_x + \mathbf{B}) - \nabla\times\mathbf{B}], \qquad (2.15)$$

$$\frac{\zeta^m}{q\gamma}\frac{DT}{Dt} = \frac{(\gamma-1)}{q\gamma}\zeta\frac{D\rho}{Dt} - [R\zeta^m + \frac{1}{\gamma q(z_0 + \tfrac{1}{2})}\{\rho - m(\gamma-1)\zeta^{m-1}T\}]V_z + \nabla^2 T, \qquad (2.16)$$

$$p = \zeta^m T + \zeta\rho, \qquad (2.17)$$

$$\nabla\cdot\mathbf{B} = 0, \qquad (2.18)$$

where $q = \kappa_0/\eta$ is the ratio of diffusivities. The primes have been omitted from T', p', ρ', etc; the motional derivatives are

$$\frac{DT}{Dt} = \frac{\partial T}{\partial t} + \mathbf{V}\cdot\nabla(T_0 + T), \qquad \frac{D\rho}{Dt} = \frac{\partial\rho}{\partial t} + \mathbf{V}\cdot\nabla(\rho_0 + \rho).$$

Clearly (2.13) is the mass conservation equation (2.6), while (2.14) expresses momentum conservation; (2.15) is the magnetic induction equation, (2.16) is the heat conduction equation, (2.17) is the gas law and (2.18) expresses conservation of magnetic flux.

In the Boussinesq limit (2.13)–(2.18) reduce to the much simpler form:

$$\nabla\cdot\mathbf{V} = 0, \qquad (2.19)$$

$$\mathbf{1}_y \times \mathbf{V} = -\nabla p + T\mathbf{1}_z + \Lambda(\nabla\times\mathbf{B})\times(\mathbf{1}_x + \mathbf{B}) + E\nabla^2\mathbf{V}, \qquad (2.20)$$

$$\frac{\partial\mathbf{B}}{\partial t} = \nabla\times[\mathbf{V}\times(\mathbf{1}_x + \mathbf{B})] + \nabla^2\mathbf{B}, \qquad (2.21)$$

$$\frac{1}{q}\frac{DT}{Dt} = \nabla^2 T - RV_z, \qquad (2.22)$$

$$\nabla\cdot\mathbf{B} = 0. \qquad (2.23)$$

3. The impossibility of steady modes

The elimination of steady modes is by *reductio ad absurdum*, i.e. we shall suppose that ω in (2.4) is zero. More precisely, we may suppose without loss of generality that solutions exist in which V_z is proportional to $\cos(kx + ly) \exp \sigma t$, where k, l and σ are real.

The linearized form of the governing equations (2.13)–(2.18) are, with E set zero [see (2.9)],

$$\nabla \cdot (\zeta^m \mathbf{V}) = 0, \qquad (3.1)$$

$$\zeta^m \mathbf{1}_y \times \mathbf{V} = -\frac{(z_0 + \frac{1}{2})}{m+1} \nabla p + \rho \mathbf{1}_z + \Lambda (\nabla \times \mathbf{B}) \times \mathbf{1}_x, \qquad (3.2)$$

$$\frac{\partial \mathbf{B}}{\partial t} = \frac{\partial \mathbf{V}}{\partial x} - (\nabla \cdot \mathbf{V}) \mathbf{1}_x + \nabla^2 \mathbf{B}, \qquad (3.3)$$

$$\frac{\zeta^m}{q\gamma} \frac{\partial T}{\partial t} = \frac{(\gamma - 1)}{q\gamma} \zeta \frac{\partial \rho}{\partial t} - R_c \zeta^m V_z + \nabla^2 T, \qquad (3.4)$$

$$p = \zeta^m T + \zeta \rho, \qquad (3.5)$$

$$\nabla \cdot \mathbf{B} = 0, \qquad (3.6)$$

R_c being the critical Rayleigh number for marginal convection.

In steady convection, T and $-V_z$ are in phase, as (3.4) shows: since V_z is proportional to $\cos(kx + ly) \exp \sigma t$, so is $-T$. In the Boussinesq approximation, ρ depends only on T, so that ρ and V_z are in phase. The maxima of V_z in any horizontal plane mark the "centers" of the cold, heavy, descending streams; the minima correspond to the centers of the hot, light, ascending plumes. The buoyancy force $\rho \mathbf{1}_z$ has the right phase to maintain that situation.

In a rapidly rotating fluid, p and V_z are out of phase as the x-component of (3.2), namely

$$\frac{\partial p}{\partial x} = -\frac{m+1}{z_0 + \frac{1}{2}} \zeta^m V_z, \qquad (3.7)$$

shows: p is proportional to $\sin(kx + ly) \exp \sigma t$ rather than to $\cos(kx + ly) \exp \sigma t$ as V_z is; this is an instance of what is sometimes called "Buys Ballot's law". The fact that p is out of phase with T is irrelevant in a Boussinesq fluid, since ρ depends only on T, and steady convection is still possible with ρ and V_z in phase, as described above. In a compressible fluid however, ρ depends both on T *and* on p; see (3.5). Thus ρ and V_z can no longer be in phase. The centers of maximum negative buoyancy no longer coincide with the maximum downward velocities. The convection pattern therefore moves, as the downward motion tries to "find" the density maxima, but creates new, translated density maxima as it does so. This result, the first surprise we encountered, does not depend on simplification 3; it is equally true for nonmagnetic compressible convection. Nevertheless, we found that the general features of the compressible modes were often not qualitatively very different from those of the corresponding Boussinesq modes, except that the former drift while the latter do not.

Our numerical results did not encourage us to believe that our original idea, that a transition from doughnut to banana convection arises through the variation with depth of ρ_0, and hence of the effective Λ, is correct. The situation might have been different had the most unstable modes been *multicellular*, rather than "unicellular", i.e. if there had been many rather than one layer of cells between the top and bottom of the layer, each of which had some degree of independence from its neighbors. Although the idea that convection preserves its coherence only over one scale height is basic to mixing-length theories of turbulent stellar convection, our marginal laminar modes appeared to show no such tendency, a fact already noted in a number of previous studies of compressible convection. Eigenfunctions are shown in Figures 3 in two cases ($\Lambda = 1$, $l = 1$, $k = 2.5$ top; $\Lambda = 16$, $l = 2$, $k = 1$ bottom), in both of which the ratio of the density at the bottom of the layer to that at the top is about 36.5, but in each of which the eigenfunctions extend right across the layer.
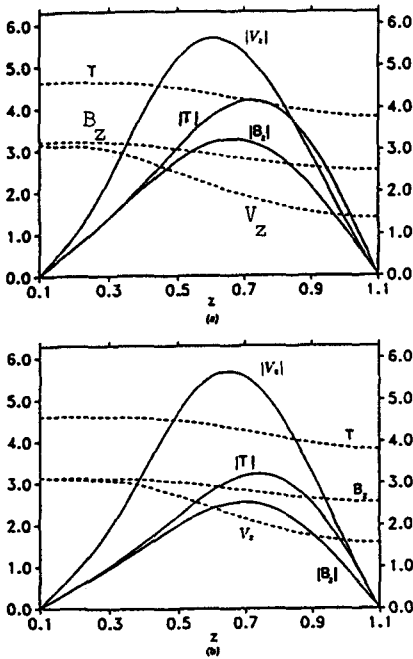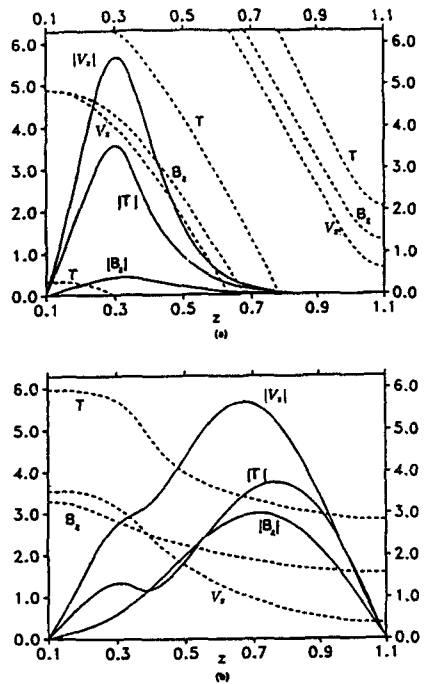

Fig. 3. One zone model

Fig. 4. Two zone model

One interesting feature emerged from the calculations however: although the phases (shown by the dashed curves) of V_z and T differ by about π at depth, as happens in the "normal" Boussinesq case (see above), they are out of phase by only $\frac{1}{2}\pi$ at the top of the layer. This suggested that, when convection is of finite amplitude convection even if that amplitude is small, the convective heat flux will be impeded near the top of the layer, so that a large mean superadiabatic

temperature gradient will develop to carry heat out of the layer. With this thought in mind, we studied a two-zone model in which the effective Rayleigh number in the upper quarter of the layer ($z_0 < z < z_0 + \frac{1}{4}$) is ten times that in the lower three quarters ($z_0 + \frac{1}{4} < z < 1 + z_0$). We brought this about by a small difference in γ between the two zones; see (2.2) and (2.10). Two kinds of eigenmodes were found, one concentrated mostly in the upper 25% of the layer, the other in the lower 75%; they are sketched in Figures 4. The ratio, k/l, at criticality is less for the upper mode than for the lower, as one would expect from its greater effective Elsasser number. The effect is not as great as one might have anticipated however: $k/l \approx 1.5$ for the upper mode and ≈ 1 for the lower, but the effective Elsasser number at the center of the upper layer is almost four times its overall value $\Lambda = 4$. For each l, the lower modes are more easily for small k, while the minimum $R(= R_m(l)$, say) for that l occurs for the upper mode at higher k; $R_m(l)$ increases with l. Smallest are $R_m(3)$ and $R_m(4)$, both of which are close to 82.

The results of the two zone model encouraged us more than those of the one zone model. The hope expressed above, that the upper layer might act independently of the lower layer, may be a vain one [see the discussion at (1.2)], at least in infinitesimal convection. But, by introducing heuristically into the linear stability problem an effect that is plausibly present at finite amplitude, namely an increase in the superadiabatic temperature gradient with height, we obtained some indications that the upper layer "prefers" doughnut structures and the lower layer banana modes. There seemed to be little doubt that, to investigate this properly, we had to investigate finite amplitude convection. Further surprises were in store, as we shall now describe.

4. Finite amplitude Boussinesq convection

Our second surprise arose from what is often called the "Taylor constraint". Consider the Boussinesq case, in the weakly nonlinear range, in which $|R/R_c - 1| \ll 0$. In the limit $E \to 0$, (2.20) becomes

$$\mathbf{1}_y \times \mathbf{V} = -\nabla p + T\mathbf{1}_z + \Lambda(\nabla \times \mathbf{B}) \times (\mathbf{1}_x + \mathbf{B}), \qquad (4.1)$$

from which Taylor's constraint follows:

$$\mathbf{F}(z,t) = 0, \qquad (4.2)$$

where

$$\mathbf{F}(z,t) \equiv \Lambda \overline{[(\nabla \times \mathbf{B}) \times \mathbf{B}]}_H; \qquad (4.3)$$

the overbar denotes the horizontal average (i.e. the average over x and y) and the H denotes horizontal components. To leading order for small $|R/R_c - 1|$, we may replace $\mathbf{B}$ in (4.3) by the eigenfunction determined from the linear eigenvalue system derived from (2.19)–(2.23) for $R = R_c$. And it is found immediately that

$$\mathbf{F}(z,t) \neq 0, \qquad (4.4)$$

in contradiction with (4.2).

This kind of difficulty can hardly be described as the next surprise. It was encountered first by Taylor (1963) and is now a commonplace in the the theory of highly rotating fluids. Inequality (4.4) states that the Lorentz force contains a component that cannot be balanced by Coriolis forces, buoyancy forces or pressure gradients. It must therefore accelerate the fluid to create a growing horizontal motion that is independent of x and y, a so-called "geostrophic flow":

$$\mathbf{U} = \mathbf{U}_H(z,t). \tag{4.5}$$

When this velocity becomes sufficiently large, it will be incorrect to treat it as part of a perturbation series; instead, it will have to be included in the zero order state, i.e. the last of (2.5) will have to be amended to $\mathbf{V}_0 = \mathbf{U}$, with the result that many new terms will have to be incorporated into (2.20)–(2.22). In particular $\mathbf{U}$ will alter the magnetic field and the Lorentz force. Taylor (1963) argued that the geostrophic flow would quickly adjust itself until the new magnetic field satisfied (4.2), and he gave a prescription that thereafter uniquely determined a $\mathbf{U}$ such that the new $\mathbf{B}$, as it evolved, would continuously satisfy (4.2). This defines the "Taylor state", in which the geostrophic flow is "O(1)" (by which we mean it is independent of ν, see below) but in which (4.2) is obeyed.

Taylor's suggestion has been welcomed widely, as being both a mathematically and a physically appealing resolution of the conflict between (4.2) and (4.4). We too expected it to apply to the present problem but (and here comes the second surprise!) we found that Taylor's idea failed: Taylor states do not exist when $\Lambda > 4.080\ldots$. This is illustrated in Figure 5. The lower part of the diagram shows the results of the original linear stability problem, with $\mathbf{U}$ absent. For small Λ, banana cells are excited first at R_c but, as Λ increases, the preferred cells become "oblique", i.e. neither k nor l are zero. These two branches of the stability curve are continued as dashed lines when they are no longer preferred. Also shown is the curve, labeled R_T, defined by the Taylor states. Evidently none such exist when Λ exceeds 4 (approximately).

It was first pointed out by Roberts and Stewartson (1975), and independently by Braginsky (1975), that the conflict between (4.2) and (4.4) might be resolved by a mechanism different from Taylor's, through the term omitted in writing down (4.1): the viscosity ν.[1] It is at once clear that, since E is small, the viscous stress can be made potent only if the velocity shear created by the unbalanced $\mathbf{F}$ is large. This shear can be predominantly in a boundary layer at a wall (see below) or distributed throughout the fluid, the possibility we examine first. But we quickly find that, in either case, only the x-component of the geostrophic flow, that aligned with $\mathbf{B}_0$, arises:

$$\mathbf{U} = U(z,t)\mathbf{1}_x. \tag{4.6}$$

The horizontally averaged (2.20) gives

---

[1] Braginsky also pointed out that, if the walls were electrically conducting, an electromagnetic coupling with the walls would also be present, which could resolve the difficulty. Here, for simplicity, we shall confine attention to the simpler mechanism that relies on viscosity.
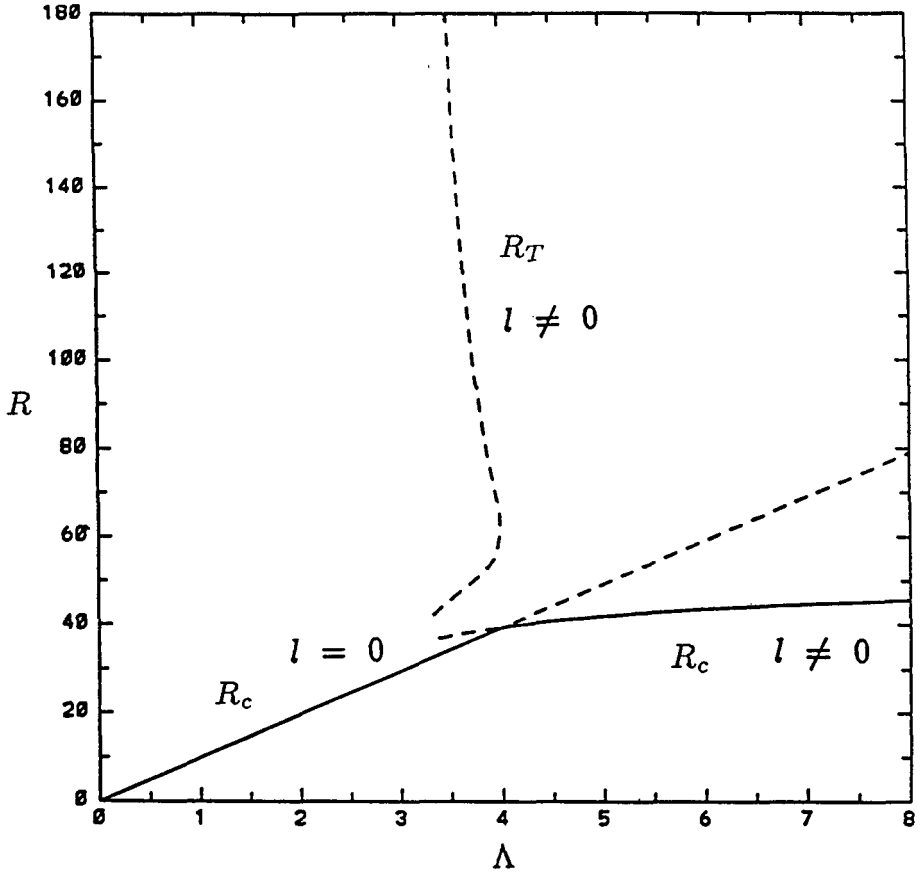
Fig. 5. Marginal and Taylor states.

$$0 = \Lambda \overline{[(\nabla\times\mathbf{B})\times\mathbf{B}]}_H + E\nabla^2\mathbf{U}, \tag{4.7}$$

and this offers one way of balancing $\mathbf{F}$: (2.3) can be scaled as

$$T \to T_0 + E^{1/2}T, \quad p \to p_0 + E^{1/2}p, \quad \mathbf{V} \to \mathbf{U} + E^{1/2}\mathbf{V}, \quad \mathbf{B} \to \mathbf{B}_0 + E^{1/2}\mathbf{B}. \tag{4.8}$$

We then obtain what has sometimes been described as a "nonlinear eigenvalue problem", that of solving the linear equations

$$\nabla\cdot\mathbf{V} = 0, \tag{4.9}$$

$$\mathbf{1}_y\times\mathbf{V} = -\nabla p + T\mathbf{1}_z + \Lambda(\nabla\times\mathbf{B})\times\mathbf{1}_x, \tag{4.10}$$

$$\frac{\partial\mathbf{B}}{\partial t} + \mathbf{U}\cdot\nabla\mathbf{B} = \mathbf{B}\cdot\nabla\mathbf{U} + \nabla\times(\mathbf{V}\times\mathbf{1}_x) + \nabla^2\mathbf{B}, \tag{4.11}$$

$$\frac{1}{q}\left(\frac{\partial T}{\partial t} + \mathbf{U}\cdot\nabla T\right) = \nabla^2 T - RV_z, \tag{4.12}$$

$$\nabla \cdot \mathbf{B} = 0, \tag{4.13}$$

in such a way that the nonlinear restriction,

$$0 = \Lambda\overline{[(\nabla \times \mathbf{B}) \times \mathbf{B}]_x} + \nabla^2 U, \tag{4.14}$$

is met. The solutions to (4.9)–(4.13) must satisfy the appropriate conditions on the upper and lower boundaries, as must the solution of (4.14) for U. As for the previous linear eigenvalue problem (the one in which $\mathbf{U}$ was assumed zero), the oblique modes $k \neq 0$ come in pairs at equal angles to $\mathbf{B}_0$, differing by the sign of l.

There is a second way in which viscosity can act, namely through Ekman layers. These exist at any point on a no–slip boundary at which $\mathbf{\Omega}$ has a nonzero normal component; in the geometry of Figure 2, this means that they do not exist on the horizontal boundaries, but only on "side walls". We supposed that these are impermeable, thermally insulating, and perfect electrical conductors. It is then necessary to assume that the two oblique rolls have equal amplitudes, so that (2.4) is replaced by (for example)

$$B_z(x, y, z, t) = ik\hat{B}(z)\exp[i(kx + \omega t)]\cos ly + *, \tag{4.15}$$

where the * denotes the complex conjugate of the expression preceding it and now $l > 0$. (Other variables are expanded similarly, but without the prefactor ik that B_z has, which proves to be convenient.) These Fourier modes satisfy all conditions on the side walls if their separation, L, is $n\pi/l$, where n is an integer, and we make the simplest choice: $n = 1$, i.e. the side walls are at $y = 0$ and $y = \pi/l$. Our infinite plane layer has become a "duct"; in terms of the original spherical model, we are confining the fluid between two particular latitudes.

When $\mathbf{U} = \mathbf{0}$, overstable modes ($\omega \neq 0$) are preferred only when $q > 1$. We did not concern ourselves with such q; in fact we usually assumed that $q = 1$. We concentrated on the case $\Lambda = 6$ and we here adopt the values of k and l found in Paper 1 to be the most unstable: $k = 1.117$ and $l = 1.668$. [2] If these walls are slippery, the solution is governed by (4.9)–(4.14), but when they are no-slip surfaces everything changes.

In the case of no–slip side walls, all the shear required to balance $\mathbf{F}$ takes place in Ekman layers on those walls. These shears are, at level z, of order $U(z)/\delta_\nu$, where $\delta_\nu = (\nu/\Omega)^{1/2}$ is the Ekman layer thickness. They create a stress of order $2\rho_0(\Omega\nu)^{1/2}U(z)$ which is required to offset the net unbalanced force, LF (per unit x-length), on the entire fluid between the walls at that level. Thus U must be of order $(L/\rho_0)(\Omega\nu)^{-1/2}F = O((L/\mu\rho_0 d)(\Omega\nu)^{-1/2}B^2)$. In fact, a more precise calculation shows that

$$U = \frac{L}{2\rho_0\mu(\Omega\nu)^{1/2}}\overline{[(\nabla \times \mathbf{B}) \times \mathbf{B}]_x}. \tag{4.16}$$

---

[2] It was realized rather late that, even though these are direct when $\mathbf{U} = \mathbf{0}$, this may not be true when $\mathbf{U} \neq \mathbf{0}$. The diagrams given below were prepared before we made the slightly different choice of paper 2, namely $k = 1.6091$, $l = \pi$, one for which direct instability is assured. The structure of the solutions is qualitatively the same.

From this $B^2/\nu^{1/2}$ dependence, we see that, in our dimensionless units, B must be $O(E^{1/4})$, so that scaling (4.8) must be replaced by

$$T \to T_0 + E^{1/4}T, \;\; p \to p_0 + E^{1/4}p, \;\; \mathbf{V} \to \mathbf{U} + E^{1/4}\mathbf{V}, \;\; \mathbf{B} \to \mathbf{B}_0 + E^{1/4}\mathbf{B}. \;\; (4.17)$$

Equations (4.9)–(4.13) hold as before; merely the dimensionless form of (4.16) replaces (4.14), but (the third surprise) this has an enormous, even *qualitative*, effect on the solutions!

Consider first the case of no-slip side walls, "the rigid duct", as we call it. When R is large, the solutions are multicellular, i.e. any vertical line across the duct meets several different cells, each operating in a different band of z. This may be seen with particular clarity in Figure 6(b) where the phases of B and T are shown. They are are almost coincident throughout the layer; that of V_z is too close to that of T to be included. The presence of many cells causes the geostrophic velocity to reverse its direction equally often with height, as is seen in Figure 6(a). The amplitudes of the perturbation variables oscillate similarly. The existence of multicellular solutions is encouraging for the main objective of the present line of study, that of finding convective modes that are banana–like at the bottom but doughnut–like at the top. But this is, of course, prevented here by the symmetry of the Boussinesq solutions about the mid–plane $z = \frac{1}{2}$.

The "free duct", as we call the duct with slippery side walls, shows no such behavior. Instead it exhibits a symmetry breaking that the rigid duct apparently lacks. As R is increased from zero, the duct first becomes unstable (when $\Lambda = 6$) for $R \approx 43.637$ with, as noted earlier, $k \approx 1.117$ and $l \approx 1.668$; this point is denoted by B_1 in Figure 7. Each roll extends from one z boundary to the other, and is completely symmetric, e.g. $V_z(1-z) \equiv V_z(z)$; all other variables are similarly symmetric or antisymmetric with respect to the mid–plane of the layer. As R is increased further, this structure is maintained by the finite amplitude convection until a second critical value of R is attained. At this point, two time-dependent solutions arise, neither of which is symmetric with respect to $z = \frac{1}{2}$, but which preserve an overall symmetry in the sense that one is the reflection of the other in the mid–plane, e.g. $V^{(1)}(1-z) \equiv V^{(2)*}(z)$. These are represented in Figure 7 by the two curves emerging from B_3 and continuing to the right. Convection occurs primarily in the upper half of the layer in one solution, and primarily in the lower half for the other. This is encouraging for the main objective of the present line of study, that of explaining banana–like convection at the bottom of the layer but doughnut–like convection at the top.

The two solutions are time-dependent, and travel in the $\pm x$-directions with equal but opposite phase velocities. After undergoing this symmetry–breaking bifurcation, denoted by B_3 in Figure 7, the original symmetric solution undergoes a second bifurcation of saddle-node type, corresponding to B_4 in Figure 7, where it links to the branch of finite amplitude solutions that emerges from the second critical Rayleigh number. This point, denoted by B_2 in Figure 7, corresponds to bicellular convection of infinitesimal amplitude, i.e. convection in which there are two layers of rolls between the horizontal boundaries.
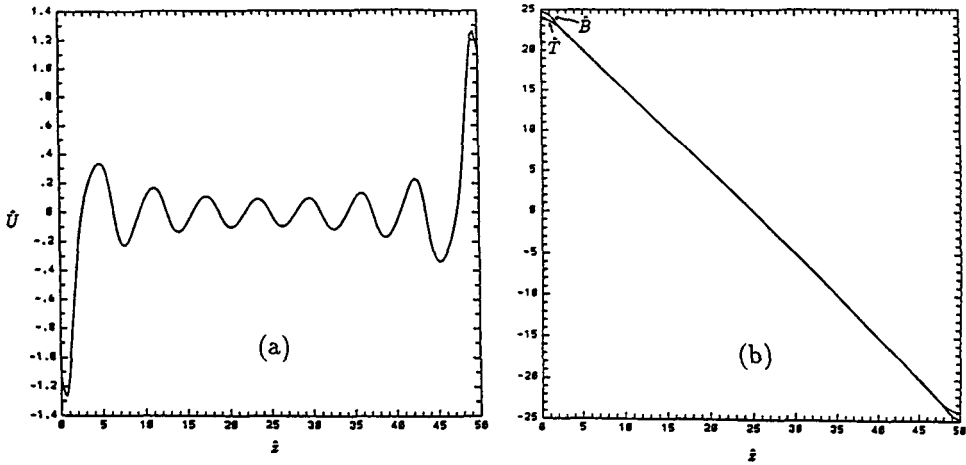
Fig. 6. Rigid Boussinesq duct (a) geostrophic flow, (b) phases of B and T. The variables carry a hat ^ to indicate that they have been scaled; see Paper 2 for details.
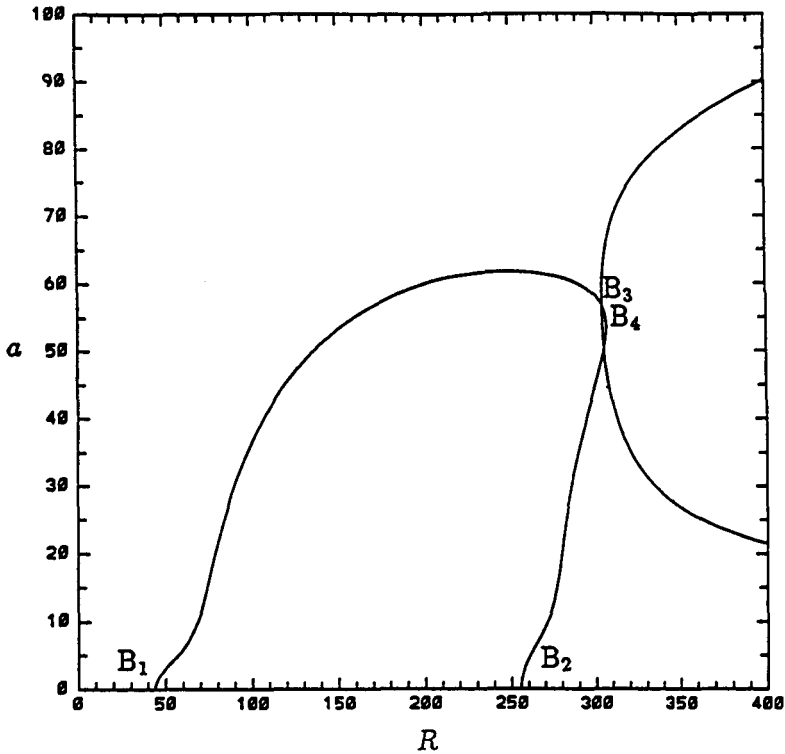

Fig. 7. Bifurcation diagram for free Boussinesq duct. The amplitude parameter a is constructed in such a way that the points representing solutions that are symmetric about $z = \frac{1}{2}$ coincide, but those which lack that symmetry are represented by different points; see Paper 2 for details.

Figures 8 give an idea of the structure of one of the two solutions when R is large. The solution hugs a boundary, in this case $z = 0$, and is of boundary layer type; ω is such that the Doppler–shifted frequency, $\omega + kU$, is nearly zero in this layer.
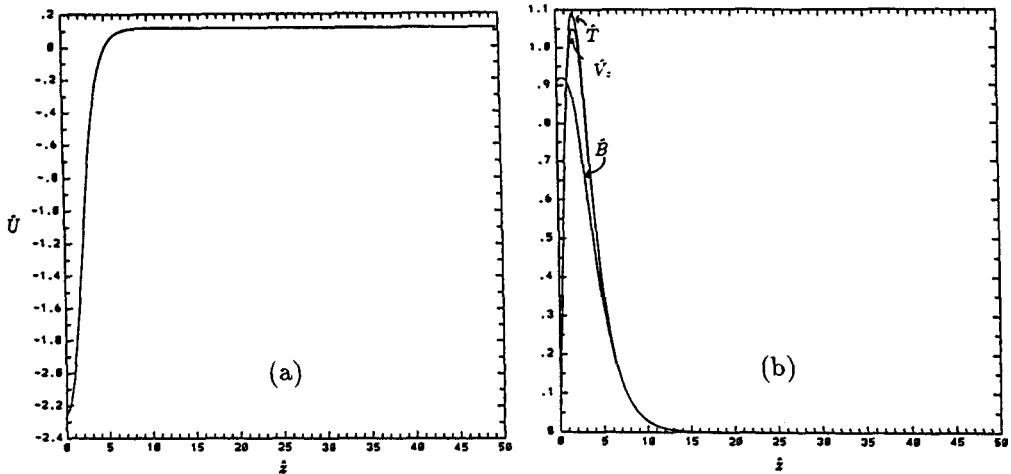
Fig. 8. Free Boussinesq duct (a) geostrophic flow, (b) amplitudes of V_z, B and T. The axes are scaled but in a different way from Figure 6; see paper 2 for details.

5. Finite amplitude compressible convection

No qualitatively new feature emerged when we studied compressible convection in a free duct. The symmetry of the Boussinesq model does not exist in a compressible fluid and, as we saw in Section 3, the convection is necessarily time-dependent. In Paper 1 we found nevertheless that the eigenfuctions of compressible convection are not totally dissimilar from those of the Boussinesq model. It is therefore not altogether surprising that, when the symmetry breaking bifurcation unfolds, one of the two branches that emerged from that pitchfork bifurcation joins smoothly to the linear unicellular solution arising at the smallest critical Rayleigh number, while the other joins to the second smallest, corresponding to bicellular convection. This is most clearly seen by the curves having alternate long and short dashes in Figure 9, which corresponds to the slightly compressible case $z_0 = 1.0$; here β_1 and β_2 correspond to the B_1 and B_2 of the Boussinesq case (full curves). It will be seen that, as R increases, one branch runs roughly parallel to each of the two branches of non-symmetric Boussinesq solutions. This is also true for the more compressible layer, $z_0 = 0.1$, denoted by the dashed curves in Figure 9, and in which b_1 and b_2 corresponds to B_1 and B_2 in the Boussinesq case. The "primary" branches, emerging from β_1 and b_1, are unicellular throughout, and correspond to convection occurring primarily in the lower part of the layer. The "secondary" branches, emerging from β_2 and b_2 are initially bicellular but, as R increases, the upper cells intensify while the lower ones evanesce, until at large R the solution is essentially unicellular, as in the Boussinesq case.
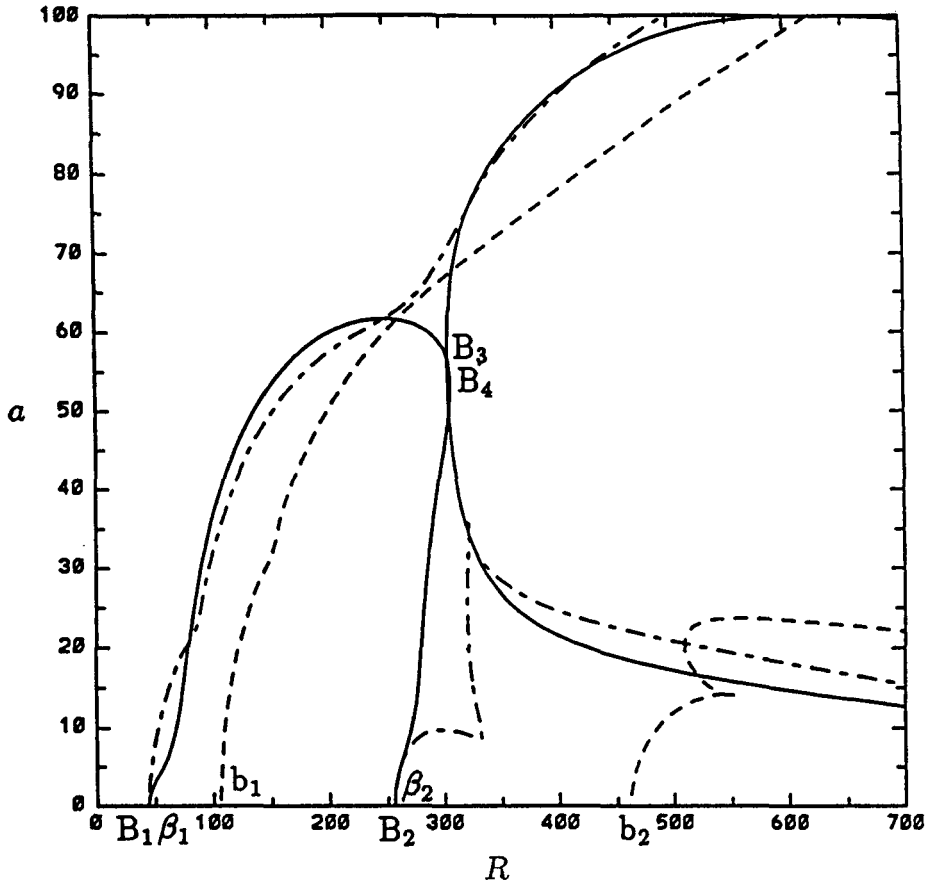
Fig. 9. Bifurcation diagram for free compressible duct.

6. Inconclusions

The present line of investigation, initially set in motion by the claims of Ribes and others (see her paper in this volume), has certainly not run along anticipated lines. There have been three major surprises:

    1: Compressible convection cannot be steady in a rotating fluid;

    2: Taylor's hypothesis fails;

    3: Conditions at distant boundaries have profound effects on the entire flow structure.

    The first surprise was pleasant in one sense, since we believe it represents a modest discovery, the first time that the impossibility of steady convection in a rotating compressible fluid has been noted and explained. In another sense it was an unpleasant surprise, since time-dependent convection is more troublesome to analyze than steady convection.

    The second surprise is of interest particularly to those who have partaken in current research on the theory of the geodynamo, in which the role of Taylor's

constraint is central. Currently the preferred approach to the MHD of this highly rotating system is to introduce a coupling between the Earth's core and its mantle. This guarantees a solution for the MHD state in all circumstances, one in which the geostrophic flow is in general large, e.g. in the case of viscous coupling it is proportional to $\nu^{-1/4}$, and ν is very small in the core ($E \ll 1$); see Section 4 above. This approach does not rule out Taylor states, which may occur naturally as solutions in which the geostrophic flow is not large, being independent of the assumed core-mantle coupling (e.g. independent of ν when the coupling is viscous), but it avoids the pitfall of seeking Taylor states in situations in which, perhaps, none exists. Until now, the discovery of a Taylor state has always been fortuitous; no general rules have been evolved that guarantee its existence or non-existence. Paper 2 represents the first exception to that situation: "anti-Taylor inequalities" were developed that ruled out Taylor states when Λ is too small.

The third surprise was something of a body blow to any like ourselves who hoped to use very simple models, ones that exclude all unnecessary mechanisms, to explain complex MHD behaviors. Amongst the idealizations has been our simplification 1: the elimination of geometrical effects, by supposing that the working fluid occupied a plane layer (later a duct), and by assuming that gravity and the prevailing magnetic field are uniform. In defence of this step, now clearly seen to have been too optimistic, we should point to the studies of the linear stability by Eltayeb and Roberts (1970) and Eltayeb (1972, 1975), which gave results qualitatively insensitive to the relative orientations of $\mathbf{g}$, $\boldsymbol{\Omega}$ and $\mathbf{B}_0$, and which initially encouraged us to believe that geometrical considerations would not be crucial. The discovery of the anti-Taylor inequalities for the present situation, was already a foretaste of an impending disappointment, for it demonstrates that our model behaves at finite amplitude quite differently from the plane layer Boussinesq model in which rotation is parallel to gravity, a configuration extensively studied by Roberts and Stewartson (1974, 1975), Soward (1980, 1986) and Skinner and Soward (1989), and in which Taylor states occur.

Still worse was in store however. Even within the framework of our model, the nature of the distant boundaries has repercussions on the entire MHD state. When the side walls of our duct are no-slip surfaces, the convection becomes increasingly multicellular as the Rayleigh number is increased; if they are stress-free surfaces, the symmetry of the solution is destroyed once R becomes large enough. These are drastic differences, greater than one might have anticipated even with prior experience behind one of the behavior of highly rotating systems, including their universal tendency towards two-dimensionality with respect to the direction of $\boldsymbol{\Omega}$, a tendency epitomised by the celebrated Proudman-Taylor theorem.

By calling the third surprise a "body blow", and describing it in such pessimistic terms, we have in mind the original aims of Papers 1 and 2, that of providing firm theoretical backing for the claims of Ribes and others. As a study of rotating MHD, the investigation has proved to be very exciting, the behaviors of the rigid and free ducts being especially fascinating. So much has emerged from the study that, speaking for myself (and I feel sure for my fellow authors too), the required time and effort have been well spent. It is only when one begins to question how far

it has achieved its stated aim that some disappointment creeps in. The behavior of the two zone model described in Section 3 does contain more than a hint that our basic idea is correct, namely that, for the reasons expounded in Section 1, strong convection in a compressible MHD layer will, in the right range of Elsasser number, be of doughnut form near the top of the layer while being banana–like at depth. The existence of two modes in the free duct described in Section 4 is at first encouraging also, but one's excitement diminishes somewhat when (Section 5) he finds that it is the bottom mode that connects to the primary branch of solutions, the one emerging from β_1 and B_1 in Figure 9.

With hindsight it seems that simplification 1 was too drastic, and has taken us too far away from the real Sun. What should the next step be? Perhaps to investigate strong convection in a spherical model, in which the imposed $\mathbf{B}_0$ is zonal? Hopefully, the behavior of such a model will, at moderate R, resemble that of our free duct. (The rigid duct would represent better a latitude band of the spherical system where the flow is brought to rest at the confining edges of the band, a situation remote from the solar application.) We look to the future with keen anticipation!

References

Braginsky, S.I.: 1975, *Geomag. Aeron.* **15**, 122

Busse, F.H.: 1970, *Astrophys. J.* **159**, 629

Carrigan, C.R., Busse, F.H.: 1983, *J. Fluid Mech.* **126**, 287

Eltayeb, I.A.: 1972, *Proc. R. Soc. Lond. A* **326**, 229

Eltayeb, I.A.: 1975, *J. Fluid Mech.* **71**, 161

Eltayeb, I.A., Roberts, P.H.: 1970, *Astrophys. J.* **162**, 699

Jones, C.A., Roberts, P.H., Galloway, D.J.: 1990, *Geophys. Astrophys. Fluid Dynam.* **53**, 145. This paper was referred to as 'Paper 1' above.

Jones, C.A., Roberts, P.H.: 1990, *Geophys. Astrophys. Fluid Dynam.*, to appear (1990). This paper was referred to as 'Paper 2' above.

Hart, J.E., Toomre, J. Deane, A. E., Hurlburt, N. E., Glatzmaier, G. A., Fichtel, G. H., Leslie, F., Fowlis, W. W., Gilman, P. A.: 1986, *Science* **234**, 61

Ribes, E., Mein, P., Mangeney, A.: 1985, *Nature* **318**, 170

Roberts, P.H.: 1968, *Phil Trans. R. Soc. Lond.* **A263**, 93

Roberts, P.H., Stewartson, K.: 1974, *Phil. Trans. R. Soc. Lond.* **A277**, 287

Roberts, P.H., Stewartson, K.: 1975, *J. Fluid Mech.* **68**, 447

Skinner, P.H., Soward, A.M.: 1989, *Geophys. Astrophys. Fluid Dynam.* **44**, 91

Soward, A.M.: 1980, *J. Fluid Mech.* **98**, 449

Soward, A.M.: 1986, *Geophys. Astrophys. Fluid Dynam.* **35**, 329

Taylor, J. B.: 1964, *Proc. R. Soc. Lond.* **A274**, 274