6

Dynamics at Moderate to High Magnetic Reynolds' Number

> . . . and to those philosophers who pursue the inquiry (of induction) zealously yet continuously, combining experiment with analogy, suspicious of their preconceived notions, paying more respect to the fact than a theory, not too hasty to generalise, and above all things, willing at every step to cross-examine their own opinions, both by reasoning and experiment, no branch of knowledge can afford so fine and ready a field for discovery as this.
>
> *Faraday (1837)*

When R_m is high there is a strong influence of $\mathbf{u}$ on $\mathbf{B}$, and so we obtain a two-way coupling between the velocity and magnetic fields. The tendency for $\mathbf{B}$ to be advected by $\mathbf{u}$, which follows directly from Faraday's law of induction, results in a completely new phenomenon, the Alfvén wave. It also underpins existing explanations for the origin of the earth's magnetic field and of the solar field. We discuss both of these topics below. First, however, it may be useful to comment on the organisation of this chapter.

The subject of high-R_m MHD is vast, and clearly we cannot begin to give a comprehensive coverage in only one chapter. There are many aspects to this subject, each of which could, and indeed has, filled textbooks and monographs. Our aim here is merely to provide the beginner with a glimpse of some of the issues involved, offering a stepping-stone to more serious study. The subject naturally falls into three or four main categories. There is the ability of magnetic fields to support inertial waves, both Alfvén waves and magnetostrophic waves. (The latter involves Coriolis forces, the former does not.) This topic is reasonably clear-cut. Then there is geodynamo theory, which attempts to explain the maintenance of the earth's magnetic field in terms of a self-excited fluid dynamo. This is anything but clear-cut! Geodynamo theory is complex and difficult and there exist many unresolved issues. Next there is plasma containment, motivated for the most part by fusion applications. Here much of the interest lies in the stability of magnetic equilibria, and this is now reasonably well understood, or at least as far as linear (small ampli-

tude) stability is concerned. Finally, there is astrophysical MHD, particularly topics such as star formation and magnetic phenomena in the sun: field oscillations, sunspots, solar flares and so on. Like geodynamo theory, the picture here is far from complete.

The layout of the chapter is as follows. We start with the simplest topic, that of wave theory. We then move to the geodynamo. This divides naturally into two parts. There is the simpler and largely understood kinematic aspect, and the altogether more difficult topic of dynamical theories. We restrict ourselves here to the kinematics of geodynamo theory, where perhaps there is less controversy. Next we give a brief and entirely qualitative tour of one or two aspects of solar MHD. There is no pretence here of a mathematical dissection of the issues involved. The discussion is purely descriptive. We end with a discussion of MHD equilibria. Although the motivation here is plasma MHD, we (artificially) restrict ourselves to incompressible fluids. The reason for this is simple: the algebra involved in developing stability theorems for even incompressible fluids is lengthy and somewhat tedious, and so it seems inappropriate in such a brief discussion to embrace all the additional complexities of compressibility.

Finally, perhaps it is worthwhile to comment on the notation employed in this chapter. Throughout this text we employ only cylindrical polar coordinates (r, θ, z). We make no use of spherical polar coordinates (r', θ', ϕ). When dealing with axisymmetric fields in cylindrical polar coordinates it is natural to divide a vector field, say $\mathbf{B}$, into *azimuthal* $(0, B_\theta, 0)$ and *poloidal* $(B_r, 0, B_z)$ components. For example, the dipole-like external field of the earth is (more or less) poloidal. The field induced by a long straight wire is azimuthal. In the geophysical and astrophysical literature it is normal to use a different terminology. The field is divided into *toroidal* and *poloidal* components. When the field is axisymmetric, toroidal is equivalent to azimuthal. Occasionally the term meridional is substituted for poloidal. We shall not employ the terms toroidal or meridional.

6.1 Alfvén Waves and Magnetostrophic Waves

6.1.1 Alfvén waves

One of the remarkable properties of magnetic fields in MHD is that they can transmit transverse inertial waves, just like a plucked string. We have already discussed the physical origin of this phenomenon. It relies on the

fact that the **B**-field and fluid are virtually frozen together when σ is high. To give an illustration, suppose that a portion of a field line is swept sideways by the lateral movement of the fluid (Figure 6.1). The resulting curvature of the field line gives rise to a restoring force, $B^2/\mu R$, as discussed in §3.9. (R is the radius of curvature of the field line.) As the curvature increases, the restoring force rises and eventually the inertia of the fluid is overcome and the lateral movement is stopped. However, the Lorentz force is still present, and so the flow now reverses, carrying the field lines back with it. Eventually, the field lines return to their equilibrium position, only now the inertia of the fluid carries the field lines past the neutral point and the whole process starts in reverse. Oscillations then develop, and this is called an Alfvén wave.

We now place our physical intuition on a formal mathematical basis. Suppose we have a uniform, steady magnetic field $\mathbf{B}_0$ which is perturbed by an infinitesimally small velocity field $\mathbf{u}$. Let $\mathbf{j}$ and $\mathbf{b}$ be the resulting perturbations in current density and $\mathbf{B}$. Then the leading order terms in the induction equation are

$$\frac{\partial \mathbf{b}}{\partial t} = \nabla \times (\mathbf{u} \times \mathbf{B}_0) + \lambda \nabla^2 \mathbf{b}, \qquad \nabla \times \mathbf{b} = \mu \mathbf{j}$$

which yields

$$\frac{\partial \mathbf{j}}{\partial t} = \frac{1}{\mu}(\mathbf{B}_0 \cdot \nabla)\boldsymbol{\omega} + \lambda \nabla^2 \mathbf{j} \tag{6.1}$$

We now consider the momentum of the fluid. Since $(\mathbf{u} \cdot \nabla)\mathbf{u}$ is quadratic in the small quantity $\mathbf{u}$ it may be neglected in the Navier–Stokes equation and so we have, to leading order in the amplitude of the perturbation,

$$\rho \frac{\partial \mathbf{u}}{\partial t} = \mathbf{j} \times \mathbf{B}_0 - \nabla p + \rho \nu \nabla^2 \mathbf{u}$$

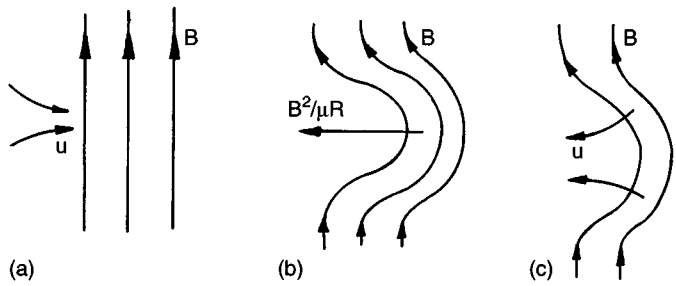

Figure 6.1 Formation of an Alfvén wave.

The equivalent vorticity equation is

$$\frac{\partial \boldsymbol{\omega}}{\partial t} = \frac{1}{\rho}(\mathbf{B}_0 \cdot \nabla)\mathbf{j} + \nu \nabla^2 \boldsymbol{\omega} \tag{6.2}$$

We now eliminate $\mathbf{j}$ from (6.2) by taking the time derivative and then substitute for $\partial \mathbf{j}/\partial t$ using (6.1). After a little algebra, we find

$$\frac{\partial^2 \boldsymbol{\omega}}{\partial t^2} = \frac{1}{\rho\mu}(\mathbf{B}_0 \cdot \nabla)^2 \boldsymbol{\omega} + (\nu + \lambda)\nabla^2\left(\frac{\partial \boldsymbol{\omega}}{\partial t}\right) - \lambda\nu\nabla^4 \boldsymbol{\omega} \tag{6.3}$$

Next we look for plane-wave solutions of the form

$$\boldsymbol{\omega} \sim \boldsymbol{\omega}_0 \exp[i(\mathbf{k} \cdot \mathbf{x} - 2\pi f t)] \tag{6.4}$$

where $\mathbf{k}$ is the wavenumber. Substituting (6.4) into our wave equation (6.3) gives the dispersion relationship

$$2\pi f = -\left[(\nu + \lambda)k^2/2\right]i \pm \left[B_0^2 k_{//}^2/(\rho\mu) - (\nu - \lambda)^2 k^4/4\right]^{1/2}$$

Here $k_{//}$ is the component of $\mathbf{k}$ parallel to $\mathbf{B}_0$. There are three special cases of interest. When $\lambda = \nu = 0$ (a perfect fluid) we obtain $2\pi f = \pm v_a k_{//}$ where v_a is the Alfvén velocity $B_0/(\rho\mu)^{1/2}$. This represents the propagation of transverse inertial waves, with phase velocity v_a. When $\nu = 0$ and λ is small but finite, which is a good approximation in stars, and for liquid-metal flows at high R_m, we find

$$2\pi f = -(\lambda k^2/2)i \pm v_a k_{//}$$

This represents a plane wave propagating with phase velocity v_a and damped by Ohmic dissipation. Finally, we consider the low-R_m case of $\nu = 0$, $\lambda \to \infty$, which characterises most of liquid-metal MHD. Here we find that

$$2\pi f = -i\lambda k^2, \qquad 2\pi f = -i\tau^{-1} k_{//}^2/k^2$$

where τ is the magnetic damping time $\left(\sigma B^2/\rho\right)^{-1}$. The first of these solutions represents a disturbance which is rapidly eradicated by Ohmic dissipation. This is of little interest. However, the second solution represents a non-oscillatory disturbance which decays rather more slowly, on a time scale of τ (Figure 6.2). This is the origin of the pseudo-diffusion phenomenon discussed in Chapter 5.

Alfvén waves are of little importance in liquid-metal MHD since R_m is usually rather modest in such cases. However, they are of considerable importance in astrophysical MHD, where they provide an effective mechanism for propagating energy and momentum. For example, it

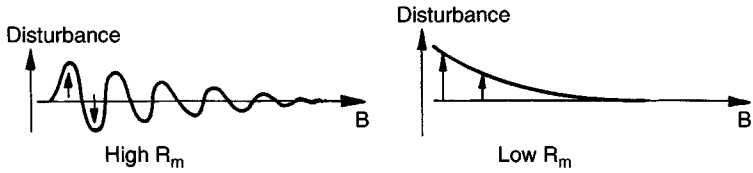

Figure 6.2 Damped Alfvén waves at low and high R_m. Note the low R_m wave is non-oscillatory.

has been suggested that they are responsible for transporting angular momentum away from the core of an interstellar cloud which is collapsing to form a star under the influence of self-gravitation.

Example: Finite-amplitude Alfvén waves

Show that finite-amplitude solutions of the ideal induction equation and Euler's equation exist in the form

$$\mathbf{u} = \mathbf{f}(\mathbf{x} - \mathbf{h}_0 t), \qquad \mathbf{h} = \mathbf{h}_0 - \mathbf{f}(\mathbf{x} - \mathbf{h}_0 t)$$

or

$$\mathbf{u} = \mathbf{g}(\mathbf{x} + \mathbf{h}_0 t), \qquad \mathbf{h} = \mathbf{h}_0 + \mathbf{g}(\mathbf{x} + \mathbf{h}_0 t)$$

where $\mathbf{h} = \mathbf{B}/(\rho\mu)^{1/2}$, $\mathbf{h}_0 =$ constant, and $\mathbf{f}$ and $\mathbf{g}$ are arbitrary solenoidal vector fields.

6.1.2 Magnetostrophic waves

There is a second type of inertial wave motion which magnetic fields can sustain. These are called magnetostrophic waves, indicating that they involve both magnetic and rotational effects. Suppose that we repeat the calculation of §6.1.1. This time, however, we let the fluid rotate and take the quiescent base state to be in a rotating frame of reference, rotating at $\boldsymbol{\Omega}$. The effect of moving into a rotating frame of reference is to introduce a centripetal acceleration, which is irrotational and so merely augments the fluid pressure, and a Coriolis force $2\mathbf{u} \times \boldsymbol{\Omega}$. Neglecting dissipative effects, our governing equations are now

$$\frac{\partial \mathbf{b}}{\partial t} = \nabla \times (\mathbf{u} \times \mathbf{B}_0)$$

$$\rho \frac{\partial \mathbf{u}}{\partial t} = 2\rho \mathbf{u} \times \boldsymbol{\Omega} + \mathbf{j} \times \mathbf{B}_0 - \nabla p$$

so that (6.1) and (6.2) become

$$\frac{\partial \mathbf{j}}{\partial t} = \frac{1}{\mu}(\mathbf{B}_0 \cdot \nabla)\boldsymbol{\omega}$$

$$\frac{\partial \boldsymbol{\omega}}{\partial t} = 2(\boldsymbol{\Omega} \cdot \nabla)\mathbf{u} + \frac{1}{\rho}(\mathbf{B}_0 \cdot \nabla)\mathbf{j}$$

We now proceed as before and eliminate $\mathbf{j}$ to give

$$\frac{\partial^2 \boldsymbol{\omega}}{\partial t^2} = 2(\boldsymbol{\Omega} \cdot \nabla)\frac{\partial \mathbf{u}}{\partial t} + \frac{1}{\rho\mu}(\mathbf{B}_0 \cdot \nabla)^2 \boldsymbol{\omega}$$

There are three special cases of interest. First, if $\boldsymbol{\Omega} = 0$ we arrive back at (6.3), representing undamped Alfvén waves. Second, if $\mathbf{B}_0 = 0$, then we find

$$\frac{\partial^2}{\partial t^2}\left(\nabla^2 \mathbf{u}\right) + 4(\boldsymbol{\Omega} \cdot \nabla)^2 \mathbf{u} = 0$$

This represents conventional *inertial waves* – waves which propagate in rotating fluids (Figure 6.3). For disturbances of the form $\hat{\mathbf{u}} \exp[j(\mathbf{k} \cdot \mathbf{x} - 2\pi f t)]$ this yields the dispersion relationship

$$2\pi f = \pm 2(\boldsymbol{\Omega} \cdot \mathbf{k})/|\mathbf{k}|$$

and a group velocity (the velocity at which wave energy propagates) of

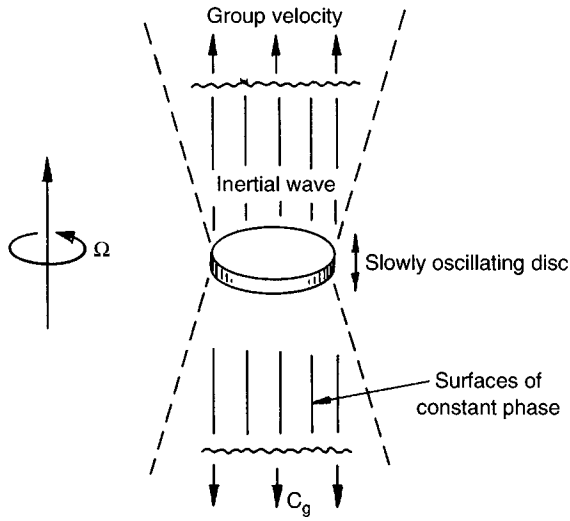

Figure 6.3 An inertial wave.

$$\mathbf{c}_g = \pm 2(k^2\boldsymbol{\Omega} - \mathbf{k}(\mathbf{k}\cdot\boldsymbol{\Omega}))k^{-3}$$

Note that the group velocity is perpendicular to $\mathbf{k}$, so that the phase velocity and group velocity are mutually perpendicular. Thus a wave appearing to travel in one direction, according to the surfaces of constant phase, is actually propagating energy in a perpendicular direction.

Evidently, slow disturbances ($f \ll |\boldsymbol{\Omega}|$) correspond to $\boldsymbol{\Omega}\cdot\mathbf{k} \approx 0$ and $\mathbf{c}_g \approx \pm 2\boldsymbol{\Omega}/|\mathbf{k}|$. Such disturbances propagate as wave packets in the $+\boldsymbol{\Omega}$ and $-\boldsymbol{\Omega}$ directions, and the net effect is that the disturbance appears to elongate along the rotational axis, leading to columnar structures called Taylor columns (Figure 6.4). More generally, the frequency of inertial waves varies from zero, when the group velocity is aligned with $\boldsymbol{\Omega}$, to $2|\boldsymbol{\Omega}|$, when the group velocity is normal to $\boldsymbol{\Omega}$.

The third case of interest is when both $\mathbf{B}_0$ and $\boldsymbol{\Omega}$ are finite but $f \ll |\boldsymbol{\Omega}|$ – slow waves. In this case

$$2(\boldsymbol{\Omega}\cdot\nabla)\frac{\partial \mathbf{u}}{\partial t} + \frac{1}{\rho\mu}(\mathbf{B}_0\cdot\nabla)^2\boldsymbol{\omega} = 0$$

which, on application of the operator $(\boldsymbol{\Omega}\cdot\nabla)\frac{\partial}{\partial t}$, yields

$$4(\boldsymbol{\Omega}\cdot\nabla)^2\frac{\partial^2 \mathbf{u}}{\partial t^2} + \left[\frac{1}{\rho\mu}(\mathbf{B}_0\cdot\nabla)^2\right]^2\nabla^2\mathbf{u} = 0$$

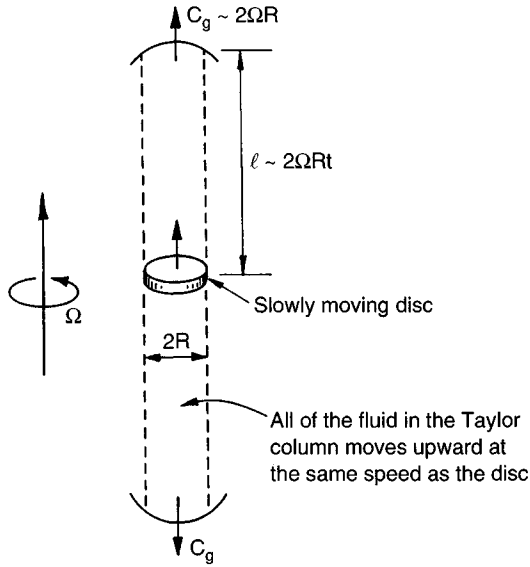

Figure 6.4 Formation of Taylor columns by inertial waves.

This is the governing equation for magnetostrophic waves. The corresponding dispersion equation is

$$2\pi f = \pm\frac{1}{\rho\mu}(\mathbf{B}_0 \cdot \mathbf{k})^2 k/[2(\boldsymbol{\Omega} \cdot \mathbf{k})]$$

Since we require $f \ll |\boldsymbol{\Omega}|$, such waves can exist only if

$$(\boldsymbol{\Omega} \cdot \mathbf{k})^2 \gg (\mathbf{B}_0 \cdot \mathbf{k})^2 k^2/\rho\mu$$

so in some sense we are considering cases where the Coriolis effect is dominant. Magnetostrophic waves are significant in solar and geophysical MHD since both the sun and the earth are rapidly rotating and the Coriolis force is dominant.

6.2 Elements of Geo-Dynamo Theory

> Where does the earth's magnetic field come from? Nobody really knows – there have only been some good guesses.
>
> *R P Feynman (1964) Lectures on Physics*

6.2.1 Why do we need a dynamo theory for the earth?

Dynamo theory is the name given to the process of magnetic field generation by the inductive action of a conducting fluid, i.e. the conversion of mechanical energy to magnetic energy through the stretching and twisting of the magnetic field lines. It is generally agreed that this is the source of the earth's magnetic field, since the temperature of the earth's interior is well above the Curie point at which ferro-magnetic material loses its permanent magnetism. Moreover, the earth's magnetic field cannot be the relic of some primordial field trapped within the interior of the earth. Such a field would long ago have decayed. To see why this is so, suppose that there is negligible motion in the earth's core. The product of **B** with Faraday's law yields the energy equation

$$\frac{\partial}{\partial t}\left(\frac{B^2}{2\mu}\right) = -\nabla \cdot [(\mathbf{E} \times \mathbf{B})/\mu] - \mathbf{J}^2/\sigma \tag{6.5}$$

(Here we have used the fact that $\mathbf{B} \cdot (\nabla \times \mathbf{E}) = (\nabla \times \mathbf{B}) \cdot \mathbf{E} + \nabla \cdot [\mathbf{E} \times \mathbf{B}]$.) We now integrate over all space and note that there is no flux of the Poynting vector, $\mathbf{E} \times \mathbf{B}$, out of a sphere of infinite radius. The result is

$$\frac{dE_B}{dt} = -\int (\mathbf{J}^2/\sigma) dV \tag{6.6}$$

where E_B is the energy of the magnetic field and the integral on the right is confined to $r < R_c$, R_c being the outer radius of the earth's conducting core. As we might have anticipated, E_B falls due to Ohmic dissipation. Now the rate of decline of E_B may be found by a normal mode analysis in which the diffusion equation defines an eigenvalue problem for **B**. It turns out that this yields a decay time of $t_d \sim R_c^2/(\lambda\pi^2)$. For the earth, we have $R_c \sim 3500\,\text{km}$ and $\lambda \sim 2\,\text{m}^2/\text{s}$, which gives $t_d \sim 10^4$ years[1]. However, the earth's magnetic field has been around for at least 10^8 years, and so there must be some additional mechanism maintaining E_B despite the Ohmic losses. Just such a mechanism was discussed in § 4.3: the stretching of flux tubes by an imposed velocity field. In fact, it is not just the earth's magnetic field which is thought to arise from dynamo action. Virtually all of the planets as well as the sun have magnetic fields, many of which are likely to be maintained by a self-excited, fluid dynamo.

Historically there have been many attempts to explain the origins of the earth's magnetic field, other than MHD. Now all abandoned, these included a magnetic mantle, the Hall effect, thermoelectric effects, rotating electrostatic charges, and even, as a last act of desperation, a proposed modification to Maxwell's equations. The electrostatic argument arises from the fact that the earth's surface is negatively charged. In fact, this charge is so great that near the earth's surface there exists an atmospheric electric field of $\sim 100\,\text{V/m}$, directed, on average, radially inward. This surface charge is maintained by lightning storms that are charging the earth, relative to the upper atmosphere, with an average current of 1800 Amperes!

One by one these theories have been abandoned, often because they failed on an order of magnitude basis. Kelvin was on the right track when, in 1867, he noted:

> We may imagine, as Gilbert did, the Earth to be wholly or in part a magnet, such as a magnet of steel, or we may conceive it to be an electromagnet, with or without a core susceptible of induced magnetism. In the present state of our knowledge this second hypothesis seems to be the more probable and indeed we now have reasons for believing in terrestrial currents.... The question which occurs now is this:- Can the magnetic phenomena at the earth's surface, and above it, be produced by an internal distribution of closed currents occupying a certain limited space below the surface.

[1] This calculation was first performed by H Lamb in 1889.

The problem, of course, is what maintains the currents. It was Larmor who, in 1919, first suggested a self-excited fluid dynamo (in the solar context) in his paper: 'How could a rotating body like the sun become a magnet?' When, in 1926, Jeffreys discovered the liquid core of the earth, Larmor's idea suddenly became very relevant to the geo-dynamo.

The general idea behind geo-dynamo theory is that fluid motion in the earth's core stretches and twists the magnetic field lines, thus intensifying the magnetic field. This relies on the advection term in the induction equation being dominant, which in turn requires that R_m is large. However, this seems quite likely. The earth has a liquid iron annulus of inner radius $\sim 10^3$ km and outer radius $\sim 3 \times 10^3$ km (see Figure 6.7). Typical scales for $\mathbf{u}$ and l are estimated to be $u \sim 2 \times 10^{-4}$ m/s and $l \sim 10^3$ km, giving $R_m \sim 100$: not massive, but large.

Mechanical analogues of a self-excited dynamo are readily constructed. A simple, and common, example is the homopolar disk dynamo, shown in Figure 6.5. Here a solid metal disk rotates with a steady angular velocity $\boldsymbol{\Omega}$ and a current path between its rim and its axis is provided by a wire twisted as shown. It is readily confirmed that, provided $\boldsymbol{\Omega}$ is large enough, any small magnetic field which exists at $t = 0$ will grow exponentially in time. First we note that rotation of the disk results in an e.m.f. of $\Omega\Phi/2\pi$ between the axis and the edge of the disk, Φ being the magnetic flux which links the disk. (This may be confirmed by the use of Faraday's law (2.28), or else by a consideration of Ohm's law.) This e.m.f. drives a current, I, which evolves according to

$$L\frac{dI}{dt} + RI = \text{e.m.f.} = \Omega MI/2\pi$$

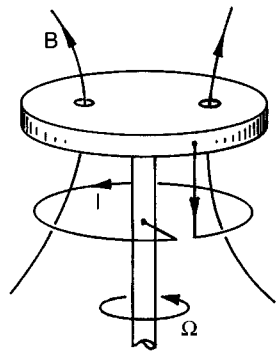

Figure 6.5 Homopolar disk dynamo.

Here M is the mutual inductance of the current loop and the rim of the disk, and L and R are the self-inductance and resistance, respectively, of the total circuit. Evidently I, and hence $\mathbf{B}$, grow exponentially whenever $\boldsymbol{\Omega}$ exceeds $2\pi R/M$. This increase in magnetic energy is accompanied by a corresponding rise in the torque, T, required to drive the disk, the source of the magnetic energy being the mechanical power, $T\boldsymbol{\Omega}$. In any real situation, however, this exponential rise in T cannot be maintained for long, and eventually the applied torque will fall below that needed to maintain a constant $\boldsymbol{\Omega}$. At this point the disk will slow down, eventually reaching the critical level of $\Omega = 2\pi R/M$. T, $\boldsymbol{\Omega}$ and $\mathbf{B}$ then remain steady.

Now this kind of mechanical device differs greatly from a fluid dynamo because the current is directed along a carefully constructed path. Nevertheless a highly conducting fluid can stretch and twist a magnetic field so as to intensify E_B (see Figure 6.6, for example). The central questions in dynamo theory are therefore: (i) can we construct a steady (or steady-on-average) velocity field which leads to dynamo action?; (ii) can such a velocity field be maintained by, say, Coriolis or buoyancy forces in the face of the Lorentz force which, presumably, tends to slow the fluid down? It is now generally agreed that the answers to these questions are 'yes' and 'probably', respectively. However, it should

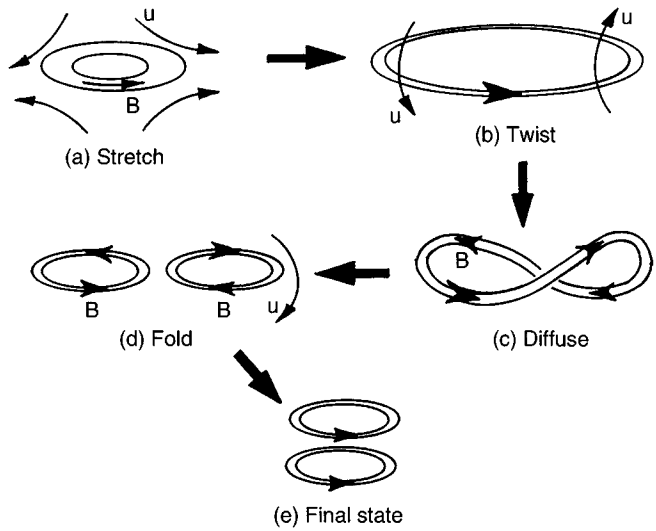

Figure 6.6 A magnetic field can be intensified by a sequence of operations: stretch + twist + diffuse + fold.

be said from the outset that there is, as yet, no self-consistent model of the geo-dynamo. In fact, the entire subject is shrouded in controversy!

In the following subsections we sketch out some of the more elementary ideas and results in geo-dynamo theory. The key points are:

(i) R_m must be large for dynamo action;
(ii) an axisymmetric geo-dynamo is not possible;
(iii) differential rotation in the earth's core can (if it exists) spiral out an azimuthal magnetic field from the familiar dipole one, and in fact this azimuthal field could well be the dominant field in the interior of the earth;
(iv) small-scale turbulence tends to tease out a (small-scale) magnetic field from the large-scale azimuthal field;
(v) this small-scale, random magnetic field is thought (by some) to organise itself in such a way as to reinforce the large-scale dipole field.

In short, one candidate for a geo-dynamo is: dipole field plus differential rotation → azimuthal field; azimuthal field plus turbulence → small-scale, chaotic field; re-organisation of small-scale field → dipole field. However, this model (called the α–$\boldsymbol{\Omega}$ model) is somewhat speculative and, as we shall see, alternatives have been proposed.

The central rôle played by the azimuthal field here is, at first sight, somewhat surprising. After all, measurements made at the surface of the earth indicate only a dipole field. However, it should be remembered that the azimuthal field is supported by poloidal currents which are confined to the core of the earth (Figure 6.7) and that Ampère's law

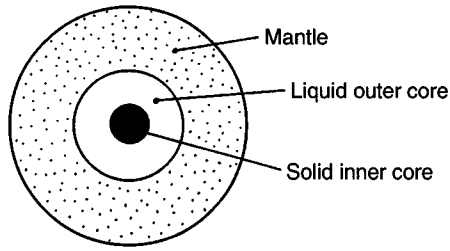

Figure 6.7 The core of the earth. The earth has solid inner core of iron/nickel alloy (radius $\sim 10^3$ km), a liquid outer core of iron plus some lighter elements (radius $\sim 3 \times 10^3$ km), and an outer mantle of ferro-magnesium silicates (radius $\sim 6 \times 10^3$ km).

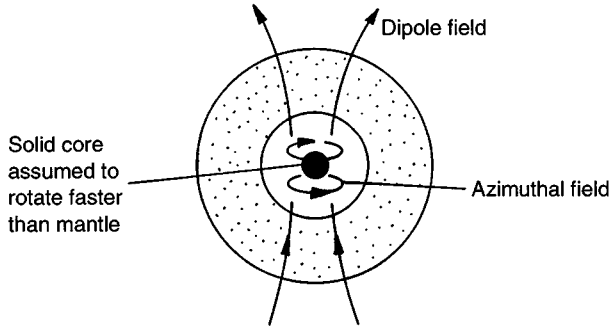

Figure 6.8 Generation of an azimuthal field by differential rotation in the core of the earth.

tells us that such a field cannot extend beyond the core–mantle boundary (see Figure 6.10). The likely source of the azimuthal field is differential rotation between the inner regions of the core and the remainder of the earth (Figure 6.8). This sweeps out B_θ from the dipole field, $\mathbf{B}_p$, and order of magnitude analysis suggests, $B_\theta \sim R_m B_p$ (see Chapter 4 Section 5.3).[2]

6.2.2 A large magnetic Reynolds number is needed

Now it is clear that dynamo action will occur only if R_m is large enough, since the intensification of E_B by flux tube stretching has to outweigh, or at least match, the decay of E_B through Ohmic losses. This may be quantified as follows. Starting with Faraday's equation, $\partial \mathbf{B}/\partial t = -\nabla \times \mathbf{E}$, we may show that,

$$\frac{\partial}{\partial t}\left(\frac{B^2}{2\mu}\right) = -\mathbf{J}\cdot\mathbf{E} - \nabla\cdot[(\mathbf{E}\times\mathbf{B})/\mu]$$

We now integrate over all space, noting that the Poynting flux integrates to zero. This yields

$$\frac{dE_B}{dt} = -\int \mathbf{J}\cdot\mathbf{E}dV$$

[2] The sense of the earth's magnetic field regularly reverses. For simplicity we shall take the field to point from south to north.

Next we assume that **J** and **u** are confined to a sphere $r < R$, and that **u** vanishes at $r = R$. Using Ohm's law to rewrite $\mathbf{J} \cdot \mathbf{E}$ as

$$\mathbf{J} \cdot \mathbf{E} = (\mathbf{J}^2/\sigma) - [\mathbf{B} \times (\nabla \times \mathbf{B})] \cdot (\mathbf{u}/\mu)$$

we have

$$\frac{dE_B}{dt} = \frac{1}{\mu}\int \mathbf{u} \cdot [\mathbf{B} \times (\nabla \times \mathbf{B})]dV - \frac{1}{\sigma}\int \mathbf{J}^2 dV = P - D \tag{6.7}$$

The first term on the right is (minus) the rate of working of the Lorentz force, while the second is, of course, the Joule dissipation. To maintain a magnetic field the first integral must be positive. The next step is to place bounds on D, the dissipation integral, and P, the (so-called) production integral. Starting with P we have

$$\mu^2 P^2 \leq u_{\max}^2 \left[\int \mathbf{B} \times (\nabla \times \mathbf{B})dV\right]^2 \leq u_{\max}^2 \int \mathbf{B}^2 dV \int (\nabla \times \mathbf{B})^2 dV$$

where $u_{\max}$ is the maximum value of $|\mathbf{u}|$ and the second inequality comes from the Schwartz inequality introduce in § 5.3. It follows that

$$|P| \leq (2/\lambda)^{1/2} u_{\max} E_B^{1/2} D^{1/2}$$

Also, by standard methods of the calculus of variations, it may be shown that

$$D \geq 2\pi^2 (\lambda/R^2) E_B$$

(The idea here is that $\int (\nabla \times \mathbf{B})^2 dV$ has a minimum value of $\sim \int \mathbf{B}^2 dV / L_{\max}^2$ where $L_{\max}$ is the maximum relevant length scale. In this case it so happens that $L_{\max}$ is R/π.) Combining the two inequalities yields

$$\frac{dE_B}{dt} \leq (2E_B D/\lambda)^{1/2} [u_{\max} - \pi\lambda/R]$$

Evidently, a necessary (though not sufficient) condition for dynamo action is

$$R_m = (u_{\max} R/\lambda) > \pi \tag{6.8}$$

Below this value of R_m the stretching of the field lines cannot compete with the Ohmic losses.

So how much larger than π must R_m be to get a dynamo, 2π or 200π? The answer is: not much larger, perhaps 5π. The simplest known dynamo is that of Ponomarenko and was developed in the 1970s. It consists of helical pipe flow of the form $\mathbf{u} = (0, \Omega r, V)$, in (r, θ, z) coordinates. Here

Ω and V are constants. For such a flow the induction equation admits separable solutions of the type $\mathbf{B} \sim \exp[j(m\theta + kz + \omega t)]$ and the resulting eigenvalue problem yields growing solutions when $R_m = R(V^2 + \Omega^2 R^2)^{1/2}/\lambda$ exceeds 17.72 (here R is the pipe radius). These growing fields are asymmetric, despite the symmetry of the base flow, with $m = 1$, $V \sim 1.31\Omega R$, and $kR \sim -0.388$.

Later, in the 1980s, generalisations of the Ponomarenko dynamo were developed, in which Ω and V are functions of r and $\Omega = V = 0$ at $r = R$. This avoids the singular behaviour at the pipe wall, inherent in Ponomarenko's dynamo. Yet another variant was developed which has a finite length, a return path being provided for the fluid. In fact, this latter model was put to the test in a laboratory in Riga, but insufficiently high values of R_m were achieved to get a self-sustaining dynamo. Undaunted, the Latvian scientists plan a second attempt at creating a fluid dynamo at the laboratory scale (albeit in a very large laboratory!). A similar experiment was undertaken in Karlsruhe, Germany with successful results, and so the next few years should prove very interesting to dynamo enthusiasts.

Note that the Ponomarenko dynamo has helicity $h = \mathbf{u} \cdot \boldsymbol{\omega}$. This is no accident. Almost all working dynamo models involve helicity. Note also that the dynamo is not symmetric. Again, as we shall see, this is no accident.

Example: Rate of change of dipole moment

It may be shown that, if currents are contained within a spherical volume, V, then the dipole moment, $\mathbf{m}$, of the current distribution is related to its associated magnetic field by

$$\mathbf{m} = \frac{1}{2}\int_V \mathbf{x} \times \mathbf{J} dV = \frac{3}{2\mu}\int_V \mathbf{B} dV$$

Use Faraday's equation and Ohm's law to show that

$$\frac{2\mu}{3}\frac{d\mathbf{m}}{dt} = \int_S \mathbf{u}(\mathbf{B} \cdot d\mathbf{S}) + \lambda \int_S (\mu\mathbf{J}) \times d\mathbf{S}$$

If $\lambda = 0$, show that a motion which tends to sweep the field lines towards the polar regions will increase $\mathbf{m}$. (This dynamo mechanism is limited though, and ceases to operate when all of the flux lines crossing S are concentrated at the poles.) Note that if $\mathbf{u}$ is zero on S then a finite diffusivity is required to increase $\mathbf{m}$.

6.2.3 An axisymmetric dynamo is not possible

The idea of a competition between flux-tube stretching and Ohmic dissipation allows us to rule out the possibility of a steady, axisymmetric dynamo. This is important because the earth's external magnetic field is essentially a dipole, field and so it is natural to look for a steady, axisymmetric, dynamo in which **B** is poloidal, $(B_r, 0, B_z)$ in (r, θ, z) coordinates, **J** is azimuthal $(0, J_\theta, 0)$ and $\mathbf{u}_p$ is also poloidal. However, a result known as Cowling's neutral-point theorem says that such a dynamo cannot exist. The proof is as follows.

Let us suppose that an axisymmetric dynamo can indeed be realised. In the steady state, Ohm's law gives $\mathbf{J} = \sigma(-\nabla V + \mathbf{u} \times \mathbf{B})$: however, $\nabla^2 V = \mathbf{B} \cdot \boldsymbol{\omega} - \mu \mathbf{u} \cdot \mathbf{J} = 0$ since $\boldsymbol{\omega}$ is azimuthal. It follows that $V = 0$ and so $\mathbf{J} = \sigma(\mathbf{u} \times \mathbf{B})$. Now in an axisymmetric, poloidal field there is always at least one so-called neutral ring, N, where **B** vanishes and the **B**-lines are closed in the neighbourhood of the ring (Figure 6.9). It is clear from Ampère's circuital law applied to a **B**-line in the vicinity of N that **J** is non-zero at N. However $\mathbf{J} = \sigma(\mathbf{u} \times \mathbf{B})$ and so we cannot have a finite current where there is no magnetic field. Such a configuration is therefore impossible. This is Cowling's neutral-point theorem.

It seems, therefore, that the earth's dynamo must involve a fairly complex flow field, and in fact the nature of this field is still a matter of considerable debate.

Now there is an alternative, less elegant, but more informative, way of establishing Cowling's theorem. In fact this second proof goes further, showing that a poloidal field cannot be intensified by flux-tube stretching whenever **u** and **B** are both axisymmetric. This includes cases where **u** comprises both poloidal and azimuthal components. The

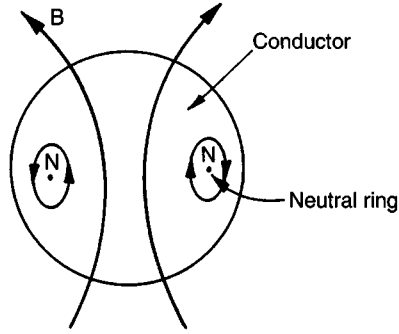

Figure 6.9 All axisymmetric, poloidal fields have a neutral ring, N.

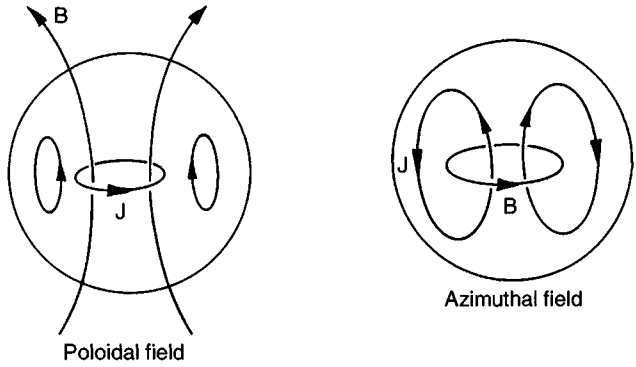

Figure 6.10 Azimuthal and poloidal magnetic fields.

first step is to decompose both **B** and **u** into poloidal and azimuthal parts:

$$\mathbf{B}(r, z) = \mathbf{B}_p + \mathbf{B}_\theta, \qquad \mathbf{u}(r, z) = \mathbf{u}_p + \mathbf{u}_\theta$$

Note that a poloidal magnetic field requires azimuthal currents to support it, while an azimuthal field requires poloidal currents (Figure 6.10). Note also that $\mathbf{B}_\theta$ is restricted to the domain in which the currents flow, which is a direct consequence of Ampère's law.

We now introduce a vector potential for the poloidal field,

$$\mathbf{B}_p = \nabla \times [(\chi/r)\hat{\mathbf{e}}_\theta] = \nabla \times (\mathbf{A}_\theta)$$

which is allowable because $\mathbf{B}_p$ is solenoidal. The parameter χ is called the flux function. It is the magnetic equivalent of the Stokes stream function, the $\mathbf{B}_p$-lines being contours of constant χ. Now it is readily confirmed that the curl of a poloidal field is azimuthal, while the curl of an azimuthal field is poloidal. It follows that the induction equation may also be divided into poloidal and azimuthal components, according to

$$\frac{\partial \mathbf{B}_p}{\partial t} = \nabla \times (\mathbf{u}_p \times \mathbf{B}_p) + \lambda \nabla^2 \mathbf{B}_p \tag{6.9}$$

$$\frac{\partial \mathbf{B}_\theta}{\partial t} = \nabla \times (\mathbf{u}_p \times \mathbf{B}_\theta) + \nabla \times (\mathbf{u}_\theta \times \mathbf{B}_p) + \lambda \nabla^2 \mathbf{B}_\theta \tag{6.10}$$

Uncurling the first of these gives

$$\frac{\partial \mathbf{A}_\theta}{\partial t} = \mathbf{u}_p \times \mathbf{B}_p + \lambda \nabla^2 \mathbf{A}_\theta$$

which yields

$$\frac{D\chi}{Dt} = \lambda \nabla^{*2} \chi \tag{6.11}$$

Here D/Dt is the convective derivative, $\partial/\partial t + (\mathbf{u}_p \cdot \nabla)$, and ∇^{*2} is the Laplacian-like operator defined by

$$\nabla^{*2} = \frac{\partial^2}{\partial z^2} + r \frac{\partial}{\partial r} \frac{1}{r} \frac{\partial}{\partial r} \tag{6.12}$$

The second component of the induction equation may be manipulated into the form

$$\frac{D}{Dt}\left(\frac{B_\theta}{r}\right) = \mathbf{B}_p \cdot \nabla\left(\frac{u_\theta}{r}\right) + \lambda\left[r^{-2}\nabla^{*2}(rB_\theta)\right] \tag{6.13}$$

(This is left as an exercise for the reader.) The important point, as far as Cowling's theorem is concerned, is that (6.11) contains no source term. The flux function is passively advected by $\mathbf{u}_p$, subject only to (a form of) diffusion. There is no term in (6.11) which might lead to an intensification of χ, and hence of $\mathbf{B}_p$. On the contrary, we may use (6.11) to show that $\mathbf{B}_p$ must always decline in accordance with

$$\frac{d}{dt}\int (\chi^2/2\lambda) dV = -\int (\nabla\chi)^2 dV$$

By contrast, (6.13) does contain a source term. Any gradient of swirl along a $\mathbf{B}_p$-line results in the generation of an azimuthal field B_θ. This is readily understood in terms of field sweeping, (see Figure 6.11a).

We might note in passing that (6.13) suggests that, at high R_m, B_θ exceeds $|\mathbf{B}_p|$ in the core of the earth, scaling as $B_\theta \sim R_m|\mathbf{B}_p|$. The point

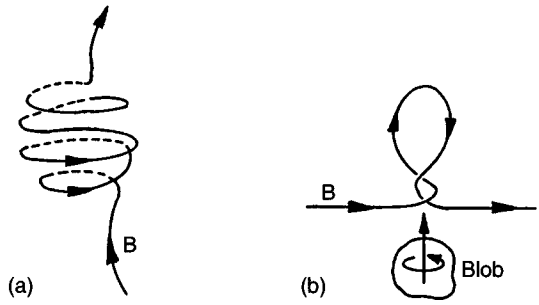

Figure 6.11 An azimuthal magnetic field can be generated by differential rotation in the core of the earth, while turbulence, such as that generated by rising spinning blobs, can tease out a poloidal field from the azimuthal magnetic field. (a) The omega effect. (b) The alpha effect.

is that the left of (6.13) integrates to zero over any volume enclosed (in the r–z plane) by a poloidal stream-line. The two terms on the right must therefore balance (in a global sense), giving $B_\theta \sim (u_\theta l/\lambda)|\mathbf{B}_p|$. Physically, differential rotation causes $\mathbf{B}_\theta$ to build up and, at first, there is nothing to oppose this. (Advection of B_θ/r merely redistributes the azimuthal field.) This process continues until B_θ is so large that diffusion of B_θ is capable of offsetting the generation of an azimuthal field by differential rotation. This happens when $B_\theta \sim (u_\theta l/\lambda)|\mathbf{B}_p|$.

In summary, we conclude that an axisymmetric velocity field cannot intensify a poloidal magnetic field, such as that of the earth, but it can sweep out a (possibly strong) azimuthal field by differential rotation. This is one of the main stumbling blocks to a self-consistent, kinematic, dynamo theory. To complete the dynamo cycle we must find a mechanism of generating a poloidal field from an azimuthal one, i.e.$\mathbf{B}_p \rightarrow \mathbf{B}_\theta \rightarrow \mathbf{B}_p$. Clearly, this last step cannot be axisymmetric.

6.2.4 The influence of small-scale turbulence: the α-effect

This impasse has been circumvented by the suggestion of a two-scale approach to the problem. On the one hand, we might envisage axisymmetric, large-scale behaviour in which $\mathbf{B}_\theta$ is swept out from $\mathbf{B}_p$ through differential rotation between the solid inner core (and its adjacent fluid) and the rest of the liquid annulus. On the other hand, we might postulate small-scale (non-axisymmetric) turbulence which teases out a poloidal field from $\mathbf{B}_\theta$. This would complete the cycle $\mathbf{B}_p \rightarrow \mathbf{B}_\theta \rightarrow \mathbf{B}_p$. This small-scale turbulence might, for example, be generated by natural convection, by shear layers (Ekman layers), by wave motion, or by buoyant, spinning blobs, randomly stretching and twisting the $\mathbf{B}_\theta$-lines as they rise up through the liquid core (Figure 6.11). Such blobs are thought to be released near the inner solid core where relatively pure iron solidifies, leaving liquid rich in an admixture of lighter elements. Whatever the source of the small-scale motion, the general idea is that a random repetition of helical, small-scale events (turbulence, waves or blobs) might lead to the systematic generation of a poloidal field from an azimuthal one. This is known as the 'alpha effect' – a rather obscure name.

Mathematically, the alpha effect may be quantified as follows. Suppose we divide $\mathbf{B}$ and $\mathbf{u}$ into mean and turbulent parts, just as we did when averaging the Navier–Stokes equation in a turbulent flow:

$$\mathbf{B}(\mathbf{x}, t) = \mathbf{B}_0(\mathbf{x}, t) + \mathbf{b}(\mathbf{x}, t)$$

$$\mathbf{u}(\mathbf{x}, t) = \mathbf{u}_0(\mathbf{x}, t) + \mathbf{v}(\mathbf{x}, t)$$

Here $\mathbf{b}$ and $\mathbf{v}$ are the turbulent components which vary rapidly in space and time, whereas $\mathbf{B}_0(\mathbf{x}, t)$ and $\mathbf{u}_0(\mathbf{x}, t)$ vary slowly in space and in time. The means, $\mathbf{B}_0$ and $\mathbf{u}_0$, might for example be defined as spatial averages over a sphere of radius much smaller than R_c (the outer core radius), yet much larger than the scale of the turbulent motion (the blob size in Figure 6.11). The induction equation may also be separated into mean and fluctuating parts:

$$\partial \mathbf{B}_0/\partial t = \nabla \times (\mathbf{u}_0 \times \mathbf{B}_0) + \nabla \times \left(\overline{\mathbf{v} \times \mathbf{b}}\right) + \lambda \nabla^2 \mathbf{B}_0$$

$$\partial \mathbf{b}/\partial t = \nabla \times (\mathbf{u}_0 \times \mathbf{b}) + \nabla \times (\mathbf{v} \times \mathbf{B}_0) + \nabla \times \left(\mathbf{v} \times \mathbf{b} - \overline{\mathbf{v} \times \mathbf{b}}\right) + \lambda \nabla^2 \mathbf{b}$$

The first of these is reminiscent of the time-averaged Navier–Stokes equation, in which the small-scale turbulence has introduced a new term, $\overline{\mathbf{v} \times \mathbf{b}}$, just as Reynolds stresses appear in the averaged momentum equation. The second of these equations is linear in $\mathbf{b}$ with $\nabla \times (\mathbf{v} \times \mathbf{B}_0)$ acting as a source of $\mathbf{b}$. Now suppose that $\mathbf{b} = 0$ at some initial instant. Then the linearity of this equation ensures that $\mathbf{b}$ and $\mathbf{B}_0$ are linearly related. It follows that $\overline{\mathbf{v} \times \mathbf{b}}$ is also linearly related to $\mathbf{B}_0$, and since the spatial scale for $\mathbf{B}_0$ is assumed to be much larger than that of $\mathbf{b}$, we might expect $\overline{\mathbf{v} \times \mathbf{b}}$ to depend mainly on the local value of $\mathbf{B}_0$. This suggests that $\left(\overline{\mathbf{v} \times \mathbf{b}}\right)_i = \alpha_{ij} B_{0j}$, where α_{ij} is some unknown tensor, analogous to an eddy viscosity in turbulence theory. If the turbulence is assumed to be statistically homogeneous and isotropic (something which is unlikely to be true in practice), then $\alpha_{ij} = \alpha \delta_{ij}$ and the mean part of the induction equation becomes

$$\boxed{\partial \mathbf{B}_0/\partial t = \nabla \times (\mathbf{u}_0 \times \mathbf{B}_0) + \alpha \nabla \times \mathbf{B}_0 + \lambda \nabla^2 \mathbf{B}_0} \qquad (6.14)$$

The turbulence has introduced a new term into the mean component of the induction equation. This is the *α-effect*. In effect, the e.m.f. $\overline{\mathbf{v} \times \mathbf{b}}$ has introduced a mean current density $\mathbf{J}$ which we model as $\mathbf{J} = \sigma \alpha \mathbf{B}_0$.

At first sight, the idea of the α-effect may seem a little implausible. Why should small-scale activity give rise to a large-scale magnetic field? One way to think about it is to consider the small-scale e.m.f, $\overline{\mathbf{v} \times \mathbf{b}}$, as like a battery, driving current through the core. If many such e.m.f's are all aligned, then we would expect a large-scale current to emerge from the

cumulative influence of many small-scale eddies. Presumably the partial alignment of the small-scale eddies (which is a dynamic effect) arises from the combined presence of rotation and of the ambient large-scale field. (This is vaguely reminiscent of the alignment of molecular dipoles in ferromagnetic material under the influence of a mean field, thus enhancing the mean field.) Actually, we have already seen this kind of 'small to large-scale' process at work. Recall the example given in § 5.3, in which low R_m turbulence is confined to a sphere of radius R. The fluid is subject to an imposed field B_0 and maintained in (turbulent) motion by some external agency. We showed that, whatever the motion, there is an induced dipole moment, $\mathbf{m}$, whenever the global angular momentum, $\mathbf{H}$, of the turbulence is non-zero:

$$\mathbf{m} = (\sigma/4)\mathbf{H} \times \mathbf{B}_0$$

This is illustrated in Figure 6.12. Now it is a standard result in magnetostatics that whenever electrical currents are confined to a sphere, then the spatial average (over the sphere) of the field associated with those currents is proportional to $\mathbf{m}$:

$$\int_{V_R} \mathbf{b}\,dV = (2\mu/3)\mathbf{m}$$

It follows that

$$\int_{V_R} \mathbf{b}\,dV = (\mathbf{H} \times \mathbf{B}_0)/(6\lambda)$$

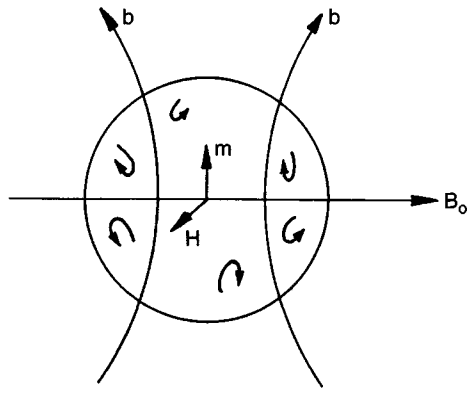

Figure 6.12 Example of small-scale turbulence generating a large-scale field.

Thus we have obtained a (weak) large-scale magnetic field from small-scale turbulence. (Actually, this is not really an α-effect but it does demonstrate the potential for small-scale motion to interact with a mean field to generate a second, large-scale field.)

The α-effect is important in the dynamo context because the poloidal field, $\mathbf{B}_p$, is governed by

$$\frac{\partial \mathbf{B}_p}{\partial t} = \nabla \times (\mathbf{u}_p \times \mathbf{B}_p) + \lambda \nabla^2 \mathbf{B}_p + \alpha \nabla \times \mathbf{B}_\theta$$

or equivalently by

$$\frac{D\chi}{Dt} = \lambda \nabla^{*2} \chi + \alpha r B_\theta$$

The α-effect allows a poloidal field to emerge from an azimuthal one through the action of turbulence. This completes the regenerative cycle $\mathbf{B}_p \rightarrow \mathbf{B}_\theta \rightarrow \mathbf{B}_p$.

The coefficient α, like Boussinesq's eddy viscosity, is a property of the turbulence. We might ask: what properties of the turbulence promote an α-effect, and can we estimate the size of α? In this respect, the most important thing to note is that α is a pseudo-scalar: that is, α changes sign under a coordinate transformation (inversion) of the form $\mathbf{x} \rightarrow -\mathbf{x}$ What do we mean by this? Consider the definition of α: $\overline{\mathbf{v} \times \mathbf{b}} = \alpha \mathbf{B}_0$. Here $\mathbf{v}$ is a polar (true) vector, in the sense that $\mathbf{v}$ always points in the same physical direction, whatever coordinate system is used to describe it. Force is another example of such a vector. However, $\mathbf{b}$ and $\mathbf{B}_0$ are examples of what are called pseudo-vectors, a strange type of vector that reverses physical direction, although it retains the same numerical values of its components, under a coordinate transformation in the form of a reflection through the origin $\mathbf{x} \rightarrow -\mathbf{x}$ (see Chapter 2). Note that such a coordinate inversion involves a change from a right-handed to a left-handed frame of reference. (To confirm that $\mathbf{B}$ is a pseudo-vector, consider its definition, $\mathbf{F} = q\mathbf{u} \times \mathbf{B}$. On inverting the coordinates the numerical values of the components of $\mathbf{F}$ and $\mathbf{u}$ reverse sign, so those of $\mathbf{B}$ cannot.)

Now under the inversion $\mathbf{x}' = -\mathbf{x}$, $\mathbf{b}$ and $\mathbf{B}_0$ retain the same component values, i.e. they reverse direction, while $\mathbf{v}$ retains the same physical direction but reverses its component values. It follows from the definition of α, $\overline{\mathbf{v} \times \mathbf{b}} = \alpha \mathbf{B}_0$, that α must change sign under a coordinate transformation of the form $\mathbf{x} \rightarrow -\mathbf{x}$, and this is what we mean by a pseudo-scalar. It is a strange kind of scalar, not at all like, say, temperature whose value cannot possibly depend on the coordinate system used to describe

space. Another example of a pseudo-scalar is helicity, $\mathbf{v} \cdot (\nabla \times \mathbf{v})$, $\nabla \times \mathbf{v}$ being a pseudo-vector.

This may all sound a little abstract, but it turns out to be important. For example, α is a statistical property of the turbulence which creates it. Thus, if α is to be non-zero, the statistical properties of the turbulence must also change sign under a reflection of the coordinates $\mathbf{x}' = -\mathbf{x}$. We say that the turbulence must lack *reflectional symmetry*, otherwise α will be zero.

The next question might be: can we estimate the size of α? We might expect α to depend on only $|\mathbf{v}|$, λ and l, where $|\mathbf{v}|$ is a measure of the eddy velocity and l is the size of the turbulent eddies or blobs. If this is so then, on dimensional grounds, $\alpha/|\mathbf{v}| = f(|\mathbf{v}|l/\lambda)$, $\alpha/|\mathbf{v}|$ and $|\mathbf{v}|l/\lambda$ being the only two dimensionless groups which we can create from these variables. Two important special cases are $R_m \gg 1$ and $R_m \ll 1$. In cases where R_m is large (on the scale of l) we might expect α not to depend on the diffusivity, λ. That is to say, diffusion should not be an important physical process in the α-effect, except at scales much smaller than l where flux tube reconnections occur. In such a case, we might expect $\alpha \sim |\mathbf{v}|$. However, this cannot be true, since α is a pseudo-scalar while $|\mathbf{v}|$, the r.m.s. turbulence velocity, is not. Given that α is independent of λ, and of order $|\mathbf{v}|$, yet reverses sign in a coordinate inversion, the simplest estimate of α one can come up with is

$$\alpha \sim -\lfloor \overline{\mathbf{v} \cdot (\nabla \times \mathbf{v})} \rfloor l/|\mathbf{v}|, \qquad (R_m \gg 1)$$

Such estimates are, in fact, commonly used. Note the minus sign. This arises because a positive helicity tends to induce **b**-loops whose associated current density, with which we associate $\sigma\alpha\mathbf{B}_0$, is anti-parallel to $\mathbf{B}_0$ (Figure 6.13). This kind of argument implies (but does not prove) that helicity is a key component of the α-effect at high R_m.

For low R_m turbulence we would expect α to depend on λ as well as on l and $|\mathbf{v}|$. In fact, the induction equation tells us $|\mathbf{b}| \sim (|\mathbf{v}|l/\lambda)|\mathbf{B}_0|$, and so

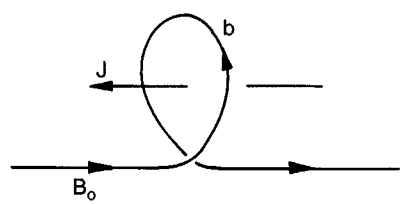

Figure 6.13 (a) Turbulent eddies with positive helicity tend to induce a current density anti-parallel to $\mathbf{B}_0$.

we might expect $\alpha \sim \mathbf{v}^2 l/\lambda$. However, as in the case above, this cannot be true since α is a pseudo-scalar while $\mathbf{v}^2 l/\lambda$ is not. We might anticipate that

$$\alpha \sim -l^2[\overline{\mathbf{v} \cdot \nabla \times \mathbf{v}}]/\lambda, \qquad (R_m \ll 1)$$

which suggests helicity is important whatever the value of R_m, as implied in Figure 6.11.

In fact, when R_m is small, we can evaluate α exactly. We require only that the turbulence be statistically homogeneous. That is, the ensemble average of any turbulent quantity, which we denote $\langle\sim\rangle$, is independent of position. Our starting point is the identity

$$\nabla(\mathbf{a} \cdot \mathbf{b}) - \mathbf{a} \cdot \nabla\mathbf{b} - \mathbf{b} \cdot \nabla\mathbf{a} = \mathbf{a} \times (\nabla \times \mathbf{b}) + \mathbf{b} \times (\nabla \times \mathbf{a})$$

where $\mathbf{a}$ is the vector potential for $\mathbf{v}$: $\mathbf{v} = \nabla \times \mathbf{a}$, $\nabla \cdot \mathbf{a} = 0$. We now substitute for $\nabla \times \mathbf{b}$ using (5.4), the low-R_m form of Ohm's law. This gives

$$\mathbf{v} \times \mathbf{b} = \lambda^{-1}[\mathbf{a} \times (\mathbf{v} \times \mathbf{B}) - \mathbf{a} \times \nabla V] - \nabla(\mathbf{a} \cdot \mathbf{b}) + \nabla \cdot (\sim)$$

Rearranging the term involving the electrostatic potential, V, yields

$$\mathbf{v} \times \mathbf{b} = \lambda^{-1}[\mathbf{a} \times (\mathbf{v} \times \mathbf{B}) - V(\nabla \times \mathbf{a})] - \nabla(\mathbf{a} \cdot \mathbf{b}) + \nabla \cdot (\sim) + \nabla \times (\lambda^{-1} V\mathbf{a})$$

We now express V in terms of $\mathbf{a}$ and $\mathbf{B}$ by taking the divergence of (5.4):

$$\nabla^2 V = \mathbf{B} \cdot \boldsymbol{\omega} = -\nabla^2(\mathbf{B} \cdot \mathbf{a})$$

Thus, to within an arbitrary harmonic function, we have $V = -\mathbf{B} \cdot \mathbf{a}$, and so our expression for $\mathbf{v} \times \mathbf{b}$ simplifies to

$$\mathbf{v} \times \mathbf{b} = \lambda^{-1}[2(\mathbf{a} \cdot \mathbf{B})\mathbf{v} - (\mathbf{a} \cdot \mathbf{v})\mathbf{B}] - \nabla(\sim) + \nabla \cdot (\sim) + \nabla \times (\sim)$$

The final step is to take ensemble averages, at which point the terms involving *grad, div* and *curl* vanish by virtue of our assumption of homogeneity. The end result is

$$\langle \mathbf{v} \times \mathbf{b} \rangle = -\lambda^{-1}\langle(\mathbf{a} \cdot \mathbf{v})\mathbf{B} - 2(\mathbf{a} \cdot \mathbf{B})\mathbf{v}\rangle$$

In terms of α_{ij} this yields

$$\boxed{\alpha_{ij} = -\lambda^{-1}\langle(\mathbf{a} \cdot \mathbf{v})\delta_{ij} - (\mathrm{a}_j \mathrm{v}_i + \mathrm{a}_i \mathrm{v}_j)\rangle} \qquad (R_m \ll 1)$$

If we define $\alpha = \alpha_{ii}/3$, which is consistent with $\alpha_{ij} = \alpha\delta_{ij}$ in the isotropic situation, then

$$\boxed{\alpha = -\langle \mathbf{a} \cdot \mathbf{v} \rangle / (3\lambda)}$$

Compare this with our previous estimate,

$$\alpha \sim -l^2 \left[\overline{\mathbf{v} \cdot \nabla \times \mathbf{v}} \right] / \lambda, \qquad (R_m \ll 1)$$

It seems that the helicity-like pseudo-scalar, $\mathbf{a} \cdot \mathbf{v}$, plays a key role in the low-R_m α-effect.

In summary then, helical turbulence can give rise to an α-effect, and when combined with differential rotation we have the possibility of a self-sustaining dynamo. Actually, integration of the induction equation, incorporating differential rotation and the α-effect, does indeed lead to a self-sustaining dynamo for sufficiently high *dynamo number*, $(\alpha l/\lambda)$ $(u_\theta l/\lambda)$. Typically, in such integrations, α is chosen to be skew-symmetric about the equator, reflecting the supposed structure of the core turbulence. These integrations often yield oscillatory dynamos when the solid inner core is ignored, and non-oscillatory dynamos when the electrical inertia of the inner core is included.

One candidate, then, for a geo-dynamo is the generation of an azimuthal field through differential rotation in the liquid core (the omega effect), supplemented by random, small-scale helical disturbances which convert the azimuthal field back into a poloidal one (the alpha effect). It has to be said, however, that this is a highly simplified picture. For example, we have not addressed the issue of why the turbulence should be dynamically pre-disposed to create an α-effect. Nor have we identified the source of this turbulence.

Example 1: The α-effect induced by helical waves

Suppose that, in Cartesian coordinates, a small amplitude, helical wave of the form

$$\mathbf{v}(\mathbf{x}, t) = Re[\mathbf{v}_0 \exp(i(\mathbf{k} \cdot \mathbf{x} - \omega t))]$$

where $\mathbf{v}_0 = v_0(-i, 1, 0)$ and $\mathbf{k} = (0, 0, k)$, travels through a uniform magnetic field $\mathbf{B}_0$. Confirm that $\mathbf{v}$ is a Beltrami field, in the sense that $\nabla \times \mathbf{v} = k\mathbf{v}$, and that the helicity density is $\mathbf{v} \cdot (\nabla \times \mathbf{v}) = k\mathbf{v}_0^2$. Now use the linearised induction equation to show that the induced magnetic field, $\mathbf{b}$, is

$$\mathbf{b} = \frac{\mathbf{B}_0 \cdot \mathbf{k}}{\omega^2 + \lambda^2 k^4} \left(\lambda k^2 \mathbf{v}^* - \omega \mathbf{v} \right)$$

where $\mathbf{v}^* = \mathrm{Re}[i\mathbf{v}_0 \exp(i(\mathbf{k}\cdot\mathbf{x} - \omega t))]$. Hence show that

$$\overline{\mathbf{v}\times\mathbf{b}} = -\frac{\lambda \mathbf{v}_0^2 k^2 (\mathbf{B}_0\cdot\mathbf{k})}{\omega^2 + \lambda^2 k^4}(0,0,1)$$

In the low R_m limit, $\lambda k^2 \gg \omega$, confirm that α_{ij} is given by

$$\alpha_{ij} = \alpha\delta_{iz}\delta_{jz}, \qquad \alpha = -\mathbf{v}_0^2/(\lambda k) = -\mathbf{v}\cdot(\nabla\times\mathbf{v})/(\lambda k^2)$$

Example 2: A dynamo wave

A two-dimensional analogue of the α-Ω equations can be constructed as follows. Suppose that $\mathbf{B}$ depends only on y and t, $B_y = 0$, and that $\mathbf{u} = \Omega x\hat{\mathbf{e}}_z$. Then

$$\frac{\partial B_z}{\partial t} = \lambda\frac{\partial^2 B_z}{\partial y^2} + \Omega B_x, \qquad \frac{\partial B_x}{\partial t} = \lambda\frac{\partial^2 B_x}{\partial y^2}$$

We might equate B_z to $\mathbf{B}_\theta$ and B_x to $\mathbf{B}_p$. Suppose now that we introduce the α-effect into the equation for B_x, while neglecting it in the B_z equation. Our governing equations now become

$$\frac{\partial B_z}{\partial t} = \lambda\frac{\partial^2 B_z}{\partial y^2} + \Omega B_x, \qquad \frac{\partial B_x}{\partial t} = \lambda\frac{\partial^2 B_x}{\partial y^2} + \alpha\frac{\partial B_z}{\partial y}$$

Show that these equations support solutions of the form $\mathbf{B} \sim \exp(iky + \omega t)$ and that these represent growing, oscillatory waves provided that a suitably defined dynamo number exceeds some threshold. What is the critical value of the dynamo number? (This is known as a dynamo wave.)

Example 3. The dependence of the α-effect on magnetic helicity

Show that, for statistically homogeneous turbulence at low R_m, $\alpha = -\lambda\overline{(\mathbf{b}\cdot\nabla\times\mathbf{b})}/\mathbf{B}_0^2$. Hint, first show that, to within a divergence, which integrates to zero, $\mathbf{b}\cdot\nabla\times\mathbf{b} = \lambda^{-1}\mathbf{b}\cdot(\mathbf{v}\times\mathbf{B}_0)$.

Example 4. Another anti-dynamo theorem

Starting with the induction equation, show that

$$\frac{D}{Dt}(\mathbf{x}\cdot\mathbf{B}) = \mathbf{B}\cdot\nabla(\mathbf{x}\cdot\mathbf{u}) + \lambda\nabla^2(\mathbf{x}\cdot\mathbf{B})$$

Now show that $(\mathbf{x}\cdot\mathbf{u})$ must be non-zero for sustained dynamo action.

6.2.5 Some elementary dynamical considerations

6.2.5.1 Preamble

So far we have restricted ourselves to kinematic aspects of dynamo theory. Of course, this is the simpler part of the problem, in the sense that we give ourselves great latitude in the choice of **u**. That is, we are free to prescribe the velocity field without any concern as to how such a motion might be sustained. Thus, in kinematic dynamo theory we ask only: 'can we find a velocity field, *any* velocity field, which will maintain **B** in the face of Joule dissipation?' It would seem that the answer to this question is 'yes', but this is a long way from providing a coherent explanation for the maintenance of the Earth's magnetic field. We must also determine which of these velocity fields is likely to arise naturally in the interior of the planets, or indeed in the sun. In short, to provide a plausible explanation for the observed planetary magnetic fields, and in particular that of the Earth, we must (somehow) obtain a self-consistent solution of both the induction equation and the momentum equation. This is a tall order and, despite great advances, dynamo enthusiasts are not yet there. Analytical theories tend to be complex and based on rather tentative foundations, while the numerical simulations cannot yet span the wide range of length and time scales inherent in a typical planetary dynamo.

The complexity of the analytical theories lies in stark contrast with the apparently ubiquitous nature of dynamo action. Consider the list of planets, and their magnetic fields, in Table 6.1.

The Earth, Jupiter, Saturn, Uranus and Neptune all have strong dipole moments, Mercury has a rather modest dipole moment, while Venus and Mars exhibit extremely weak (possibly zero) magnetic fields (Venus is probably non-magnetic). It is thought that many of these planetary magnetic fields are self-sustaining dynamos. Yet the constitution, size and rotation rate of the planets vary considerably. The magnetic planets have rotation periods, T, ranging from 0.4 to 59 days, radii which span the range 2400 to 71 000 km and dipole moments from 10^{19} to 10^{27} amps/m^2.

The magnitudes of the planetary fields also vary considerably. It is possible to estimate the mean magnetic field in the planets using the relationship

$$\int_{V_R} \mathbf{B} dV = (2\mu/3)\mathbf{m}$$

Table 6.1. *Properties of the planets*

Planet	Material & density of core (10^3 kg/m^3)	Rotation period, T (days)	Equatorial radius (10^3 km)	Core radius (10^3 km)	Dipole moment (10^{22} amp m^2)	Mean B_z in core (Gauss)	Mean B_z in planet (Gauss)	$(B_x)_p T$ (Gauss-days)
Mercury	Iron, 7.6	59	2.44	1.84	0.005	0.016	0.007	0.60
Venus	Iron, 10.6	243	6.05	3.15	0?	0?	0?	0?
Earth	Iron, 10.6	1	6.38	3.49	7.9	3.7	0.60	0.60
Mars	Iron, 7.5	1.03	3.40	1.50	< 0.002	?	?	?
Jupiter	Liquid hydrogen, 1.3	0.41	71.4	55	150 000	18	8.2	3.4
Saturn	Liquid hydrogen, 0.7	0.43	60.3	28	4700	4.3	0.42	0.18
Uranus	?	0.72	25.6	?	390	?	0.46	0.34
Neptune	?	0.66	24.8	?	200	?	0.26	0.16

where $\mathbf{m}$ is the dipole moment and V_R is any spherical volume which encloses the core currents. This suggests that the mean axial field in the core is of order

$$\bar{B}_z \sim \mu|\mathbf{m}|/(2\pi R_c^3)$$

Estimates of B_z are given in Table 6.1, based on both the core and equatorial radii. Evidently, the mean axial field in the magnetic planets varies from $\sim 10^{-2}$ to ~ 10 Gauss. We might note in passing that, by and large, those planets with the highest rotation rates exhibit the largest magnetic fields, as indicated by the final column in Table 6.1. The main point, though, is that the magnetic planets are all rather different. If it is true that planetary dynamos are so common, yet manifest themselves in such varied circumstances, then one might have hoped that an explanation of dynamo action would be both simple and robust. Not a bit of it! Dynamo theories are complex and, as yet, incomplete.

We shall now outline some of the more elementary dynamical issues. Of course, we have one eye on whether or not the α–$\boldsymbol{\Omega}$ dynamo survives scrutiny. In particular, the α–$\boldsymbol{\Omega}$ model relies on significant and sustained differential rotation in the core, and requires a separation of scales in the core motion, with a significant amount of kinetic energy at the small scales. We shall see that, while (weak) differential rotation probably does exist, there is little support for a formal separation of scales.

6.2.5.2 Typical time scales in the core

Let us suppose that we have both large-scale motion, for which $L \sim 2000$ km, and small-scale eddies, of size no greater than, say, 2 km. (It may turn out that the small scale-scale motion is both weak and transitory – too weak to contribute to an α-effect. However, we should at least make some provision for such a motion since it is a key ingredient of the α–$\boldsymbol{\Omega}$ model.) A common estimate of $|\mathbf{u}|$, based on variations of the surface magnetic field, is 0.2 mm/s. Thus the large-scale motion, which might include differential rotation, buoyant up-wellings and magnetostrophic waves, has a large magnetic Reynolds number, $R_m = uL/\lambda \sim 200$. The small-scale motion, which might be associated with turbulence generated in shear layers, or perhaps small, buoyant plumes, has a relatively low value of R_m, say $R_m = vl/\lambda \sim 0.2$. (We shall take the large-scale

[3] Some caution must be exercised when making the low-R_m approximation. For example, we have seen that Alfvén waves occur when R_m based on the Alfvén speed, v_a, exceeds $\sim \pi$. Typically this wave speed is somewhat greater than v, and so we can have high-R_m phenomena (waves), even though vl/λ is small.

motion, u, and the small-scale velocity, v, to be of similar magnitudes.) A natural starting point, therefore, is to characterise the large-scale phenomena, such as the Ω-effect, as high-R_m, while the small-scale dynamics might be treated as low-R_m.[3]

Let us also allow for differential rotation in the core, as required by the α–Ω model. The probable origin of this relative rotation is discussed more fully in Section 6.2.5.3. We merely note here that differential rotation is observed in certain numerical simulations and in some seismic studies (although the interpretation of the seismic data is not always clear cut). Both the numerical and experimental evidence suggests that the inner core has a rotation rate which is around one degree per year faster than that of the mantle.[4] This differential rotation is thought to be maintained by the secular cooling of the earth, and resisted by viscous coupling of the core and the mantle. That is, cooling causes the solid inner core to grow by solidification, precipitating the release of latent heat and solute-rich, buoyant fluid at the inner-core boundary. The resulting thermal and compositional buoyancy drives a large-scale motion, convecting angular momentum across the core in a systematic manner. The end result is a slight difference in rotation between the inner core and the mantle.

The most important consequence of differential rotation is the inevitable shearing of the dipole field, which sweeps out an azimuthal field of magnitude $B_\theta \sim (u_\theta L/\lambda)B_z$. Thus the dominant field in this picture is azimuthal. Given that $B_z \sim 4$ Gauss in the core, and that $(u_\theta L/\lambda) \sim 200$ we might anticipate that $B_\theta \sim 800$ Gauss in regions of intense differential rotation, i.e. near the inner core. Of course, B_θ does not penetrate beyond the core-mantle boundary, so we have no way of verifying this. We must fall back on the (imperfect) numerical simulations. These suggest that 800 Gauss is an overestimate, and that ~ 50 Gauss is more realistic near the inner core, where the differential rotation is strongest, while B_θ is somewhat weaker in the rest of the core, say, 20 Gauss.

Thus our picture of the large-scale field is one dominated by B_θ. We might further suppose that non-axisymmetric, large-scale convection exists which advects the azimuthal field, forcing low-wavenumber oscillations (magnetostrophic waves) in the large–scale field. This is also a high-R_m process, operating on the scale L. Thus, in this picture, the large-scale field is weakly non-axisymmetric and predominantly azimuthal, as shown in Figure 6.13(b). If we arbitrarily take the internal dipole field to point

[4] A differential rotation of 1 degree per year translates to a velocity of 0.6 mm/s, which is consistent with the estimate of u above.

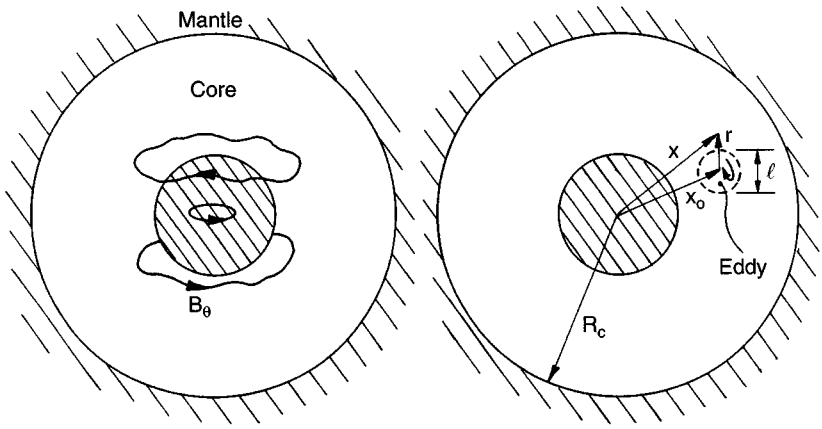

Figure 6.13 (b) (1) The dominant large-scale field is assumed to be azimuthal, produced by differential rotation. It contains a non-axisymmetric, low-wavenumber oscillation caused by large-scale convective motions. Both the differential rotation and the large-scale convective motions occur at large R_m. (2) A small eddy of size l teases out a small-scale field, **b**. This occurs at low R_m.

north (at present it points south), then enhanced rotation of the core relative to the mantle implies that B_θ is positive in the southern hemisphere and negative in the north, as shown.

In addition to these large-scale structures we shall suppose that we have a range of small-scale eddies of size $\sim l$. For lack of a better phrase we might call these small eddies 'turbulence'. In the α–$\boldsymbol{\Omega}$ model, the role of these small eddies is to tease out a small-scale field, **b**, from B_θ, thus regenerating the dipole field $\mathbf{B}_p$.

Let us now try to estimate the characteristic times associated with the large and small-scale structures. To focus thoughts we shall (somewhat arbitrarily) take: $u \sim v \sim 0.2\,\text{mm/s}$, $L \sim 2000\,\text{km}$, $l \sim 2\,\text{km}$, $\lambda \sim 2\,\text{m}^2/\text{s}$, $\nu \sim 3 \times 10^{-6}\,\text{m}^2/\text{s}$, $\rho \sim 10^4\,\text{kg/m}^3$, $|\mathbf{B}_p| \sim 4$ Gauss, $B_\theta \sim 50$ Gauss (near the inner core), and $B_\theta \sim 20$ Gauss (elsewhere).

Table 6.2. *Approximate time scales for large-scale phenomena in the core*

Decay/diffusion time for **B**, $t_d \sim R_c^2/\lambda\pi^2$	10^4 years
Period of magnetostrophic waves ,T	10^4 years
Convective time scale, L/u	300 years

Table 6.3. *Approximate time scales for small-scale phenomena in the core*

Eddy turn-over time, l/v	100 days
Time required to form a Taylor column, $t_\Omega = L/(2\Omega l)$	80 days
Damping time for an Alfvén wave, $l^2/(2\lambda\pi^2)$	1 day
Low-R_m magnetic damping time away from the inner core, $4\tau = 4(\sigma B^2/\rho)^{-1}$	8 hours
Low-R_m magnetic damping time near the inner core, $4\tau = 4(\sigma B^2/\rho)^{-1}$	1 hour

Consider first the large-scale phenomena. There are at least three time scales of interest here: (i) the convective time scale L/u; (ii) the period, T, of the magnetostrophic waves which propagate along the B_θ-lines (see section 6.1.2); (iii) the large-scale diffusion/decay time for **B**, $t_d \sim R_c^2/\lambda\pi^2$. Estimates of these are given in Table 6.2.

The key point is that all of these time scale are relatively large. For example, it takes 10^4 years for a fluctuation in **B** to diffuse through the core!

Now consider the time scales associated with a small-scale eddy. If we can categorise its behaviour as low-R_m (and it is not clear that this is always valid – see footnote at the beginning of this sub-section), then there are three time scales of interest. These are: (i) the eddy turn-over time, l/v; (ii) the time required for an inertial wave to propagate across the core and thus form a Taylor column, $t_\Omega = L/(2\Omega)l$ (iii) the low-R_m damping time, $4\tau = 4(\sigma B^2/\rho)^{-1}$. (Here Ω is the angular velocity of the Earth.) For cases where $v_a l/\alpha >\sim \pi$, v_a being the Alfvén speed, we must add a fourth time scale – that of the damping time for Alfvén waves, $l^2(2\lambda\pi^2)$. (For $v_a l/\alpha <\sim \pi$ these waves do not exist.) Thus the key time scales are as given in Table 6.3.

Of course, we do not really know what B_θ or u are in the core, and so these estimates must be regarded with extreme caution. Nevertheless, if these are at all indicative of the real time scales then they give considerable food for thought. For example, the small-scale processes seem to be extremely rapid by comparison with the large-scale phenomena. Thus, if a buoyant plume left the inner core on the day that Newton first picked up his pen to write *Principia*, it would only just be arriving at the mantle now! Yet small-scale inertial waves can traverse the core in a month or so, while small, neutrally buoyant eddies located near the inner core are

annihilated in a matter of hours! This separation of time scales suggests that we might picture the small-scale eddies as evolving in a pseudo-static, large-scale environment.

We might also note that the turn-over time of a small eddy is large by comparison with the magnetic damping time. This implies that the inertial forces associated with a small eddy are negligible by comparison with the Lorentz forces. The immediate implication is that there is little small-scale turbulence (in the conventional meaning of the word) since the non-linear inertial forces, which are responsible for the turbulent cascade, are absent.

It seems probable, therefore, that the dominant forces acting on a small eddy are the Lorentz, Coriolis and buoyancy forces. The ratio of the Lorentz to Coriolis forces, or equivalently the ratio of the inertial wave period to the magnetic damping time, is represented by the *Elsasser number*, $A = \sigma B^2/(2\rho\Omega)$. If $B_\theta \sim 50$ Gauss, as it might be near the inner core, then $A \sim 7$, and if $B_\theta \sim 20$ Gauss, then $A \sim 1$. Thus, if our estimates of B_θ are reasonable, it would seem that the Coriolis and Lorentz forces are of similar magnitudes in the core. In regions where the Coriolis force wins out we might anticipate quasi-two-dimensional structures, two-dimensionality being enforced by the rapid propagation of inertial waves across the core (see Figure 6.4). In regions where the Lorentz force is dominant, on the other hand, we might expect heavily damped Alfvén waves (if $v_a l/\alpha >\sim \pi$) or else non-oscillatory, low-R_m damping of the type discussed in Section 5.2 (when $v_a l/\alpha <\sim \pi$). In either case, eddies are smeared out along the B_θ-lines while undergoing intense dissipation. The low-R_m damping of eddies is discussed at length in Chapter 9. However, for the present purposes it is sufficient to note that, in the absence of buoyancy, the kinetic energy of an eddy declines as $(t/\tau)^{-1/2}$, where $\tau = (\sigma B^2/\rho)^{-1}$.

The broad picture which emerges, then, is that there is a wide range of time scales. Small eddies are either damped by B_θ or else extruded into Taylor columns by inertial wave propagation. Both processes take place in a matter of days. The convective transport of momentum or magnetic flux is much slower, taking several hundred years to traverse the core. Finally, the large-scale magnetic phenomena (diffusion, magnetostrophic waves) occur on vast time scales, of the order of 10^4 years. There are two important implications of this. First, full numerical simulations are difficult to realise because a wide range of scales have to be resolved. Second, the efficiency with which the particularly small eddies are damped (by Joule dissipation) raises some doubt as to the likely-hood of an energetic small-

scale motion, as required by the α–$\boldsymbol{\Omega}$ model. We shall return to this second issue in Section 6.2.5.4. First, however, we consider the large scales.

6.2.5.3 The large-scale dynamics

The discussion above raises at least two questions relating to the large-scale motion. First, why should compositional or thermal buoyancy give rise to differential rotation? Second, if the inertial forces are so small, is it possible to achieve a quasi-static balance between Coriolis, Lorentz and buoyancy forces? This second question leads to something known as *Taylor's constraint*. Let us start, however, with the issue of differential rotation.

Perhaps the simplest way to picture how differential rotation might arise is to consider an axisymmetric motion, consisting of an up-welling of buoyant fluid rising vertically upward from the inner core. Let us suppose that inertial and viscous forces are negligible (we shall justify this shortly). Moreover, for simplicity, we shall ignore the Lorentz force. (This is completely unjustified. However, the Lorentz force, while modifying the motion, plays no part in the mechanism we are about to describe.) In a frame of reference rotating with the Earth, we have

$$2\mathbf{u} \times \boldsymbol{\Omega} - \nabla(p/\rho) + (\delta\rho/\rho)\mathbf{g} = \text{inertial forces} \approx 0$$

where $\delta\rho$ is the perturbation in density which drives the convection, $\boldsymbol{\Omega}$ is the angular velocity of the Earth, and $\mathbf{g}$ is the gravitational vector which points inwards. Taking the curl of this force balance yields

$$2\Omega \frac{\partial u_\theta}{\partial z} = -\frac{\partial}{\partial z}\left(\frac{\delta\rho}{\rho}\right)g_r + \frac{\partial}{\partial r}\left(\frac{\delta\rho}{\rho}\right)g_z$$

(Remember that we using cylindrical polar coordinates.) Now it is observed in some numerical simulations that the regions above and below the inner core tend to consist of relatively light, buoyant fluid and that, consequently, $\delta\rho$ rises as r increases. The implication is that $\partial u_\theta/\partial z$ is negative in the north and positive in the south. The fluid near the inner core then rotates faster than the mantle, and magnetic coupling, via the dipole field, causes the inner core to rotate at a speed close to that of the surrounding fluid. (The inner core has a relatively small moment of inertia and so reacts almost passively to the magnetic forces which couple it to the surrounding fluid.) The net effect, therefore, is a difference in rotation between the inner core and the mantle.

Thus it appears that there are plausible grounds for believing in differential rotation. Indeed, recent seismic evidence has tended to confirm that the inner core has a prograde rotation relative to the mantle of between 0.3 and 3 degrees per year (although the interpretation of this evidence has been disputed). The energy which maintains this differential rotation (in the face of viscous coupling) comes from the slow growth of the inner core. That is to say, as relatively pure iron solidifies on the inner core, latent heat and solute-rich buoyant fluid are released. The resulting thermal and compositional convection drives the differential rotation, compositional convection probably being the more important of the two.

Let us now turn to the broader issue of how, in the absence of inertial and viscous forces, the Lorentz, buoyancy and Coriolis forces all balance. We have already seen that $\mathbf{u} \cdot \nabla \mathbf{u}$ is negligible by comparison with the Lorentz and Coriolis forces. For example, the ratio of the inertial to Coriolis terms is, $u/2\Omega L \sim 10^{-6}$. The viscous stresses are also small (outside the boundary layers) since $\mathrm{Re} = uL/\nu \sim 10^8$, while the so-called *Ekman number*, $E = \nu/(2\Omega L^2)$, is of the order of 10^{-14}. (The Ekman number represents the ratio of viscous to Coriolis forces.) It appears, therefore, that the viscous and inertial forces are negligible outside the boundary layers.

Now, we have already seen (in Section 3.7) that differential rotation between a fluid and an adjacent solid surface sets up an Ekman boundary layer of thickness $(\nu/\Omega)^{1/2}$ provided, of course, that the flow is laminar. Such layers might be expected to form on the inner core and on the mantle. Indeed, it is the viscous coupling between the core and the mantle which moderates the differential rotation. However, the Ekman layers in the core cannot be laminar since the estimate $\delta \sim (\nu\Omega)^{1/2}$ leads to the bizarre conclusion that $\delta \sim 20$ cms, yet all other relevant length scales are measured in kilometres. For example, the surface of the inner core is thought to consist of a 'forest' of dendritic crystals, about 1 km deep. It seems probable, therefore, that the Ekman thickness is controlled by surface roughness and by turbulence. Some authors allow for this by replacing ν by an eddy viscosity, ν_t. It should be noted, however, that an effective Ekman number based on ν_t is not expected to exceed $\sim 10^{-9}$, so that turbulent stresses are negligible outside the Ekman layers.

The neglect of inertial and viscous forces has profound implications for the large-scale motion. Consider a control-volume, V, confined to the core and bounded by the cylindrical surface $r = r_0$ and by the mantle. In the absence of the non-linear inertial term we have

$$\frac{\partial \mathbf{u}}{\partial t} = 2\mathbf{u} \times \boldsymbol{\Omega} - \nabla(p/\rho) + (\delta\rho/\rho)\mathbf{g} + \rho^{-1}\mathbf{J} \times \mathbf{B} + \nu\nabla^2\mathbf{u}$$

which yields the angular momentum equation

$$\frac{\partial(\mathbf{x} \times \mathbf{u})}{\partial t} = 2\mathbf{x} \times (\mathbf{u} \times \boldsymbol{\Omega}) + \nabla \times (p\mathbf{x}/\rho) + \rho^{-1}\mathbf{x} \times (\mathbf{J} \times \mathbf{B}) + \nu\mathbf{x} \times (\nabla^2\mathbf{u})$$

If we now integrate the z-component of this equation over the volume V, then the pressure torque integrates to zero while the Coriolis term, which can be expanded as a vector triple product, also vanishes, i.e.

$$\int_V 2(\mathbf{x} \times (\mathbf{u} \times \boldsymbol{\Omega}))_z dV = 2\Omega \int_V (zu_z - \mathbf{x} \cdot \mathbf{u}) dV = \Omega \int_V \nabla \cdot \left((z^2 - \mathbf{x}^2)\mathbf{u}\right) dV$$

$$= -r_0^2 \Omega \oint \mathbf{u} \cdot d\mathbf{S} = 0$$

Thus we are left with,

$$\frac{d}{dt}\int (\mathbf{x} \times \mathbf{u})_z dV = \rho^{-1} \int (\mathbf{x} \times (\mathbf{J} \times \mathbf{B}))_z dV + (\text{viscous torque on mantle})$$

It is conventional to reformulate this so that it applies to a thin annulus at $r = r_0$. The end result is

$$\frac{d}{dt}\int u_\theta dA = \rho^{-1} \int (\mathbf{J} \times \mathbf{B})_\theta dA + (\text{viscous torque})$$

Here dA is an element of the surface $r = r_0$. Now the viscous contribution to this equation is of order $\nu_t(u_\theta/\delta)$. Thus the ratio of the Lorentz to the viscous torques is $\sim AE_t^{-1/2}$, where A is the Elsasser number and E_t is the effective Ekman number based on a turbulent eddy viscosity. Since $A \sim 1$, we have

$$\frac{d}{dt}\int u_\theta dA = \rho^{-1} \int (\mathbf{J} \times \mathbf{B})_\theta dA + 0(E_t^{1/2})$$

Thus, in the steady state, the Lorentz torque must satisfy

$$T(r_0) = \rho^{-1} \int (\mathbf{J} \times \mathbf{B})_\theta dA = 0(E_t^{1/2}) \approx 0$$

This is known as Taylor's constraint. In short, the annulus cannot support a sizeable Lorentz torque since there are no significant forces (inertial or viscous) to balance such a torque. If, at some initial instant, this

constraint is broken, torsional oscillations are thought to ensue between adjacent annuli, these annuli being coupled by the B_r field. Damping of the oscillations then causes the flow to evolve towards an equilibrium state, called a *Taylor state*.

There are several ways in which the Taylor constraint can be satisfied. It turns out that one of these is to ensure that B_r is suitably small in the core. That is, in terms of Maxwell stresses,

$$\int (r_0 B_r B_\theta / \mu) dA = \int (\mathbf{x} \times (\mathbf{J} \times \mathbf{B}))_z dV$$

(There is no contribution to the integral on the left from the core–mantle boundary since B_θ is zero there.) Thus the Taylor constraint is satisfied if B_r is suitably small in the core, i.e. the poloidal field lines are almost axial in the core. (This idea has led to a clutch of models known collectively as *Model z*.)

The limitations imposed by the Taylor constraint are quite profound, and in fact this has dominated much of the recent literature on the geodynamo. It might be noted, however, that the net toque arising from a closed system of currents interacting with its 'self field' is necessarily zero. Thus, in a global sense, the Taylor constraint is always satisfied (if we ignore any currents in the mantle).

6.2.5.4 The small-scale dynamics

We now turn to the small-scale motion in the core. The main issue here is whether or not the small-scale structures, which are so important in the α–$\boldsymbol{\Omega}$ model, can survive the relatively intense Joule dissipation and so contribute, via the α-effect, to the global dipole field.

We have already seen that, at scales of $1 \sim 2$ km the magnetic Reynolds number is less than unity. This suggests that many of the small-scale eddies are subject to low-R_m damping of the type discussed in Chapters 5 and 9. Such eddies will decay rather rapidly unless they are maintained by some external force, such as buoyancy. For example, it is shown in Chapter 9 that, in the absence of buoyancy, the kinetic energy of a low-R_m eddy evolving in a uniform magnetic field declines as $(t/\tau)^{-1/2}$, where τ is the Joule damping time $(\sigma B^2/\rho)^{-1}$. Moreover, in the absence of Coriolis and buoyancy forces, an eddy whose axis of rotation is parallel to $\mathbf{B}$ conserves its angular momentum, while one whose axis is normal to $\mathbf{B}$ loses its angular momentum at a rate $\mathbf{H}_\perp \sim \mathbf{H}_{\perp 0} \exp(-t/4\tau)$. In the former case, the eddy evolves into an elongated, cigar-like structure

whose main axis is aligned with **B** (see Figure 5.3b), while in the latter case the eddy loses its angular momentum by disintegrating into a network of plate-like structures whose planes are orientated parallel to **B** (see Figure 9.12). In both cases the eddy elongates in the direction of **B** at a rate $(t/\tau)^{1/2}$.

The situation is more complicated when both Lorentz and Corlios forces act on a small eddy. Some hint as to how Coriolis forces might influence the decay of a low-R_m eddy is furnished by the following simple model problem.

Suppose we have a small, localised, low-R_m eddy sitting in a locally uniform field $\mathbf{B} = B\mathbf{e}_x$. It has finite angular momentum and is subject to Coriolis and Lorentz forces (with $\boldsymbol{\Omega} = \Omega\mathbf{e}_z$). However, viscous, gravitational and non-linear inertial forces are neglected. Thus the momentum equation simplifies to

$$\frac{\partial \mathbf{u}}{\partial t} = 2\mathbf{u} \times \boldsymbol{\Omega} - \nabla(p/\rho) + \rho^{-1}\mathbf{J} \times \mathbf{B}$$

from which we have

$$\frac{\partial(\mathbf{x} \times \mathbf{u})}{\partial t} = 2\mathbf{x} \times (\mathbf{u} \times \boldsymbol{\Omega}) + \nabla \times (p\mathbf{x}/\rho) + \rho^{-1}\mathbf{x} \times (\mathbf{J} \times \mathbf{B})$$

Using (5.22) to rearrange the Lorentz force, and a variant of (5.22) to recast the Coriolis force, we find that

$$\frac{\partial(\mathbf{x} \times \mathbf{u})}{\partial t} = (\mathbf{x} \times \mathbf{u}) \times \boldsymbol{\Omega} + \nabla \times (p\mathbf{x}/\rho) + (2\rho)^{-1}(\mathbf{x} \times \mathbf{J}) \times \mathbf{B} + \nabla \cdot (\sim \mathbf{u}) + \nabla \cdot (\sim \mathbf{J})$$

Next we integrate over a large spherical volume, and insist that $\mathbf{u}.d\mathbf{S}$ and $\mathbf{J}.d\mathbf{S}$ are zero on some remote boundary. This yields

$$\frac{d\mathbf{H}}{dt} = \mathbf{H} \times \boldsymbol{\Omega} + \mathbf{m} \times \mathbf{B}/\rho$$

where **m** is the dipole moment induced by the interaction of the eddy with **B**. Finally, following the arguments leading up to (5.23), and on the assumption (which is not always valid in the core) that the low-R_m form of Ohm's law applies, we recast **m** in terms of **H** to give

$$\frac{d\mathbf{H}}{dt} = \mathbf{H} \times \boldsymbol{\Omega} - \frac{\mathbf{H}_\perp}{4\tau}, \qquad \mathbf{H}_\perp = (0, H_y, H_z)$$

It appears that H_z declines exponentially. H_y and H_x, on the other hand, decay in a sinusoidal fashion if $A = \sigma B^2/(2\rho\Omega)$ is less than 4, and decay

exponentially if A exceeds 4. In either case the characteristic decay time is 4τ. Now it is readily confirmed that the Coriolis force does not change the rate of decline of energy and so the energy of the eddy falls as $(t/\tau)^{-1/2}$. An algebraic decline in energy, yet exponential decline in angular momentum, requires that the eddy adopts a spatial structure in which the angular momentum alternates in sign and integrates to zero. So we might anticipate that, whatever the value of A, plate-like structures will emerge, as shown in Figure 9.12. Moreover, when A is small, an eddy presumably undergoes a substantial elongation before being destroyed, the Coriolis force extruding the eddy into a Taylor column. Thus the eddy shown on the left of Figure 9.12 will grow at a rate $(t/\tau)^{1/2}$ parallel to $\mathbf{B}$ and at a rate Ωt parallel to $\boldsymbol{\Omega}$, forming a set of platelets of alternating vorticity which, when added together, have zero net angular momentum.

So what does all this mean in the context of the Earth's core? For $B_\theta \sim$ 20 Gauss we have a dissipation time scale of $4\tau \sim 1$ day. This is very rapid by comparison with the other relevant time scales, and so it is by no means clear that these eddies can be replenished as fast as they are destroyed.

At present, the prevailing view is that, as far as planetary dynamos are concerned, the α-effect is, at best, a pedagogical idealisation.

6.2.6 Competing kinematic theories for the geo-dynamo

Two-scale α–$\boldsymbol{\Omega}$ models have been around for some time now. They represented a significant breakthrough in dynamo theory because they circumvented the fundamental limitations imposed by Cowling's theorem while providing a theoretical framework for constructing possible dynamo mechanisms. Their weaknesses, however, are three-fold. First, the α-effect is essentially a kinematic theory. Why should the turbulence in the earth's core be dynamically predisposed to reconstruct a large-scale dipole field from a random small-scale field? Second, they rely on significant differential rotation in the core, which requires the inner core and the mantle to rotate at different rates. As noted above, there is some indication that this is indeed the case. However, the evidence is not yet conclusive. Third, they presuppose a two-scale structure for $\mathbf{u}$ and $\mathbf{B}$, with significant energy in the small-scale turbulence. There is no real evidence that this is the case and, as we have seen, there are arguments to the contrary.

Many other dynamo mechanisms have been proposed. For example, if we accept the notion of a two-scale approach, but reject the idea of strong differential rotation, then we can still get a dynamo through the α-effect. That is, B_θ can be generated from $|\mathbf{B}_p|$ by an α-effect, which is then converted back into a poloidal field, again by the α-effect. This is called an α^2-dynamo (see example below). Alternatively, we might abandon the two-scale picture altogether and consider large-scale convective motions driven by buoyancy and Coriolis forces. Indeed, there have been many computer simulations of that type. However, despite all of this research, there is still no self-consistent theory which explains the observations.

In any event, it looks like the search for an entirely self-consistent model of the geo-dynamo will continue for some time. Great advances have been made, yet there is still some resonance in Maxwell's comment:

> ... we are not yet fully acquainted with one of the most powerful agents in nature, the scene of whose activity lies in those inner depths of the earth, to the knowledge of which we have so few means of access.
>
> *(1873)*

Example: An α^2-dynamo

Consider the averaged induction equation in cases where $\alpha =$ constant and $\mathbf{u}_0 = 0$:

$$\frac{\partial \mathbf{B}_0}{\partial t} = \alpha \nabla \times \mathbf{B}_0 + \lambda \nabla^2 \mathbf{B}_0$$

Consider solutions of this equation of the form $\mathbf{B}_0 = \hat{\mathbf{B}}(\mathbf{x})e^{pt}$where $\hat{\mathbf{B}}(\mathbf{x})$is a so-called 'force-free' field satisfying,

$$\nabla \times \hat{\mathbf{B}} = k\hat{\mathbf{B}}$$

Show that, for suitable initial conditions, this is a solution of the averaged induction equation, and that

$$p = \alpha k - \lambda k^2$$

Deduce the criterion for a self-sustaining α^2-dynamo.

6.3 A Qualitative Discussion of Solar MHD

One lot cogitates on the way of religion, Another ponders on the path of mystical certainty; But I fear one day the cry will go up, 'Oh you fools, neither this nor that is the way!'

Omar Khayyam

The capricious behaviour of sunspots has been the source of speculation since the first observations in ancient China. Considered debate in the West probably dates back to the early 17th century and to the development of the telescope by Galileo. Indeed it was Galileo's *Letters on Solar Spots*, published in Rome in 1613, which precipitated the clash between Galileo and the church. Thus the battle between science and religion began; a skirmish which had still not abated by 1860 when Huxley and Bishop Wilberforce debated Darwin's *Origin of the Species*.

Records of sunspot appearances have been kept more or less continuously since Galileo's time. By 1843 it was realised that the appearance of spots followed an eleven-year cycle (although there was a curious dearth of sunspots during the reign of the Roi Soleil in France!). The reason for the eleven-year cycle remained a mystery for some time, but it was clear by the end of the 19th century that there was an electromagnetic aspect to the problem. As Maxwell noted in 1873, when discussing terrestrial magnetic storms: '*It has been found that there is an epoch of maximum disturbance every eleven years, and that this coincides with the epoch of maximum number of sunspots in*

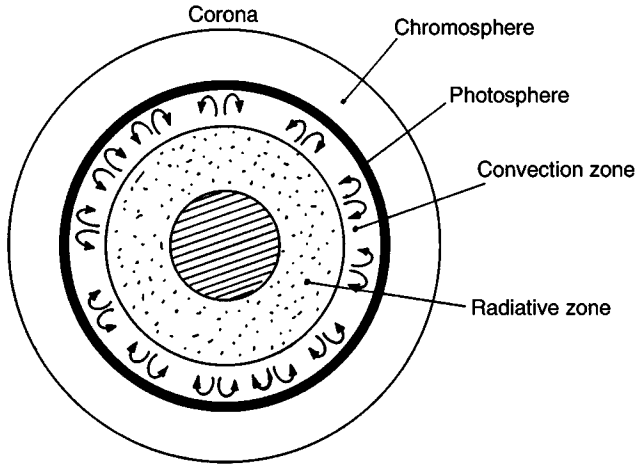

Figure 6.14 (a) The structure of the sun.

the sun.' Maxwell was on the right track, but it was not until the development of MHD that we have begun to understand some of the observations.

The following paragraphs give a brief qualitative introduction to solar MHD. The focus is on sunspots and solar flares. The discussion is purely descriptive, intended only to give a glimpse of some of the more intriguing phenomena involved. To date, many problems concerned with the solar dynamo have not been solved, and models which have been proposed require for their understanding mathematics beyond the level of this book. Some reading suggestions are given at the end of the chapter which fearless readers may consult.

6.3.1 The structure of the sun

The sun's interior is conventionally divided into three zones (Figure 6.14(a). The central core of radius $\sim 2 \times 10^5$ km is the seat of thermonuclear fusion. This is surrounded by the so-called radiative zone which extends up to a radius of $\sim 5 \times 10^5$ km Here heat is transported diffusively by radiation and conditions are hydrodynamically stable. The outer region is called the convection zone. It is approximately 2×10^5 km deep, convectively unstable and so in a state of constant motion. Heat is carried to the surface via convection.

The solar atmosphere is also divided into three regions. The 'surface' of the sun is called the photosphere. This is a thin transparent layer of relatively dense material about 500 km deep. Above this lies the hotter, lighter chromosphere which is around 2500 km deep. The outermost layer is the corona, which has no clear upper boundary and extends in the form of the solar wind out to the planets. There is a dramatic rise in temperature in passing from the chromosphere to the corona.

The existence of the solar convection zone is evident in the granular appearance of the photosphere. In photographs it looks like a gravel path and is reminiscent of multi-cellular Bénard convection. The granules (convection cells), which are continually evolving on a time-scale of minutes, have a typical diameter of $\sim 10^3$ km and are bright at the centre, where hot plasma is rising to the surface, and darker at the cell boundaries where cooler plasma falls. Because they are influenced by surface radiation, these granules are not necessarily representative of the scale of the internal motion deep within the convection zone.

As was mentioned in Chapter 4, the sun does not rotate as a rigid body. The average surface rotation is faster at the equator than it is near the poles but the radiative zone rotates more or less like a rigid body at a rate somewhere between the equatorial and polar surface rates. This differential rotation is crucial to much of solar MHD.

6.3.2 Is there a solar dynamo?

The natural decay time for a magnetic dipole field in the sun is around 10^9 years, which is about the age of the solar system itself. This is not inconsistent with the notion that the field is the relic of some galactic field which was trapped in the solar gas at the time of the sun's formation. However, the sun's magnetism is constantly varying in a manner that cannot be explained by some frozen-in primordial field. Sunspot activity could be attributable to transient, small-scale processes, but the periodic (22-year) variation of the sun's global field suggests that theories based on a frozen-in relic are incorrect. It seems likely, therefore, that the explanation of solar magnetism lies in dynamo action within the convective zone. Note that while dynamo theory is invoked to explain the unexpected persistence of the earth's magnetic field, it is invoked in the solar context to explain the rapid evolution of the sun's field.

The dynamo theories which have been developed in the context of the earth and the sun are, however, very different. In the core of the earth, velocities are measured in fractions of a millimetre per second, and as a result R_m is rather modest, $R_m \leq 100$. In the convective zone of the sun, on the other hand, velocity fluctuations are around 1 km/s, giving $R_m \sim 10^7$. While concerns in geodynamo theory often centre around finding turbulent motions which have an R_m high enough to induce significant field stretching, in the solar dynamo the problem is of the opposite nature. R_m is so high that molecular diffusivity becomes very weak, and so extremely large gradients in the magnetic field must develop in order to allow the flux tube reconnections needed to explain the observed behaviour.

6.3.3 Sunspots and the solar cycle

We have already discussed sunspots in Chapter 4. They are a manifestation of unstable, buoyant flux tubes which float up through the convective zone and erupt into the solar atmosphere (see Figure 4.2). These dark

spots appear in pairs and are the foot-points of the flux tube in the solar surface, where the intense local magnetic field (~ 3000 G) suppresses fluid motion and cools the surface. The spots are typically 10^4 km in diameter (much larger than a granule) and they appear mainly near the equatorial plane. Often they occur in groups (an 'active region') and this gives rise to an increased brightness, called photospheric faculae.

The intensity of sunspot activity fluctuates on a regular 11-year cycle. At the sunspot minimum there may be no sunspots, while at the maximum there are typically around a hundred. After their rapid initial formation, sunspot pairs may survive for some time, disintegrating over a period of days or weeks, the so-called 'following spot' vanishing first. The fragments of the flux tube which caused the spot are then convected around by the photospheric flows accumulating along granular boundaries.

The area of the photosphere covered by sunspots varies during the 11-year cycle. At the minimum point new spots appear first at latitudes of $\sim \pm 30°$. The number of active zones then increases, gathering towards the equator, until finally the sunspot activity dies away. The magnetic field in an active region is predominantly azimuthal (field lines which circle the solar axis), so that the sunspot pairs are aligned (almost) with a line of latitude. They rotate with the surface of the sun, the leading spot being slightly closer to the equator than the following spot. Leading and following spots are observed to have opposite orientations of **B**, as you would expect. However, all pairs in one hemisphere have the same orientation (e.g. **B** directed outward in the leading spot and **B** directed inward in the following spot) and this orientation reverses as we move from one hemisphere to the other. This suggests that the sub-surface azimuthal field is unidirectional in each hemisphere and antisymmetric about the equator: a picture which is consistent with an azimuthal field being swept out from a dipole field by differential rotation (see §4.5.3). Crucially, however, the field orientation in the sunspot pairs reverses from one 11-year cycle to the next, suggesting a periodic variation of the subsurface azimuthal field every 22 years. If this azimuthal field is generated from a dipole field by differential rotation, then this, in turn, suggests a periodicity in the dipole field, or else a periodic reversal in the differential rotation which sweeps out the azimuthal field. If the latter were true there would be no need for a solar dynamo to explain the 22-year cycle. However, observation suggests that it is the first explanation which is correct. The sun's poloidal (dipole) field appears to reverse at the sunspot maximum, strongly sug-

gesting (but not proving) that the solar magnetic field is maintained by dynamo action in the convective zone.

6.3.4 The location of the solar dynamo

It might be thought that the entire convective zone contributes equally to dynamo action, and indeed this was once taken to be the case. However, recently it has been suggested that much of the dynamo action occurs in a relatively thin layer at the interface of the radiative and convection zones. In part, this change in view arose from measurements of rotation in the sun which suggest that differential rotation is concentrated at this interface and so this thin layer is likely to be the location of intense azimuthal field generation.

Mathematical models of the solar dynamo have been proposed based on this idea, including, for example, an α–$\boldsymbol{\Omega}$ model. In this picture, strong azimuthal fields build up at the base of the convective zone due to differential rotation (the $\boldsymbol{\Omega}$-effect). These fields support low-frequency magnetostrophic waves (see Section 6.1) which, when combined with buoyancy-driven motion, regenerate a dipole field from the azimuthal one (the α-effect). However, such a model is, perhaps, a little idealistic, representing a convenient conceptual framework which captures key physical mechanisms, but not really providing a truly predictive model of the solar dynamo. As with the geo-dynamo, much remains to be done.

6.3.5 Solar flares

The solar atmosphere is anything but passive. It is threaded with vast magnetic flux tubes which arch up from the photosphere into the corona and which are constantly evolving, being jostled by the convective motions in the photosphere (see Figure 4.3). Some of these flux tubes are associated with sunspots, others are associated with so-called *prominences*. Prominences extend from the chromosphere up into the corona, appearing as arch-like, tubular structures of length $\sim 10^5$ km and thickness $\sim 10^4$ km. They contain cold, chromospheric gas, perhaps 300 times colder than the surrounding coronal gas. This relatively cold plasma is threaded by a magnetic field of ~ 10 Gauss, which is much weaker than that in a sunspot, but larger than the mean surface field of ~ 1 Gauss. A prominence is itself immersed in, and surrounded by,

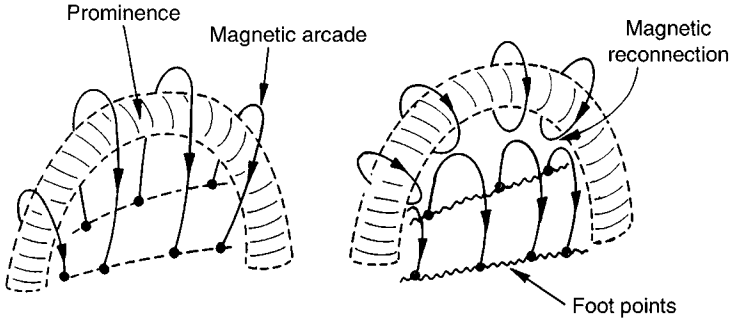

Figure 6.14 (b) A cartoon of a two-ribbon solar flare.

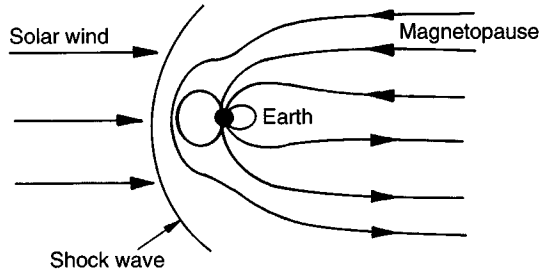

Figure 6.14 (c) The solar wind.

thinner flux tubes which arch up from the photosphere, criss-crossing the prominence. Some flux tubes lie below the prominence, providing a magnetic cushion. Others lie above, pushing down on the prominence. The flux tubes which overlie the prominences are sometimes referred to as a *magnetic arcade*.

Quiescent prominences are stable, long-lived structures which survive for many weeks, while *explosively eruptive prominences* give rise to spectacular releases of mass and energy in relatively short periods of time (hours). The mass which is propelled from the sun in this way is called a coronal mass ejection (CME), and the sudden release of energy is called a *solar flare*.

As yet, there is no self-consistent model for solar flares, although all current theories agree that the power source is stored magnetic energy whose release is triggered by magnetic reconnection. The largest flares

are called *two-ribbon-flares* and they are thought to arise as follows. Consider a prominence which is supported by a magnetic cushion and has a magnetic arcade overlying it. Now suppose that the prominence starts to rise, perhaps because of a build-up of magnetic pressure in the magnetic cushion (which itself might arise from the photospheric jostling of the flux tube roots). The field lines in the overlying arcade, which also have their roots in the photosphere, will become increasingly stretched. Eventually, large gradients in **B** will build up, allowing magnetic reconnections to occur. This, in turn, allows the arcade flux tubes to 'pinch off', releasing magnetic energy, as shown in Figure 6.14(b). When this occurs the downward force associated with the overlying flux tubes is suddenly removed and so the prominence is propelled explosively upward by the magnetic pressure in the underlying magnetic cushion. Some of this energy is also propagated down the arcade field lines to their foot points in the chromosphere and photosphere. The footprints of these field lines then appear as two highly energetic 'ribbons' in the chromosphere – hence the name.

It has to be said, however, that this picture is rather simplistic. Recent measurements suggest that there is not a one-to-one correspondence between coronal mass ejection and solar flares. Often CMEs occur without flares while flares need not be accompanied by a CME. Clearly, the entire process is much more complicated than that suggested above.

Whatever the true explanation of solar flares, it cannot be denied that they are spectacular events. They are vast in scale, extending over $\sim 10^5$ km, and release prodigious amounts of energy, of the order of $\sim 10^{25}$ J. This sudden release of mass and energy enhances the solar wind which, even in quiescent times, spirals radially outward from the sun. At times of vigorous solar activity (at sunspot maximum) the concentration of particles in the solar wind can increase from $\sim 5 \times 10^6\,\mathrm{m}^{-3}$ to $\sim 10^7\,\mathrm{m}^{-3}$, and their velocity rises from around 400 km/s to 900 km/s. The mass released by these solar flares sweeps through the solar system and one or two days after a large flare is observed the earth is buffeted by magnetic storms.

Such storms can cause significant damage, as one Canadian power company discovered to its embarrassment in 1989. Around the 11th of March 1989 a large solar flare burst from the surface of the sun, and as dawn broke on the 13th of March six million Canadians found themselves without power!

6.4 Energy-Based Stability Theorems for Ideal MHD

One of the major successes of high-R_m MHD lies in the area of stability theory. This has its roots, not in liquid-metal MHD, but rather in plasma physics. A question which is often asked in fluid mechanics is: 'Is a given equilibrium or steady motion stable to small disturbances?' That is to say, if a steady flow is disturbed, will it evolve into a radically different form or will it remain close (in some sense) to its initial distribution. The method used most often to answer this question is so-called normal mode analysis. This proceeds by looking for small amplitude disturbances which are of the separable form $\delta\mathbf{u}(\mathbf{x}, t) = \mathbf{u}_0(\mathbf{x})e^{pt}$. When quadratic terms in the small disturbance are neglected the governing equations of motion become linear in $\delta\mathbf{u}$, and this defines an eigenvalue problem for the amplitude of the disturbance, $\mathbf{u}_0(\mathbf{x})$. The eigenvalues of this equation determine p, and the motion is deemed to be unstable if any p can be found which has a real positive part. This works well when the geometry of the base flow is particularly simple, possessing a high degree of symmetry, e.g. one-dimensional flow. However, if there is any significant complexity to the base flow (it is two- or three-dimensional) this procedure rapidly becomes very messy, requiring numerical methods to determine the eigenvalues.

In MHD an alternative method has been developed, which relies on the conservation of energy. This has the advantage that it may be applied to equilibria of arbitrary complexity, but it has two major short-comings. First, it applies only to non-dissipative systems ($\lambda = \nu = 0$), which we might call *ideal MHD*. Second, it usually provides sufficient, but not necessary, conditions for stability. Thus often an equilibrium may be proved stable, but it cannot be shown to be unstable. We shall describe this energy method here. First, however, we shall discuss the motivation for developing special stability methods in MHD.

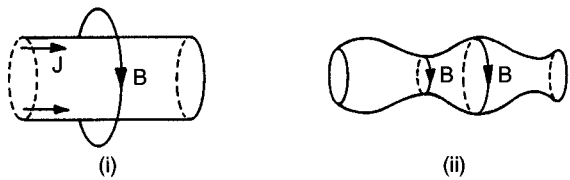

Figure 6.15 The linear pinch. (i) The confinement principle. (ii) Instability of the pinch.

6.4.1 The need for stability theorems in ideal MHD: plasma containment

In the 1950s the quest for controlled thermonuclear fusion began in earnest. This required that the (very) hot plasma be confined away from material surfaces, and since these plasmas are good conductors, magnetic pressure seemed the obvious confinement mechanism. A simple confinement system is shown in Figure 6.15. An axial current is induced in the surface of the plasma, which is in the form of a cylinder, and the resulting azimuthal field creates a radial Lorentz force which is directed inward. (To form a more compact confinement system, the cylinder could be deformed into a torus.) This configuration is known as the linear pinch. Regrettably, it is unstable. Let I be the total current passing along the column. Then the surface field is $B_\theta = \mu I/2\pi R$ where R is the radius of the column. If R locally decreases for some reason, then B_θ rises by an amount $\delta B_\theta = B_\theta \delta R/R$. A 'sausage-mode' instability then develops because there is a rise in magnetic pressure, $\delta p_m = B_\theta \delta B_\theta/\mu$, at precisely those points where the radius reduces.

This sausage-mode instability may be stabilised by trapping a longitudinal magnetic field, B_L, within the plasma. The idea is the following. If λ is very small then this longitudinal field is frozen into the plasma, so if R reduces locally to $R - \delta R$, the longitudinal magnetic field will increase by an amount $\delta B_L = 2B_L \delta R/R$, the total longitudinal flux remaining constant. The magnetic pressure due to B_L therefore increases by $\delta p_m = B_L \delta B_L/\mu = 2B_L^2 \delta R/\mu R$ and this tends to counterbalance the rise in 'pinch pressure' $\delta p_m = B_\theta^2 \delta R/\mu R$. The column is then stable to axisymmetric disturbances provided that $B_L^2 > B_\theta^2/2$.

Unfortunately, this is not the end of the story. The column is unstable to non-axisymmetric disturbances even in the presence of a longitudinal field. This is known as the *kink instability*. Suppose that the column is bent slightly, as shown in Figure 6.16. The field lines are pressed together on the concave side, and spaced out on the other side. Thus the magnetic field, and hence the magnetic pressure, is increased on the concave side of

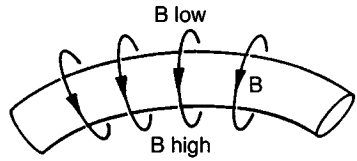

Figure 6.16 The kink instability.

the column and reduced on the convex side. This produces a net sideways force which accentuates the initial disturbance.

In fact, confining plasmas using magnetic fields turns out to be altogether rather tricky. It is not just the linear pinch which is unstable. In the late 1950s, plasma physicists were faced with the problem of deciding which confinement schemes were unstable. Conventional, normal-mode techniques seemed cumbersome and so a new stability theory was developed (primarily at Princeton), first for magnetostatic equilibria, such as that shown in Figure 6.15, and shortly afterwards for any steady solution of the equations of ideal MHD, static or otherwise. This new method is based on the conservation of energy, and in fact it is more in line with our intuitive notions of stability than conventional normal-mode analysis. For example, it predicts that a magnetostatic equilibrium is stable if its magnetic energy is a minimum at equilibrium. Unfortunately, though, the proof of these new stability theorems requires a great deal of vector manipulation. Consequently, the proofs which follow are not for the impatient or the faint-hearted. The end result, though, is rewarding.

6.4.2 The energy method for magnetostatic equilibria

To get an idea of how conservation of energy may be used in a stability analysis, we first consider the simpler problem of the magnetostatic equilibrium of an ideal, incompressible fluid. The fluid and magnetic field are both assumed to be contained in a volume, V, with a solid surface, S, and the equilibrium is governed by

$$\mathbf{J}_0 \times \mathbf{B}_0 = \nabla P_0, \qquad \mathbf{B}_0 \cdot d\mathbf{S} = 0$$

Here the subscript 0 indicates a steady, base configuration whose stability is in question, and $d\mathbf{S}$ is an element of the boundary, S. Now suppose that this equilibrium is slightly disturbed, and that during the initial disturbance the magnetic field is frozen into the fluid. Let $\boldsymbol{\zeta}(\mathbf{x}, t)$ be the displacement of a particle, p, from its equilibrium position $\mathbf{x}$,

$$\boldsymbol{\zeta}(\mathbf{x}, t) = \mathbf{x}_p(t) - \mathbf{x}_p(0), \qquad \mathbf{x}_p(0) = \mathbf{x}$$

Following the initial disturbance some motion will ensue, perhaps in the form of an oscillation, e.g. Alfvén waves, or perhaps something rather more drastic. In any event, $\mathbf{B}$ will be frozen into the fluid and the resulting velocity field, $\mathbf{u}(\mathbf{x},t)$, is related to the instantaneous particle displacement, $\boldsymbol{\zeta}$, by

$$\frac{\partial \boldsymbol{\zeta}}{\partial t} = \mathbf{u}(\mathbf{x} + \boldsymbol{\zeta}, t) = \mathbf{u}(\mathbf{x}) + \boldsymbol{\zeta} \cdot \nabla \mathbf{u} + \ldots \tag{6.15}$$

Let us now evaluate the change in magnetic energy, E_B, which results from the particle displacement, $\boldsymbol{\zeta}$. We first expand E_B in a series

$$E_B(\boldsymbol{\zeta}, t) = \int (\mathbf{B}^2/2\mu) dV = E_{B0} + \delta^1 E_B + \delta^2 E_B + \ldots$$

Here $\delta^1 E_B$ and $\delta^2 E_B$ are the first- and second-order changes in E_B, $\boldsymbol{\zeta}$ being assumed small at all times. We shall see shortly that $\delta^1 E_B = 0$, while the stability of the magnetostatic equilibrium is determined by the sign of $\delta^2 E_B$. Specifically, if $\delta^2 E_B$ is positive, so that E_B is a minimum at equilibrium, the magnetic field is stable. The question, then, is how to evaluate $\delta^1 E_B$ and $\delta^2 E_B$. We now employ a trick. E_B depends only on the instantaneous position of the fluid particles and not their previous histories. That is, E_B is completely determined by the instantaneous spatial distribution of $\mathbf{B}$. There are infinitely many ways in which each particle could get from $\mathbf{x}$ to $\mathbf{x} + \boldsymbol{\zeta}$, but, since E_B does not care about the history of the particles, we shall consider the simplest. Suppose that we apply an imaginary, steady velocity field, $\mathbf{v}(\mathbf{x})$, to the fluid for a short time τ. We choose $\mathbf{v}(\mathbf{x})$ such that it shifts the fluid from its equilibrium configuration to $\mathbf{x} + \boldsymbol{\zeta}$. Since the fluid is incompressible, $\mathbf{v}(\mathbf{x})$ must be solenoidal. We shall call $\mathbf{v}(\mathbf{x})$ a *virtual velocity field* (Figure 6.17). Since $\mathbf{B}$ is frozen into the fluid during the application of $\mathbf{v}$ we have

$$\frac{\partial \mathbf{B}}{\partial t} = \nabla \times (\mathbf{v} \times \mathbf{B}), \qquad 0 < t < \tau \tag{6.16}$$

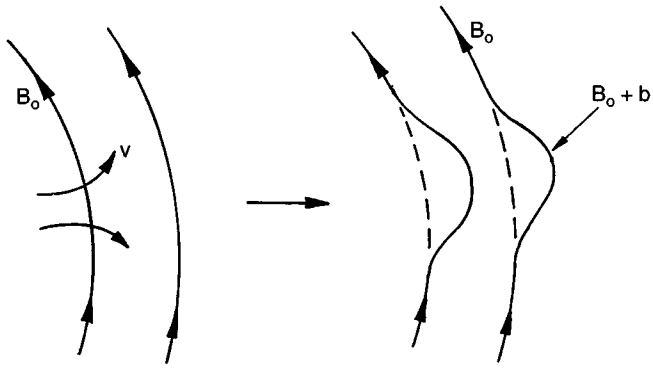

Figure 6.17 Perturbation of $\mathbf{B}_0$ by a virtual velocity field.

It follows that the first- and second-order changes in **B** are

$$\delta^1\mathbf{B} = \nabla \times (\boldsymbol{\eta} \times \mathbf{B}_0) \tag{6.17a}$$

$$\delta^2\mathbf{B} = \frac{1}{2}\nabla \times (\boldsymbol{\eta} \times \delta^1\mathbf{B}) \tag{6.17b}$$

where $\boldsymbol{\eta} = \mathbf{v}\tau$. This new fields satisfies

$$\nabla \cdot \boldsymbol{\eta} = 0, \qquad \boldsymbol{\eta} \cdot d\mathbf{S} = 0 \tag{6.18}$$

We shall call $\boldsymbol{\eta}$ the *virtual displacement field* in order to distinguish it from the Lagrangian particle displacement $\boldsymbol{\zeta}$. Note that $\boldsymbol{\eta}$ and $\boldsymbol{\zeta}$ are not identical. During the application of our imaginary velocity field, $\mathbf{v}(\mathbf{x})$, we have, from (6.15),

$$\frac{\partial \boldsymbol{\zeta}}{\partial t} = \mathbf{v}(\mathbf{x} + \boldsymbol{\zeta}) = \mathbf{v}(\mathbf{x}) + \boldsymbol{\zeta} \cdot \nabla \mathbf{v} + \ldots$$

It follows that

$$\boldsymbol{\zeta} = \boldsymbol{\eta} + \frac{1}{2}\boldsymbol{\eta} \cdot \nabla \boldsymbol{\eta} + \ldots \tag{6.19a}$$

$$\boldsymbol{\eta} = \boldsymbol{\zeta} - \frac{1}{2}\boldsymbol{\zeta} \cdot \nabla \boldsymbol{\zeta} + \ldots \tag{6.19b}$$

Thus, the particle displacement and the virtual displacement are equal only at first order. Let us now use (6.17) to evaluate the changes in E_B which result from the application of $\boldsymbol{\eta}$. The first-order change is

$$\delta^1 E_B = \frac{1}{\mu}\int (\mathbf{B}_0 \cdot \delta^1\mathbf{B}) dV = \frac{1}{\mu}\int \mathbf{B}_0 \cdot \nabla \times [\boldsymbol{\eta} \times \mathbf{B}_0] dV$$

which we might anticipate is zero. To show that this is indeed the case, we note that the integrand may be rewritten as

$$\mathbf{B}_0 \cdot \nabla \times [\boldsymbol{\eta} \times \mathbf{B}_0] = (\boldsymbol{\eta} \times \mathbf{B}_0) \cdot (\nabla \times \mathbf{B}_0) + \nabla \cdot [(\boldsymbol{\eta} \times \mathbf{B}_0) \times \mathbf{B}_0]$$

Rearranging the scalar triple product and expanding the vector triple product yields

$$\mathbf{B}_0 \cdot \nabla \times [\boldsymbol{\eta} \times \mathbf{B}_0] = -\mu(\mathbf{J}_0 \times \mathbf{B}_0) \cdot \boldsymbol{\eta} + \nabla \cdot \left[(\mathbf{B}_0 \cdot \boldsymbol{\eta})\mathbf{B}_0 - \mathbf{B}_0^2 \boldsymbol{\eta}\right]$$

The divergence integrates to zero and so

$$\delta^1 E_B = -\int \boldsymbol{\eta} \cdot (\nabla P_0) dV = -\int \nabla \cdot [P_0 \boldsymbol{\eta}] dV = 0$$

The first-order change in energy is evidently zero, as stated above. The second-order change is

$$\delta^2 E_B = \frac{1}{\mu}\int\left[\frac{1}{2}\left(\delta^1\mathbf{B}\right)^2 + \mathbf{B}_0 \cdot \delta^2\mathbf{B}\right]dV$$

from which,

$$\boxed{\delta^2 E_B = \frac{1}{2\mu}\int\left[\mathbf{b}^2 + \mathbf{B}_0 \cdot \nabla \times [\boldsymbol{\eta} \times \mathbf{b}]\right]dV, \qquad \mathbf{b} = \nabla \times [\boldsymbol{\eta} \times \mathbf{B}_0]} \quad (6.20a)$$

Now $\boldsymbol{\eta}$ is an imaginary displacement resulting from our virtual velocity field. However, we have $\boldsymbol{\eta} = \boldsymbol{\zeta} - \frac{1}{2}\boldsymbol{\zeta} \cdot \nabla\boldsymbol{\zeta} + \text{H.O.T.}$ and so when we substitute for $\boldsymbol{\eta}$ and discard cubic and higher-order terms we find that

$$\boxed{\delta^2 E_B = \frac{1}{2\mu}\int\left[\mathbf{b}^2 + \mathbf{B}_0 \cdot \nabla \times [\boldsymbol{\zeta} \times \mathbf{b}]\right]dV, \qquad \mathbf{b} = \nabla(\boldsymbol{\zeta} \times \mathbf{B}_0)} \quad (6.20b)$$

This gives us the instantaneous perturbations in magnetic energy and magnetic field (to leading order) at any instant in terms of the Lagrangian particle displacement field, $\boldsymbol{\zeta}(x, t)$. Now we expect the equilibrium to be stable if E_B is a minimum at equilibrium, i.e. $\delta^2 E_B > 0$. We now show that this is indeed the case. First we note that the total energy is conserved in ideal MHD, that is, the momentum equation gives us

$$\frac{D}{Dt}\left(\frac{1}{2}\rho\mathbf{u}^2\right) = -\nabla \cdot (P\mathbf{u}) + (\mathbf{J} \times \mathbf{B}) \cdot \mathbf{u}$$

while the dot product of $\mathbf{B}$ with the induction equation yields, after a little work,

$$\frac{\partial}{\partial t}\left(\mathbf{B}^2/(2\mu)\right) = (\mathbf{u} \times \mathbf{B}) \cdot \mathbf{J} + \nabla \cdot [(\mathbf{u} \times \mathbf{B}) \times (\mathbf{B}/\mu)]$$

Rearranging the scalar triple product and combining the two we have

$$\frac{\partial}{\partial t}\left[\frac{B^2}{2\mu} + \frac{\rho\mathbf{u}^2}{2}\right] = -\nabla \cdot \left[\left(p + \frac{1}{2}\rho\mathbf{u}^2\right)\mathbf{u} + (\mathbf{u} \times \mathbf{B}) \times (\mathbf{B}/\mu)\right]$$

If we take $\mathbf{B} \cdot d\mathbf{S} = \mathbf{u} \cdot d\mathbf{S} = 0$ at the boundary (we are assuming $\mathbf{B}$ is contained within V), then this gives us conservation of energy in the form

$$E = E_u + E_B = \frac{1}{2}\int\left[\rho\mathbf{u}^2 + \mathbf{B}^2/\mu\right]dV = \text{constant} \quad (6.21)$$

It follows that, for our perturbed magnetostatic equilibrium,

$$E - E_0 = \frac{1}{2}\int [\rho \mathbf{u}^2] dV + \delta^2 E_B = \text{constant}$$

(cubic and higher-order terms have been neglected here). We also have, to leading order in $\boldsymbol{\zeta}$, $\mathbf{u}(\mathbf{x}, t) = \dot{\boldsymbol{\zeta}}(\mathbf{x}, t)$. Conservation of energy therefore gives us

$$\boxed{\int \left[\frac{1}{2}\rho \dot{\zeta}^2\right] dV + \delta^2 E_B = \text{constant} = \Delta E} \tag{6.22}$$

where $\dot{\boldsymbol{\zeta}}$ indicates a partial derivative with respect to time. We can now, at last, discuss stability. We take as our definition of stability the condition that the kinetic energy of the disturbance is always bounded from above by the initial energy of the disturbance, ΔE. In effect, this limits the size of the resulting velocity field. It follows that an equilibrium is stable if $\delta^2 E_B$ is positive for all possible shapes of disturbances. That is to say, stability is ensured if $\delta^2 E_B > 0$ for all possible $\boldsymbol{\zeta}$ (or $\boldsymbol{\eta}$). Thus, to show that a magnetostatic equilibrium is stable we merely need to demonstrate that (6.20) is positive for any choice of $\boldsymbol{\eta}$. We have, in effect, a form of stability test.

All of this is in accord with our intuitive notions of stability. We may think of E_B as potential energy, in the sense that it is the conserved energy of a force acting on the fluid. Like a ball sitting on a hillside, the fluid (or ball) is in equilibrium if the potential energy is stationary, $\delta^1 E_B = 0$, and it is stable if the potential (i.e. magnetic) energy is a minimum (Figure 6.18).

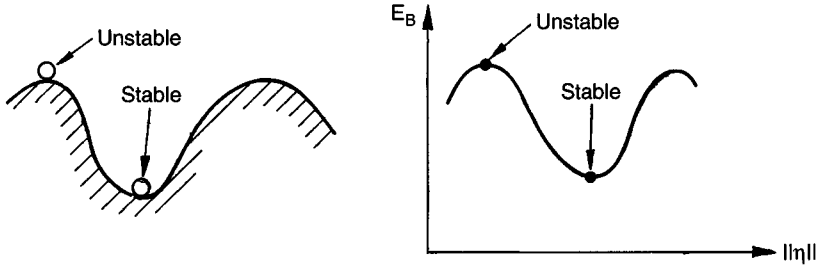

Figure 6.18 Analogy between magnetostatic and mechanical equilibria.

6.4.3 An alternative method for magnetostatic equilibrium

Now there is a different, though ultimately equivalent, route to establishing this stability criterion. This alternative method proves more useful when working with non-static equilibria, and so we shall describe it in some detail. The idea is to develop a dynamic equation for the disturbance. This time we work, not with the virtual displacement field, $\boldsymbol{\eta}$, but rather with the particle displacement, $\boldsymbol{\zeta}$. Of course, to leading order in the amplitude of the disturbance, $\boldsymbol{\zeta}$ and $\boldsymbol{\eta}$ are equal. We also work only with first-order quantities, such as $\mathbf{b} = \delta^1\mathbf{B}$, and discard all higher-order terms. The induction and momentum equations then give us the disturbance equations

$$\frac{\partial \mathbf{b}}{\partial t} = \nabla \times (\mathbf{u} \times \mathbf{B}_0)$$

$$\rho \frac{\partial \mathbf{u}}{\partial t} = \mathbf{j} \times \mathbf{B}_0 + \mathbf{J}_0 \times \mathbf{b} - \nabla p$$

Here lower-case letters represent perturbed quantities, e.g. $\mathbf{J} = \mathbf{J}_0 + \mathbf{j}$, and quadratic terms in the disturbance, such as $\mathbf{u} \cdot \nabla \mathbf{u}$ or $\mathbf{j} \times \mathbf{b}$, are neglected. We also have, to leading order,

$$\dot{\boldsymbol{\zeta}}(\mathbf{x}, t) = \mathbf{u}(\mathbf{x}, t), \qquad \nabla \cdot \boldsymbol{\zeta} = 0, \qquad \boldsymbol{\zeta} \cdot d\mathbf{S} = 0$$

The perturbation equations then give us

$$\mathbf{b} = \nabla \times (\boldsymbol{\zeta} \times \mathbf{B}_0) \tag{6.23}$$

$$(\rho\mu)\ddot{\boldsymbol{\zeta}} = (\nabla \times \mathbf{b}) \times \mathbf{B}_0 + (\nabla \times \mathbf{B}_0) \times \mathbf{b} - \nabla p \tag{6.24}$$

The first of these is a restatement of (6.17a), since $\boldsymbol{\zeta} = \boldsymbol{\eta}$ to leading order. The second equation may be rewritten as

$$(\rho\mu)\ddot{\boldsymbol{\zeta}} = \mathbf{F}(\boldsymbol{\zeta}) + \nabla(\cdot) \tag{6.25}$$

where $\nabla(\cdot)$ denotes the gradient of some scalar function whose value does not concern us, and

$$\mathbf{F}(\boldsymbol{\zeta}) = (\nabla \times \mathbf{b}) \times \mathbf{B}_0 + (\nabla \times \mathbf{B}_0) \times \mathbf{b}, \qquad \mathbf{b} = \nabla \times (\boldsymbol{\zeta} \times \mathbf{B}_0) \tag{6.26}$$

It is straightforward, but tedious, to show that the linear force operator $\mathbf{F}(\boldsymbol{\zeta})$ is self-adjoint, in the sense that

$$\int \boldsymbol{\zeta}_1 \cdot \mathbf{F}(\boldsymbol{\zeta}_2) dV = \int \boldsymbol{\zeta}_2 \cdot \mathbf{F}(\boldsymbol{\zeta}_1) dV \tag{6.27}$$

We now multiply (6.25) by $\dot{\zeta}$, and invoke (6.27) in the form, $\zeta_1 = \zeta$, $\zeta_2 = \dot{\zeta}$ The result is an energy-like equation

$$\frac{d}{dt}\int\left(\frac{1}{2}\rho\dot{\zeta}^2\right)dV = \frac{d}{dt}\left[\frac{1}{2\mu}\int \mathbf{F}(\zeta)\cdot\zeta dV\right] \tag{6.28}$$

The next step is to evaluate the integral on the right. In fact, it may be shown that

$$W(\zeta) = -\frac{1}{2\mu}\int \mathbf{F}(\zeta)\cdot\zeta dV = \delta^2 E_B \tag{6.29}$$

which, when combined with (6.28), gets us back to the energy stability criterion (6.22). The proof of (6.29) is a little involved and so we give here a schematic outline only.

Schematic proof of (6.29)

First we need a vector identity based on the equilibrium equation $\mathbf{J}_0 \times \mathbf{B}_0 = \nabla P_0$

$$\mathbf{J}_0 \times [\nabla \times (\mathbf{q} \times \mathbf{B}_0)] + [\nabla \times (\mathbf{q} \times \mathbf{J}_0)] \times \mathbf{B}_0 = -\nabla(\mathbf{q}\cdot\nabla P_0) \tag{6.30}$$

where $\mathbf{q}$ is any solenoidal field. (We shall not pause to prove (6.30).) Next we take $\mathbf{q} = \zeta$ and rewrite (6.26) as

$$\mathbf{F}(\zeta) = \nabla \times [\nabla \times (\zeta \times \mathbf{B}_0) - \zeta \times (\nabla \times \mathbf{B}_0)] \times \mathbf{B}_0 + \nabla(\cdot)$$

from which

$$\mathbf{F}(\zeta)\cdot\zeta = -(\zeta \times \mathbf{B}_0)\cdot\nabla \times [\nabla \times (\zeta \times \mathbf{B}_0) - \zeta \times (\nabla \times \mathbf{B}_0)] + \zeta\cdot\nabla(\cdot)$$

After a little algebra we find

$$\int \mathbf{F}(\zeta)\cdot\zeta dV = -\int [\mathbf{b}^2 + \mathbf{B}_0\cdot\nabla \times [\zeta \times \mathbf{b}]]dV \tag{6.31}$$

It is evident from (6.20b) that the right-hand integral is equal to $-2\mu\delta^2 E_B$ and so

$$W(\zeta) = -\frac{1}{2\mu}\int \mathbf{F}(\zeta)\cdot\zeta dV = \delta^2 E_B$$

as required.

It follows from (6.29) and our energy-like equation (6.28) that the sum of the kinetic energy of the disturbance, plus $\delta^2 E_B$, is conserved in the linear approximation, i.e.

$$\int \frac{1}{2}\rho\dot{\zeta}^2 dV + \delta^2 E_B = \text{constant} \tag{6.32}$$

Of course, this is identical to (6.22). Once again we conclude that the magnetostatic equilibrium is stable if $\delta^2 E_B$ is positive for all possible $\boldsymbol{\zeta}$ (or equivalently, all possible $\boldsymbol{\eta}$). This is known as *Bernstein's stability criterion.*

6.4.4 Proof that the energy method provides both necessary and sufficient conditions for stability

This second proof of (6.32) is less elegant than the first. However, it does set the scene for our more general stability analysis ($\mathbf{u}_0 \neq 0$). In fact, we can push this second method a little further. It can be used to establish both necessary and sufficient conditions for stability. That is to say, a magnetostatic equilibrium is stable if and only if E_B is a minimum. The proof of the necessity of $\delta^2 E_B > 0$ is as follows: suppose that $W(\boldsymbol{\zeta}) < 0$ for some $\boldsymbol{\zeta} = \boldsymbol{\zeta}^*$. Then we introduce a constant, γ, defined by

$$W(\boldsymbol{\zeta}^*) = -\gamma^2 \int \left[\frac{1}{2}\rho \boldsymbol{\zeta}^{*2}\right] dV$$

Next we note that (6.24) is second order in t, and so $\boldsymbol{\zeta}$ and $\dot{\boldsymbol{\zeta}}$ may be specified separately at $t = 0$. We choose $\boldsymbol{\zeta}(0) = \boldsymbol{\zeta}^*$ and $\dot{\boldsymbol{\zeta}}(0) = \gamma \boldsymbol{\zeta}^*$ The total disturbance energy is then zero and so, for all t,

$$\frac{1}{2}\int \rho \dot{\boldsymbol{\zeta}}^2 dV = -W(\boldsymbol{\zeta}) \tag{6.33}$$

We now return to (6.25) which, on multiplication by $\boldsymbol{\zeta}$ and integration over V, yields

$$\frac{1}{2}\ddot{I} = \frac{1}{2}\int \rho \dot{\boldsymbol{\zeta}}^2 dV - W(\boldsymbol{\zeta})$$

where $I = \frac{1}{2}\rho \int \boldsymbol{\zeta}^2 dV$. Combining this with (6.33) gives us

$$\frac{1}{2}\ddot{I} = \int \rho \dot{\boldsymbol{\zeta}}^2 dV \tag{6.34}$$

Now the Schwartz inequality tells us that

$$\dot{I}^2 \leq 2I \int \rho \dot{\boldsymbol{\zeta}}^2 dV$$

and so (6.34) may be rewritten as, $I\ddot{I} \geq \dot{I}^2$. This, in turn, ensures the exponential growth of any disturbance at a rate, $I \geq I(0)\exp[2\gamma t]$. [The proof of this last statement can be verified by making the substitution $y = \ln(I/I(0))$, which yields $\ddot{y} \geq 0$. Integrating this equation subject to $\dot{y}(0) = 2\gamma$ and $y(0) = 0$ gives $y \geq 2\gamma t$.] Thus a magnetostatic equilibrium

is stable if and only if E_B is a minimum. Unfortunately, when we extend the energy method to the stability of non-static equilibria ($\mathbf{u}_0 \neq 0$), we obtain only sufficient conditions for stability. That is to say, we can prove that a given flow is stable, but not that it is unstable.

6.4.5 *The stability of non-static equilibria*

We now repeat these arguments, but for equilibria in which $\mathbf{u}_0$ is non-zero. Our aim is to use conservation of energy to provide a sufficient condition for stability. To avoid carrying the constants ρ and μ all the way through the analysis, we take $\rho = \mu = 1$. (In effect, we rescale $\mathbf{B}$ as $\mathbf{B}/(\rho\mu)^{1/2}$.) Also, in the interests of simplicity, we shall take $\mathbf{B}$ to be confined to the fluid domain, V. Now the development of a more general stability criterion turns out to be no more difficult than the magnetostatic case, at least at a conceptual level. However, the algebra is long and tedious. We shall therefore give a schematic proof only. We start with

$$\frac{\partial \mathbf{u}}{\partial t} = \mathbf{u} \times \boldsymbol{\Omega} - \mathbf{B} \times \mathbf{J} - \nabla C, \qquad \mathbf{u} \cdot d\mathbf{S} = 0$$

$$\frac{\partial \mathbf{B}}{\partial t} = \nabla \times (\mathbf{u} \times \mathbf{B}), \qquad \mathbf{B} \cdot d\mathbf{S} = 0$$

which are the governing equations of ideal MHD. Here C is Bernoulli's function and $\boldsymbol{\Omega}$ is the vorticity. It is readily confirmed that these are consistent with the conservation of energy $E = \frac{1}{2}\int (\mathbf{u}^2 + \mathbf{B}^2) dV$ (see (6.21)). Steady flows are governed by

$$\mathbf{u}_0 \times \mathbf{B}_0 = \nabla D \tag{6.35}$$

$$\mathbf{u}_0 \times \boldsymbol{\Omega}_0 - \mathbf{B}_0 \times \mathbf{J}_0 = \nabla C_0 \tag{6.36}$$

and it is the stability of equilibria governed by these equations which concern us here. At this point it is convenient to introduce two vector identities, analogous to (6.30), which stem directly from the equilibrium equations. If $\mathbf{q}$ is any solenoidal field, then it may be shown that

$$\mathbf{u}_0 \times [\nabla \times (\mathbf{q} \times \mathbf{B}_0)] + [\nabla \times (\mathbf{q} \times \mathbf{u}_0)] \times \mathbf{B}_0 = -\nabla(\mathbf{q} \cdot \nabla D) \tag{6.37}$$

$$\begin{aligned}\mathbf{u}_0 \times [\nabla \times (\mathbf{q} \times \boldsymbol{\Omega}_0)] + [\nabla \times (\mathbf{q} \times \mathbf{u}_0)] \times \boldsymbol{\Omega}_0 + \mathbf{J}_0 \times [\nabla \times (\mathbf{q} \times \mathbf{B}_0)] \\ + [\nabla \times (\mathbf{q} \times \mathbf{J}_0)] \times \mathbf{B}_0 = -\nabla(\mathbf{q} \cdot \nabla C_0)\end{aligned} \tag{6.38}$$

We shall not pause to prove these uninspiring-looking relationships, but they will be used in the analysis which follows.

We now consider small-amplitude perturbations in $\mathbf{u}_0$ and $\mathbf{B}_0$, in which $\mathbf{B} = \mathbf{B}_0 + \mathbf{b}$, $(\mathbf{b} \cdot d\mathbf{S} = 0)$ and $\mathbf{u} = \mathbf{u}_0 + \delta\mathbf{u}$, $(\delta\mathbf{u} \cdot d\mathbf{S} = 0)$. Related quantities are $\mathbf{j} = \nabla \times \mathbf{b}$ and $\omega = \nabla \times (\delta\mathbf{u})$. In the analysis which follows we shall ignore all quantities which are quadratic, or of higher order, in the amplitude of the disturbance. As with the magnetostatic stability analysis, our first step is to introduce the particle displacement field $\boldsymbol{\zeta}(\mathbf{x}, t)$, defined by

$$\boldsymbol{\zeta}(\mathbf{x}, t) = \mathbf{x}_p(t) - \mathbf{x}_{p0}(t)$$

where $\mathbf{x}_{p0}$ is the position vector of particle p in the base flow and $\mathbf{x}_p$ is the position of the same particle in the perturbed flow. The generalisation of (6.15) is then

$$\frac{D\boldsymbol{\zeta}}{Dt} = \mathbf{u}(\mathbf{x} + \boldsymbol{\zeta}, t) - \mathbf{u}_0(\mathbf{x})$$

In the linear (small-amplitude) approximation, this becomes

$$\frac{\partial \boldsymbol{\zeta}}{\partial t} + \mathbf{u}_0 \cdot \nabla\boldsymbol{\zeta} = \delta\mathbf{u}(\mathbf{x}, t) + \mathbf{u}_0(\mathbf{x} + \boldsymbol{\zeta}) - \mathbf{u}_0(\mathbf{x})$$

which, using the approximation $\mathbf{u}_0(\mathbf{x} + \boldsymbol{\zeta}) - \mathbf{u}_0(\mathbf{x}) = \boldsymbol{\zeta} \cdot \nabla\mathbf{u}_0$, simplifies to

$$\frac{\partial \boldsymbol{\zeta}}{\partial t} = \delta\mathbf{u}(\mathbf{x}, t) - \nabla \times (\boldsymbol{\zeta} \times \mathbf{u}_0) \tag{6.39}$$

Note that, in the small-amplitude approximation, $\boldsymbol{\zeta}$ is solenoidal. Also, since $\mathbf{u} \cdot d\mathbf{S} = 0$, we have $\boldsymbol{\zeta} \cdot d\mathbf{S} = 0$. We now turn to the perturbation equations $\mathbf{b}$ and $\delta\mathbf{u}$. When we discard quadratic and higher-order terms in the perturbation, we find

$$\frac{\partial \mathbf{b}}{\partial t} = \nabla \times [\delta\mathbf{u} \times \mathbf{B}_0 + \mathbf{u}_0 \times \mathbf{b}] \tag{6.40}$$

$$\frac{\partial (\delta\mathbf{u})}{\partial t} = \delta\mathbf{u} \times \boldsymbol{\Omega}_0 + \mathbf{u}_0 \times \omega - \mathbf{b} \times \mathbf{J}_0 - \mathbf{B}_0 \times \mathbf{j} - \nabla c \tag{6.41}$$

We concentrate first on the induction equation. Introducing, $\Delta\mathbf{B} = \mathbf{b} - \nabla \times (\boldsymbol{\zeta} \times \mathbf{B}_0)$, this may be rewritten as

$$\frac{\partial}{\partial t}(\Delta\mathbf{B}) = \nabla \times [\mathbf{u}_0 \times \Delta\mathbf{B} + \mathbf{u}_0 \times (\boldsymbol{\zeta} \times \mathbf{B}_0) + \nabla \times (\boldsymbol{\zeta} \times \mathbf{u}_0) \times \mathbf{B}_0]$$

which, by virtue of (6.37), simplifies to

$$\frac{\partial}{\partial t}(\Delta\mathbf{B}) = \nabla \times [\mathbf{u}_0 \times \Delta\mathbf{B}] \tag{6.42}$$

Evidently, if we set $\Delta\mathbf{B} = 0$ at some initial instant, then $\Delta\mathbf{B}$ remains zero for all time. Let us assume that this is so. We then have

$$\mathbf{b} = \nabla \times (\boldsymbol{\zeta} \times \mathbf{B}_0) \tag{6.43}$$

which is identical to (6.23). Setting $\Delta\mathbf{B} = 0$ in the initial condition is therefore equivalent to assuming that $\mathbf{B}$ is frozen into the fluid during the initial disturbance. Such a disturbance might be triggered by, say, a pressure pulse travelling through the fluid. In the analysis which follows, therefore, we shall assume that $\Delta\mathbf{B} = 0$ at $t = 0$ so that (6.43) holds at all times.

We now turn to the momentum equation (6.41). Substituting for $\mathbf{b}$ and $\delta\mathbf{u}$ using (6.39) and (6.43) we find, after a little algebra, that

$$\boxed{\ddot{\boldsymbol{\zeta}} + 2\mathbf{u}_0 \cdot \nabla(\dot{\boldsymbol{\zeta}}) = \mathbf{F}(\boldsymbol{\zeta}) + \nabla(\cdot)} \tag{6.44}$$

Here $\mathbf{F}$ is given by

$$\boxed{\mathbf{F}(\boldsymbol{\zeta}) = (\nabla \times \mathbf{b}) \times \mathbf{B}_0 + (\nabla \times \mathbf{B}_0) \times \mathbf{b} - (\nabla \times \hat{\mathbf{u}}) \times \mathbf{u}_0 - (\nabla \times \mathbf{u}_0) \times \hat{\mathbf{u}}} \tag{6.45}$$

and $\mathbf{b}$ and $\hat{\mathbf{u}}$ are defined by

$$\boxed{\mathbf{b} = \nabla \times (\boldsymbol{\zeta} \times \mathbf{B}_0); \qquad \hat{\mathbf{u}} = \nabla \times (\boldsymbol{\zeta} \times \mathbf{u}_0)} \tag{6.46}$$

It should be noted that, while $\mathbf{b}$ represents the perturbation in $\mathbf{B}$, $\hat{\mathbf{u}}$ does not represent the perturbation in $\mathbf{u}$. Rather, $\hat{\mathbf{u}}$ is the difference between $\delta\mathbf{u}$ and $\dot{\boldsymbol{\zeta}}$: $\hat{\mathbf{u}} = \delta\mathbf{u} - \dot{\boldsymbol{\zeta}}$.

Now compare (6.44) and (6.45) with (6.25) and (6.26). It is clear that we have extended the dynamical equation for $\boldsymbol{\zeta}$ from equilibria in which $\mathbf{u}_0$ is zero to those where it is not. Note that (6.44) and (6.45) reduce to the magnetostatic perturbation equations (6.25) and (6.26) when $\mathbf{u}_0 = 0$, as they should. Note also the skew-symmetric rôles played by $\mathbf{B}_0$ and $\mathbf{u}_0$ in (6.45). It is now a small step to obtain a sufficient condition for stability. In effect, we simply repeat the arguments used in the magnetostatic case. As before, $\mathbf{F}$ is self-adjoint:

$$\int \mathbf{F}(\boldsymbol{\zeta}_1) \cdot \boldsymbol{\zeta}_2 dV = \int \mathbf{F}(\boldsymbol{\zeta}_2) \cdot \boldsymbol{\zeta}_1 dV \tag{6.47}$$

and so (6.44), multiplied by $\dot{\boldsymbol{\zeta}}$, yields

$$\frac{d}{dt}\int \frac{1}{2}\dot{\zeta}^2 dV = \frac{d}{dt}\left[\frac{1}{2}\int \mathbf{F}(\zeta)\cdot\zeta dV\right] \tag{6.48}$$

Also, using (6.37) and (6.38), we can determine the analogue of (6.29). After a little work we find

$$\boxed{W(\zeta) = -\frac{1}{2}\int \mathbf{F}(\zeta)\cdot\zeta dV = \delta^2 E_B - \frac{1}{2}\int\left[(\hat{\mathbf{u}})^2 + \mathbf{u}_0\cdot\nabla\times[\zeta\times\hat{\mathbf{u}}]\right]dV} \tag{6.49}$$

which leads to the conservation equation

$$\boxed{\frac{1}{2}\int\dot{\zeta}^2 dV + \delta^2 E_B - \frac{1}{2}\int\left[(\hat{\mathbf{u}})^2 + \mathbf{u}_0\cdot\nabla\times(\zeta\times\hat{\mathbf{u}})\right]dV = \text{constant}} \tag{6.50}$$

or

$$\boxed{\frac{1}{2}\int\dot{\zeta}^2 dV + W(\zeta) = e = \text{constant}} \tag{6.51}$$

If we take $\int\dot{\zeta}^2 dV$ as a measure of our disturbance, then the equilibrium flow $(\mathbf{u}_0, \mathbf{B}_0)$ is stable whenever $W(\zeta)$ is positive for all possible choices of ζ. This is, in effect, a generalisation of Bernstein's criterion and was developed in the 1960s by researchers working at Princeton on plasma containment.

Unfortunately, we cannot extend the argument to give a necessary condition for stability. The term involving $\mathbf{u}_0$ on the right of (6.44) prevents us from repeating the arguments used for the magnetostatic case. In fact, it is not difficult to construct flows which are stable yet admit negative values of $W(\zeta)$. Consider the axisymmetric flow $\mathbf{u}_0 = \Omega r\hat{\mathbf{e}}_\theta$, $\mathbf{B}_0 = \alpha\mathbf{u}_0$ (for some constants α, Ω) which is confined to $r < R$. If the two-dimensional stability of this flow is examined then (6.44) leads (eventually) to the dynamic equation,

$$\left[\frac{\partial}{\partial t} + \Omega(1+\alpha)\frac{\partial}{\partial\theta}\right]\left[\frac{\partial}{\partial t} + \Omega(1-\alpha)\frac{\partial}{\partial\theta}\right]\nabla^2(r\zeta_r) = 0$$

The resulting solutions are stable Alfvén waves travelling (clockwise or anti-clockwise) along the $\mathbf{B}_0$-lines and riding on the back of the base flow.

$$\zeta_r = \zeta_r(\theta - \Omega(1 \pm \alpha)t)$$

However, it is readily confirmed that $W(\zeta)$ is negative whenever $|\alpha| < 1$, and so this is an example of a flow which violates our stability criterion, yet is perfectly stable.

In summary then, equilibrium solutions of the ideal MHD equations are stable to small disturbances provided that $W(\zeta)$, defined by

$$W(\zeta) = \frac{1}{2}\int [\mathbf{b}^2 + \mathbf{B}_0 \cdot \nabla \times (\boldsymbol{\zeta} \times \mathbf{b})]\, dV - \frac{1}{2}\int [\hat{\mathbf{u}}^2 + \mathbf{u}_0 \cdot \nabla \times (\boldsymbol{\zeta} \times \hat{\mathbf{u}})]\, dV \tag{6.52}$$

$$(\mathbf{b} = \nabla \times (\boldsymbol{\zeta} \times \mathbf{B}_0), \qquad \hat{\mathbf{u}} = \nabla \times (\boldsymbol{\zeta} \times \mathbf{u}_0))$$

is positive for all possible choices of $\boldsymbol{\zeta}$. This is a remarkably general result which has been rediscovered many times by alternative means. It covers magnetostatics, ideal MHD, and inviscid flows in the absence of a magnetic field. Unfortunately, there are relatively few three-dimensional flows for which $W(\zeta)$ can be shown to be positive. However, there are many two-dimensional flows which may be shown to be stable by this method.

It turns out that a simpler derivation of (6.51) and (6.52) may be formulated by appealing directly to Lagrange's equation, and this is described in Appendix 2.

6.5 Conclusion

This concludes our brief exploration of high-R_m dynamics. The subject is an attractive one, rich in physical phenomena and full of unresolved problems. For example, we have discussed stability criteria only in the context of incompressible flows. Yet ideal MHD only really holds in plasma MHD, not liquid-metal MHD, and so actually we want stability criteria for compressible fluids. Then there is dynamo theory. While kinematic aspects of the subject seem well understood, there are many unanswered questions concerning the dynamics of the geo-dynamo and solar dynamo. The interested reader is urged to consult the references given below.

Suggested Reading

H K Moffatt, *Magnetic field generation in electrically conducting fluids*, 1978. Cambridge University Press. (Chapters 6–12 for a very detailed discussion of dynamo theory.)

P H Roberts, *An introduction to magnetohydrodynamics,* 1967, Longmans. (Chapter 3 for dynamo theory (without the α-effect), Chapter 5 for Alfvén waves, Chapters 8 and 9 for stability theory.)

M R E Proctor & A D Gilbert, *Lectures on solar and planetary dynamos*, 1994. Cambridge University Press. (Chapter 1, by P H Roberts, for an introduction to geo-dynamo theory, Chapter 2, by N O Weiss, for solar MHD.)

D Biskamp, *Nonlinear magnetohydrodynamics*, 1993. Cambridge University Press. (Chapter 4 for stability theory.)

Examples

6.1 In an ideal fluid ($\lambda = \nu = 0$) there exist a velocity field $\mathbf{u}$ and a scaled magnetic field $\mathbf{h} = \mathbf{B}/(\rho\mu)^{1/2}$. Now consider the alternative fields, $\mathbf{v}_1 = \mathbf{u} + \mathbf{h}$ and $\mathbf{v}_2 = \mathbf{u} - \mathbf{h}$. Show that these fields are governed by

$$\partial \mathbf{v}_1/\partial t + (\mathbf{v}_2 \cdot \nabla)\mathbf{v}_1 = -\nabla(p/\rho)$$

$$\partial \mathbf{v}_2/\partial t + (\mathbf{v}_1 \cdot \nabla)\mathbf{v}_2 = -\nabla(p/\rho)$$

where p is the sum of the fluid pressure and the magnetic pressure ($\mathbf{v}_1$ and $\mathbf{v}_2$ are known as Elsasser variables).

6.2 Suppose that a uniform magnetic field permeates an (almost) inviscid, (almost) perfectly conducting fluid and is orientated at right angles to a plane solid surface which forms the boundary of the semi-infinite fluid domain. A constant current sheet, J, is suddenly applied in the stationary wall giving rise to a tangential field, $\mathbf{B}_{//}$, at the wall. Show that jumps in $\mathbf{u}_{//}$ and $\mathbf{B}_{//}$ across the Hartmann-vortex sheet at the wall are related by

$$|\Delta \mathbf{B}_{//}|/(\rho\mu)^{1/2} = |\Delta \mathbf{u}_{//}|(\nu/\lambda)^{1/2}$$

Determine the subsequent motion and field distribution for the case where $\nu \ll \lambda$.

6.3 Show that W in (6.52) is always sign-indefinite for three-dimensional equilibria in which $\mathbf{B}$ and $\mathbf{u}$ are not aligned. (The inference is that such equilibria are usually unstable.)

6.4 Consider a two-dimensional magnetic field which is in equilibrium and sits in a steady, two-dimensional velocity field. Show that W in (6.52) is always sign-indefinite if $B^2 < \rho\mu u^2$ at any point. (Restrict the analysis to two-dimensional stability.)