

Solar and Stellar Dynamos

Paul Charbonneau

Paul Charbonneau: *Solar and Stellar Dynamos* in Magnetic Fields of Stars: From the Sun to Compact Objects, Saas-Fee Advanced Course 39 of the Swiss Society of Astrophysics & Astronomy, pp. 1-xx (2010).

Contents

Solar and Stellar Dynamos	1
Paul Charbonneau	
1 Magnetohydrodynamics	7
0.3 The fluid approximation	8
0.4 Essentials of hydrodynamics	10
0.5 The magnetohydrodynamical induction equation ...	19
0.6 Scaling analysis	21
0.7 The Lorentz force	23
0.8 Joule heating	26
0.9 The full set of MHD equations	26
0.10 MHD waves	27
0.11 Magnetic energy	28
0.12 Magnetic flux freezing and Alfvén's theorem	29
0.13 The magnetic vector potential	30
0.14 Magnetic helicity	31
0.15 Force-free magnetic fields	31
0.16 The ultimate origin of astrophysical magnetic fields	32
0.17 The astrophysical dynamo problem(s)	35
2 Decay and Amplification of Magnetic Fields	43
0.18 Resistive decays of magnetic fields	43
0.19 Magnetic field amplification by stretching and shearing	50
0.20 Magnetic field evolution in a cellular flow	59
0.21 Two anti-dynamo theorems	72
0.22 The Roberts cell dynamo	75
0.23 The CP flow and fast dynamo action	84
0.24 Dynamo action in turbulent flows	91

3	Dynamo models of the solar cycle	97
0.25	The solar magnetic field	98
0.26	Mean-field dynamo models	106
0.27	Babcock-Leighton models	131
0.28	Models based on HD and MHD instabilities	144
0.29	Global MHD simulations	148
4	Fluctuations, intermittency and predictivity	159
0.30	Observed patterns of solar cycle variations	160
0.31	Cycle modulation through stochastic forcing	170
0.32	Cycle modulation through the Lorentz force	175
0.33	Cycle modulation through time delays	178
0.34	Intermittency	180
0.35	Model-based cycle predictions	184
5	Stellar dynamos	193
0.36	Early-type stars	195
0.37	A-type stars	201
0.38	Solar-type stars	204
0.39	Fully convective stars	213
0.40	Pre- and post-main-sequence stars	214
0.41	Compact objects	216
1	Galaxies and beyond	217
	Appendix A: Useful identities and theorems from vector calculus ..	222
	A.1 Vector Identities	222
	A.2 The Gradient Theorem	223
	A.3 The Divergence Theorem	223
	A.4 Stokes' Theorem	223
	Appendix B: Coordinate Systems and the Fluid Equations	223
	B.1 Cylindrical coordinates (s, ϕ, z)	224
	B.1.1 Conversion to cartesian coordinates	224
	B.1.2 Infinitesimals	225
	B.1.3 Vector operators	225
	B.1.4 The divergence of a second-order tensor	226
	B.1.5 Components of the viscous stress tensor	226
	B.1.6 Equations of motion	226
	B.1.7 The energy equation	227
	B.1.8 The MHD induction equation	227
	B.2 Spherical coordinates (r, θ, ϕ)	228
	B.2.1 Conversion to cartesian coordinates	228
	B.2.2 Infinitesimals	229
	B.2.3 Operators	229
	B.2.4 The divergence of a second-order symmetric tensor	230
	B.2.5 Components of the viscous stress tensor	231
	B.2.6 Equations of motion	231

B.2.7 The energy equation	232
B.2.8 The MHD induction equation	232
Appendix C: Physical and astronomical constants	233
C.1 Physical constants	233
C.2 Astronomical constants	233
Appendix D: Maxwell's equations and physical units	233
D.1 Maxwell's equations	234
D.2 Conversion of units	236 {pch_sec:1.15}

Chapter 1

Magnetohydrodynamics

{pch_chap:mhd}

From a long view of history —seen from, say, ten thousand years from now— there can be little doubt that the most significant event of the 19th century will be judged as Maxwell’s discovery of the laws of electrostatics.

*Richard Feynman
The Feynman Lectures on Physics, vol. II (1964)*

To sum it all up in a single sentence, *magnetohydrodynamics* (hereafter MHD) is concerned with the behavior of electrically conducting but globally neutral fluids flowing at non-relativistic speeds and obeying Ohm’s Law. Remarkably, most astrophysical fluids meet these apparently stringent requirements, the most glaring exception being the relativistic inflows and outflows powered by compact objects such as black holes or neutron stars.

The focus of these lectures is on the amplification of solar and stellar magnetic fields through the inductive action of fluid flows, a process believed to be well-described by MHD for physical conditions characterizing the interior of the sun and (most) stars. Before we dive into MHD proper, we will first clarify what we mean by “fluid” (§0.3), and review the fundamental physical laws governing the flow of unmagnetized fluid, i.e., classical hydrodynamics (§0.4). We then introduce magnetic fields into the fluid picture (§§0.5—0.15), and close by reflecting upon the ultimate origin of astrophysical magnetic fields (§0.16), and establishing the various incarnations of the so-called dynamo problem (§0.17) which will occupy our attention in the subsequent chapters.

0.3 The fluid approximation

{pch_sec:fluid}

0.3.1 Matter as a continuum

{pch_ssec:continuum}

It did take some two thousand years to figure it out, but we now know that Democritus was right after all: matter is composed of small, microscopic “atomic” constituents. Yet on our daily macroscopic scale, things sure look smooth and continuous. Under what circumstances can an assemblage of microscopic elements be treated as a continuum? The key constraint is that there be a good *separation of scales* between the “microscopic” and “macroscopic”.

Consider the situation depicted on Figure 1, corresponding to an amorphous substance (spatially random distribution of microscopic constituents). Denote by λ the mean interparticle distance, and by L the macroscopic scale of the system; we now seek to construct macroscopic variables defining fluid characteristics at the macroscopic scale. For example, if we are dealing with an assemblage of particles of mass m , then the *density* (ρ) associated with a cartesian volume element of linear dimensions l centered at position \mathbf{x} would be given by something like:

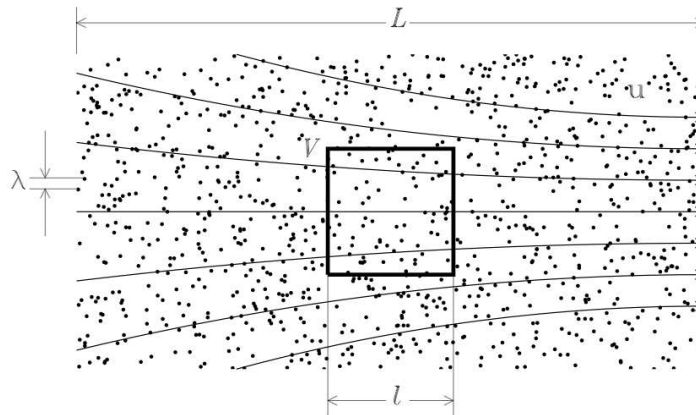


Fig. 1 fig:continuum Microscopic view of a fluid. In general the velocity of microscopic constituents is comprised of two parts: a randomly-oriented thermal velocity, and a systematic drift velocity, which, on the macroscopic scale amounts to what we call a flow \mathbf{u} . A fluid representation is possible if the mean inter-particle distance λ is much smaller than the global length scale L .

{eq:density}

$$\rho(\mathbf{x}) = \frac{1}{l^3} \sum_k m_k, \quad [\text{kg m}^{-3}], \quad (1)$$

where the sum runs over all particles contained within the volume element. One often hears or reads that for a continuum representation to hold, it is only necessary that the particle density be “large”. But how large, and large with respect to what? For the above expression to yield a well-defined quantity, in the sense that the numerical value of ρ so computed does not depend sensitively on the size and location of the volume element, or on time if the particles are moving, it is essential that a great many particles be contained within the element. Moreover, if we want to be writing differential equations describing the evolution of ρ , the volume element better be infinitesimal, in the sense that it is much smaller than the macroscopic length scale over which global variables such as ρ may vary. These two requirements translate in the double inequality:

$$\lambda \ll l \ll L . \tag{2} \quad \{\text{eq:scales}\}$$

Because the astrophysical systems and flows that will be the focus of our attention span a very wide range of macroscopic sizes, the continuum/fluid representation will turn out to hold in circumstances where the density is in fact minuscule, as you can verify for yourself upon perusing the collection of astrophysical systems listed in Table 1 below¹. In all cases, a very good separation of scales does exist between the microscopic (λ) and macroscopic (L).

Table 1 Spatial scales of some astrophysical objects and flows

{pch_tab1}

System/flow	ρ [kg/m ³]	N [m ⁻³]	λ [m]	L [km]
Solar interior	100	10 ²⁹	10 ⁻¹⁰	10 ⁵
Solar atmosphere	10 ⁻⁴	10 ²³	10 ⁻⁸	10 ³
Solar corona	10 ⁻¹¹	10 ¹⁷	10 ⁻⁶	10 ⁵
Solar wind (1 AU)	10 ⁻²¹	10 ⁷	0.006	10 ⁵
Molecular cloud	10 ⁻²⁰	10 ⁷	0.001	10 ¹⁴
Interstellar medium	10 ⁻²¹	10 ⁶	0.01	10 ¹⁶

¹ All density-related estimates assume a gas of fully ionized Hydrogen ($\mu = 0.5$) for the Sun, of neutral Hydrogen for the interstellar medium ($\mu = 1$), and molecular Hydrogen ($\mu = 2$) for molecular clouds. Solar densities are for the base of the convection zone (solar interior), optical depth unity (atmosphere), and typical coronal loop (corona). N is the number density of microscopic constituents. The length scale listed for the solar atmosphere is the granulation dimension, for the corona it is the length of a coronal loop, for the solar wind the size of Earth’s magnetosphere, and that for the interstellar medium is the thickness of the galactic (stellar) disk; all rounded to the nearest factor of ten.

0.3.2 Solid versus fluid

Most continuous media can be divided into two broad categories, namely *solids* and *fluids*. The latter does not just include the usual “liquids” of the vernacular, but also gases and plasmas. Physically, the distinction is made on the basis of a medium’s response to an applied *stress*, as illustrated on Figure 2. A volume element of some continuous substance is subjected to a shear stress, i.e., two force acting tangentially and in opposite directions on two of its parallel bounding surface (black arrows). A *solid* will immediately generate a restoring force (white arrows), ultimately due to electrostatic interactions between its microscopic constituents, and vigorously resist deformation (try shearing a brick held between the palms of your hands!). The solid will rapidly reach a new equilibrium state characterized by a finite deformation, and will relax equally rapidly to its initial state once the external stress vanishes. A *fluid*, on the other hand, can offer no resistance to the applied stress, at least in the initial stages of the deformation.

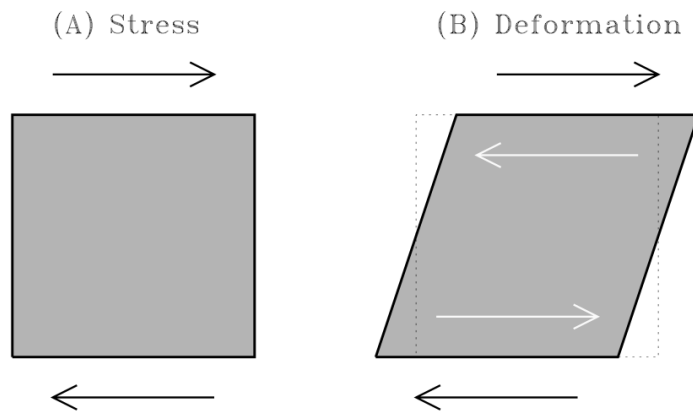


Fig. 2 ^{fig:stress} Deformation of a mass element in response to a stress pattern producing an horizontal shear (black arrows). A solid will rapidly reach an equilibrium where internal stresses (white arrows) produced by the deformation will equilibrate the applied shear. A fluid at rest cannot generate internal stresses, and so will be increasingly deformed for as long as the external shear is applied.

{pch_sec:hydro}

0.4 Essentials of hydrodynamics

The governing principles of classical hydrodynamics are the same as those of classical mechanics, transposed to continuous media: conservation of mass, linear momentum, angular momentum and energy. The fact that these prin-

ciples must now be applied not to point-particles, but to spatially extended and deformable volume elements (which may well be infinitesimal, but they are still finite!) introduces some significant complications, mostly with regards to the manner in which forces act. Let's start with the easiest of our conservation statements, that for mass, as it exemplifies very well the manner in which conservation laws are formulated in moving fluids.

0.4.1 Mass: the continuity equation

{pch_ssec:mcons}

Consider the situation depicted on Figure 3, namely that of an arbitrarily shaped fictitious surface S fixed in space and enclosing a volume V embedded in a fluid of density $\rho(\mathbf{x})$ moving with velocity $\mathbf{u}(\mathbf{x})$. The *mass flux* associated with the flow across the (closed) surface is

$$\Phi = \oint_S \rho \mathbf{u} \cdot \hat{\mathbf{n}} dS, \quad [\text{kg s}^{-1}] \quad (3) \quad \{\text{eq:mcons1}\}$$

where $\hat{\mathbf{n}}$ is a unit vector everywhere perpendicular to the surface, and by

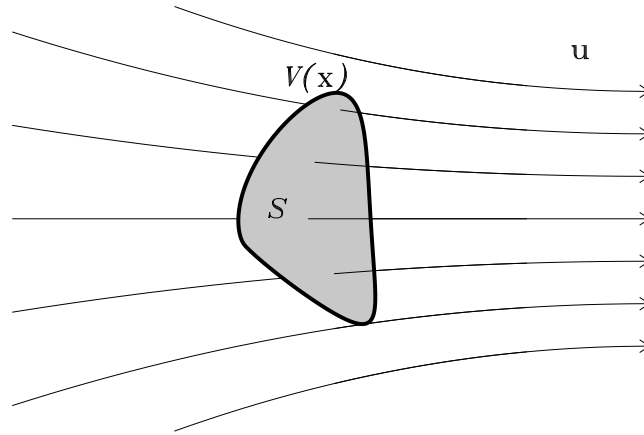


Fig. 3 ^{fig:mcons1} An arbitrarily shaped volume element V bounded by a closed surface S , both fixed in space, and traversed by a flow \mathbf{u} .

convention oriented towards the exterior. The mass of fluid contained within V is simply

$$M = \int_V \rho dV. \quad [\text{kg}] \quad (4) \quad \{\text{eq:mcons3}\}$$

This quantity will evidently vary if the mass flux given by eq. (3) is non-zero:

$$\frac{\partial M}{\partial t} = -\Phi . \quad (5) \quad \{\text{eq:mcons4}\}$$

Here the minus sign is a direct consequence of the exterior orientation of $\hat{\mathbf{n}}$. Inserting eq. (3) and eq. (4) into (5) and applying the divergence theorem to the RHS of the resulting expression yields:

$$\{\text{eq:mcons6}\} \quad \frac{\partial}{\partial t} \int_V \rho dV = - \int_V \nabla \cdot (\rho \mathbf{u}) dV . \quad (6)$$

Because V is fixed in space, the $\partial/\partial t$ et \int_V operators commute, so that

$$\{\text{eq:mcons7}\} \quad \int_V \left[\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{u}) \right] dV = 0 . \quad (7)$$

Because V is completely arbitrary, in general this can only be satisfied provided that

$$\{\text{eq:mcons}\} \quad \boxed{\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{u}) = 0} . \quad (8)$$

This expresses mass conservation in differential form, and is known in hydrodynamics as the *continuity equation*.

Incompressible fluids have constant densities, so that in this limiting case the continuity equation reduces to

$$\{\text{eq:incompress}\} \quad \nabla \cdot \mathbf{u} = 0 , \quad [\text{incompressible}] . \quad (9)$$

Water is perhaps the most common example of an effectively incompressible fluid (under the vast majority of naturally occurring conditions anyway). The gaseous nature of most astrophysical fluids may lead you to think that incompressibility is likely to be a pretty lousy approximation in cases of interest in this course. It turns out that the incompressibility can lead to a pretty good approximation of the behavior of compressible fluids provided that the flow's Mach number (ratio of flow speed to sound speed) is much smaller than unity.

`{pch_ssec:doper}` **0.4.2 The D/Dt operator**

Suppose we want to compute the time variation of some physical quantity (Z , say) at some fixed location \mathbf{x}_0 in a flow $\mathbf{u}(\mathbf{x})$. In doing so we must take into account the fact that Z is in general both an explicit and implicit function of time, because the volume element “containing” Z is moving with the fluid, i.e., $Z \rightarrow Z(t, \mathbf{x}(t))$. We therefore need to use the chain rule and write:

$$\{\text{eq:doper2}\} \quad \frac{dZ}{dt} = \frac{\partial Z}{\partial t} + \frac{\partial Z}{\partial x} \frac{\partial x}{\partial t} + \frac{\partial Z}{\partial y} \frac{\partial y}{\partial t} + \frac{\partial Z}{\partial z} \frac{\partial z}{\partial t} . \quad (10)$$

Noting that $\mathbf{u} = d\mathbf{x}/dt$, this becomes

$$\frac{dZ}{dt} = \frac{\partial Z}{\partial t} + \frac{\partial Z}{\partial x}u_x + \frac{\partial Z}{\partial y}u_y + \frac{\partial Z}{\partial z}u_z = \frac{\partial Z}{\partial t} + (\mathbf{u} \cdot \nabla)Z . \quad (11)$$

This corresponds to the time variation of Z following the fluid element as it is carried by the flow. It is a very special kind of derivative in hydrodynamics, known as the *Lagrangian derivative*, which will be represented by the operator:

$$\boxed{\frac{D}{Dt} \equiv \frac{\partial}{\partial t} + (\mathbf{u} \cdot \nabla)} . \quad (12)$$

Note in particular that the Lagrangian derivative of \mathbf{u} yields the acceleration of a fluid element:

$$\mathbf{a} = \frac{D\mathbf{u}}{Dt} , \quad (13)$$

a notion that will soon come very handy when we write $F = ma$ for a fluid.

A *material surface* is defined as an ensemble of points that define a surface, all moving along with the flow. Therefore, in a local frame of reference S' co-moving with any infinitesimal element of a material surface, $\mathbf{u}' = 0$. The distinction between material surfaces, as opposed to surfaces fixed in space such as in eq. (3), has crucial consequences with respect to the commuting properties of temporal and spatial differential operators. In the latter case \int_V commutes with $\partial/\partial t$, whereas for material surfaces and volume elements it is D/Dt that commutes with \int_V (and \oint_S , etc.).

0.4.3 Linear momentum: the Navier-Stokes equations

{pch_ssec:NS}

A force \mathbf{F} acting on a point-object of mass m is easy to deal with; it simply produces an acceleration $\mathbf{a} = \mathbf{F}/m$ in the same direction as the force (sounds simple but it still took the genius of Newton to figure it out...). In the presence of a force acting on the surface of a spatially extended fluid element, the resulting fluid acceleration will depend on both the orientation of the force and the surface. We therefore define the net force \mathbf{t} in terms of a *stress tensor*:

$$\mathbf{t}_x = \hat{\mathbf{e}}_x s_{xx} + \hat{\mathbf{e}}_y s_{xy} + \hat{\mathbf{e}}_z s_{xz} , \quad (14)$$

$$\mathbf{t}_y = \hat{\mathbf{e}}_x s_{yx} + \hat{\mathbf{e}}_y s_{yy} + \hat{\mathbf{e}}_z s_{yz} , \quad (15)$$

$$\mathbf{t}_z = \hat{\mathbf{e}}_x s_{zx} + \hat{\mathbf{e}}_y s_{zy} + \hat{\mathbf{e}}_z s_{zz} , \quad (16)$$

where “ s_{xy} ” denotes the force per unit area acting in the y -direction on a surface perpendicular to the x -direction, \mathbf{t}_x is the net force acting on the surfaces perpendicular to the x -direction, and similarly for the other components. Consider now a unit vector perpendicular to a surface arbitrarily oriented in space:

$$\hat{\mathbf{n}} = \hat{\mathbf{e}}_x n_x + \hat{\mathbf{e}}_y n_y + \hat{\mathbf{e}}_z n_z, \quad n_x^2 + n_y^2 + n_z^2 = 1. \quad (17)$$

The net force along this direction is simply

$$\mathbf{t}_{\hat{\mathbf{n}}} = (\hat{\mathbf{n}} \cdot \hat{\mathbf{e}}_x) \mathbf{t}_x + (\hat{\mathbf{n}} \cdot \hat{\mathbf{e}}_y) \mathbf{t}_y + (\hat{\mathbf{n}} \cdot \hat{\mathbf{e}}_z) \mathbf{t}_z = \hat{\mathbf{n}} \cdot \mathbf{s}. \quad (18)$$

We can now use the Lagrangian acceleration to write “ $\mathbf{a} = \mathbf{F}/m$ ” for the fluid element:

$$\frac{D}{Dt} \int_V \mathbf{u} dV = \frac{1}{\rho} \oint_S \mathbf{s} \cdot \hat{\mathbf{n}} dS. \quad (19)$$

where the LHS represents the mean acceleration of the fluid element, and ρ on the RHS is its mean density. We now pull the same tricks as in §0.4.1: use the divergence theorem to turn the surface integral into a volume integral, commute the temporal derivative and volume integral on the RHS, invoke the arbitrariness of the actual integration volume V , and finally make good use of the continuity equation (8), to obtain the differential equation for \mathbf{u} :

$$\{\text{eq:NS0}\} \quad \frac{D\mathbf{u}}{Dt} = \frac{1}{\rho} \nabla \cdot \mathbf{s}, \quad [\text{m s}^{-2}]. \quad (20)$$

We now define the *pressure* (units: pascal; $1 \text{ Pa} \equiv 1 \text{ N m}^{-2}$) as the isotropic part of the force acting perpendicularly on the volume’s surfaces, and separate it explicitly from the stress tensor:

$$\mathbf{s} = -p\mathbf{I} + \boldsymbol{\tau}, \quad (21)$$

where \mathbf{I} is the identity tensor, and the minus sign arises from the convention that pressures acts on the bounding surface towards the interior of the volume element, and $\boldsymbol{\tau}$ will presently become the *viscous stress tensor*. Since $\nabla \cdot (p\mathbf{I}) = \nabla p$, eq. (20) becomes

$$\{\text{eq:Navier-Stokes}\} \quad \boxed{\frac{D\mathbf{u}}{Dt} = -\frac{1}{\rho} \nabla p + \frac{1}{\rho} \nabla \cdot \boldsymbol{\tau}}. \quad (22)$$

This is the celebrated *Navier-Stokes equation*. Any additional volumetric body forces (gravity, Lorentz force, etc.) are simply added to the RHS.

The next step is to obtain expressions for the components of the tensor $\boldsymbol{\tau}$. The viscous force, which is what $\boldsymbol{\tau}$ stands for, can be viewed as a form of friction acting between contiguous laminae of fluid moving with different

velocities, so that we expect it to be proportional to velocity *derivatives*. Consider now the following decomposition of a velocity gradient:

$$\frac{\partial u_k}{\partial x_l} = \underbrace{\frac{1}{2} \left(\frac{\partial u_k}{\partial x_l} + \frac{\partial u_l}{\partial x_k} \right)}_{D_{kl}} + \underbrace{\frac{1}{2} \left(\frac{\partial u_k}{\partial x_l} - \frac{\partial u_l}{\partial x_k} \right)}_{\Omega_{kl}}. \quad (23) \quad \{\text{eq:vstress10}\}$$

The first term on the RHS is a *pure shear*, and is described by the (symmetric) *deformation tensor* D_{kl} ; the second is a *pure rotation*, and is described by the antisymmetric *vorticity tensor* Ω_{kl} . It can be shown that the latter causes no deformation of the fluid element, *therefore the viscous force can only involve* D_{kl} . A *Newtonian fluid* is one for which the (tensorial) relation between $\boldsymbol{\tau}$ and D_{kl} is linear:

$$\tau_{ij} = f_{ij}(D_{kl}), \quad i, j, k, l = (1, 2, 3) \equiv (x, y, z) \quad (24) \quad \{\text{eq:Newtonfl}\}$$

Since $\boldsymbol{\tau}$ and \mathbf{D} are both symmetric tensors, this linear relationship can involve up to 36 independent numerical coefficients. The next step is to invoke the invariance of the physical laws embodied in eq. (24) under rotation of the coordinate axes to set some of these coefficients to zero. The mathematics is rather tedious, but wirth the effort because at the end of the day you end up with:

$$\tau_{xx} = 2\mu D_{xx} + (\mu_\vartheta - \frac{2}{3}\mu)(D_{xx} + D_{yy} + D_{zz}) \quad (25)$$

$$\tau_{yy} = 2\mu D_{yy} + (\mu_\vartheta - \frac{2}{3}\mu)(D_{xx} + D_{yy} + D_{zz}) \quad (26)$$

$$\tau_{zz} = 2\mu D_{zz} + (\mu_\vartheta - \frac{2}{3}\mu)(D_{xx} + D_{yy} + D_{zz}) \quad (27)$$

$$\tau_{xy} = 2\mu D_{xy} \quad (28)$$

$$\tau_{yz} = 2\mu D_{yz} \quad (29)$$

$$\tau_{zx} = 2\mu D_{zx} \quad (30) \quad \{\text{eq:transf}\}$$

which now involves only two numerical coefficients, μ and μ_ϑ , known as the *dynamical viscosity* and **bulk viscosity**, respectively. It is often convenient to define a coefficient of *kinematic viscosity* as

$$\nu = \frac{\mu}{\rho}, \quad [\text{m}^2 \text{s}^{-1}]. \quad (31)$$

In an incompressible flow, the terms multiplying μ_ϑ vanish and it is possible to rewrite the Navier-Stokes equation in the simpler form:

$$\frac{D\mathbf{u}}{Dt} = -\frac{1}{\rho}\nabla p + \nu\nabla^2\mathbf{u}. \quad [\text{incompressible}] \quad (32) \quad \{\text{eq:NSincompress}\}$$

Note here the presence of a Laplacian operator acting on a *vector* quantity (here \mathbf{u}); this is only equivalent to the Laplacian acting on the scalar components of \mathbf{u} in the special case of cartesian coordinates.

Incompressible or not, the behavior of viscous flows will often hinge on the relative importance of the advective and dissipative terms in the Navier-Stokes equation:

$$\{\text{eq:NR2}\} \quad \rho(\mathbf{u} \cdot \nabla)\mathbf{u} \quad \leftrightarrow \quad \nabla \cdot \boldsymbol{\tau} . \quad (33)$$

Introducing characteristic length scales u_0 , L , ρ_0 and ν_0 , dimensional analysis yields:

$$\{\text{eq:NR3}\} \quad \rho_0 \frac{u_0^2}{L} \quad \leftrightarrow \quad \frac{1}{L} \rho_0 \nu_0 \frac{u_0}{L} , \quad (34)$$

where we made use of the fact that the viscous stress tensor has dimensions $\mu \times D_{ik}$, with $\mu = \rho\nu$ and the deformation tensor D_{ik} has dimension of velocity per unit length (cf. eq. 23). The ratio of these two terms is a dimensionless quantity called the *Reynolds Number*:

$$\{\text{eq:NR}\} \quad \boxed{\text{Re} = \frac{u_0 L}{\nu_0}} . \quad (35)$$

This measures the importance of viscous forces versus fluid inertia. It is a key dimensionless parameter in hydrodynamics, as it effectively controls fundamental processes such as the transition to turbulence, as well as more mundane matters such as boundary layer thicknesses.

A few words on boundary conditions; in the presence of viscosity, the flow speed must vanish wherever the fluid is in contact with a rigid surface S :

$$\{\text{eq:NSBC}\} \quad \mathbf{u}(\mathbf{x}) = 0 , \quad \mathbf{x} \in S . \quad (36)$$

This remains true even in the limit where the viscosity is vanishingly small. For a *free surface* (e.g., the surface of a fluid sphere floating in a vacuum), the normal components of both the flow speed and viscous stress must vanish instead:

$$\{\text{eq:NSBCfs}\} \quad \mathbf{u} \cdot \hat{\mathbf{n}}(\mathbf{x}) = 0 , \quad \boldsymbol{\tau} \cdot \hat{\mathbf{n}} = 0 , \quad \mathbf{x} \in S . \quad (37)$$

`{pch_ssec:vort}` **0.4.4 Angular momentum: the vorticity equation**

The “rotation” and “angular momentum” of a fluid system cannot simply be reduced to simple scalars such as angular velocity and moment of inertia, because the application of a torque to a fluid element can alter not just its rotation rate, but also its shape and mass distribution. A more useful measure of “rotation” is the *circulation* Γ about some closed contour γ embedded in and moving with the fluid:

$$\{eq:vort1\} \quad \Gamma(t) = \oint_{\gamma} \mathbf{u}(\mathbf{x}, t) \cdot d\boldsymbol{\ell} = \int_S (\nabla \times \mathbf{u}) \cdot \hat{\mathbf{n}} dS = \int_S \boldsymbol{\omega} \cdot \hat{\mathbf{n}} dS, \quad (38)$$

where the second equality follows from Stokes' theorem, and the third from the definition of *vorticity* :

$$\boldsymbol{\omega} = \nabla \times \mathbf{u}. \quad (39) \quad \{eq:vortdef\}$$

Thinking about flows in terms of vorticity $\boldsymbol{\omega}$ rather than speed \mathbf{u} can be useful because of *Kelvin's theorem*, which states that in the inviscid limit $\nu \rightarrow 0$ (or, equivalently, $\text{Re} \rightarrow \infty$), the circulation Γ along any closed loop γ advected by the moving fluid is a conserved quantity:

$$\frac{D\Gamma}{Dt} = 0. \quad (40) \quad \{eq:kelvin11\}$$

Applying again Stokes' theorem yields the equivalent expression

$$\boxed{\frac{D}{Dt} \int_S \boldsymbol{\omega} \cdot \hat{\mathbf{n}} dS = 0}, \quad (41) \quad \{eq:kelvin12\}$$

stating that the flux of vorticity across any material surface S bounded by γ is also a conserved quantity, both in fact being integral expressions of angular momentum conservation.

An evolution equation for $\boldsymbol{\omega}$ can be obtained via the Navier-Stokes equation, in a particularly illuminating manner in the case of an incompressible fluid ($\nabla \cdot \mathbf{u} = 0$) with constant kinematic viscosity ν , in which case eq. (32) can be rewritten as

$$\frac{D\mathbf{u}}{Dt} = -\nabla \left(\frac{p}{\rho} + \Phi \right) - \nu \nabla \times (\nabla \times \mathbf{u}), \quad [\text{incompressible}] \quad (42) \quad \{eq:vorteq1\}$$

where it was assumed that gravity can be expressed as the gradient of a (gravitational) potential. Taking the curl on each side of this expression then yields:

$$\nabla \times \left(\frac{\partial \mathbf{u}}{\partial t} \right) + \nabla \times (\mathbf{u} \cdot \nabla \mathbf{u}) = \underbrace{\nabla \times \left[\nabla \left(\frac{p}{\rho} + \Phi \right) \right]}_{=0} - \nu \nabla \times \nabla \times (\nabla \times \mathbf{u}) \quad (43) \quad \{eq:vorteq3\}$$

then, commuting the time derivative with $\nabla \times$ and making judicious use of some vector identities to develop the second term on the LHS, remembering also that $\nabla \cdot \boldsymbol{\omega} = 0$, eventually leads to:

$$\boxed{\frac{D\boldsymbol{\omega}}{Dt} - \boldsymbol{\omega} \cdot \nabla \mathbf{u} = \nu \nabla^2 \boldsymbol{\omega}}, \quad [\text{incompressible}]. \quad (44) \quad \{eq:vorteq\}$$

This is the *vorticity equation*, expressing in differential form the conservation of the fluid’s angular momentum. A useful vorticity-related quantity is the *kinetic helicity*, defined as

$$\{eq:fluidhel\} \quad h = \mathbf{u} \cdot \boldsymbol{\omega} , \quad (45)$$

which measures the amount of twisting in a flow. This will prove an important concept when investigating magnetic field amplification by fluid flows.

{pch_sec:NSentropy} 0.4.5 Energy: the entropy equation

Omitting to begin with the energy dissipated in heat by viscous friction, the usual accounting of energy flow into and out of a volume element V fixed in space leads to the following differential equation expressing the conservation of the plasma’s *internal energy* per unit mass (e , in units J/kg):

$$\{eq:econs\} \quad \frac{De}{Dt} + (\gamma - 1)e\nabla \cdot \mathbf{u} = \frac{1}{\rho} \nabla \cdot [(\chi + \chi_r)\nabla T] , \quad (46)$$

where for a perfect gas we have

$$\{eq:inte\} \quad e = \frac{1}{\gamma - 1} \frac{p}{\rho} = \frac{1}{\gamma - 1} \frac{kT}{\mu m} , \quad (47)$$

with $\gamma = c_p/c_v$ the ratio of specific heats, and $(\chi + \chi_r)\nabla T$ the heat flux in or out of the fluid element, with χ and χ_r the coefficients of thermal and radiative conductivity, respectively (units: $\text{JK}^{-1}\text{m}^{-1}\text{s}^{-1}$). Equation (46) expresses that any variation of the specific energy in a plasma volume moving with the flow (LHS) is due to heat flowing in or out of the volume by conduction or radiation (here in the diffusion approximation). The “extra” term $\propto \nabla \cdot \mathbf{u}$ on the LHS of eq (46) embodies the work done against (or by) the pressure force in compressing (or letting expand) the volume element.

It is often convenient to rewrite the energy conservation equation in terms of the plasma’s *entropy* $S \propto \rho^{-\gamma}p$, which allows to express eq. (46) in the more compact form:

$$\{eq:difec4\} \quad \boxed{\rho T \frac{DS}{Dt} = \nabla \cdot [(\chi + \chi_r)\nabla T]} , \quad (48)$$

which states, now unambiguously, that any change in the entropy S as one follows a fluid element (LHS) can only be due to heat flowing out of or into the domain by conduction or radiation (RHS).

While this is seldom an important factor in astrophysical flows, in general we must add to the RHS of eq. (48) the heat produced by viscous dissipation (and, as we shall see later, by Ohmic dissipation). This is given by the so-

called (volumetric) *viscous dissipation function*:

$$\{\text{eq:viscenrg3c}\} \quad \phi_\nu = \frac{\mu}{2} \left(\frac{\partial u_i}{\partial x_k} + \frac{\partial u_k}{\partial x_i} - \frac{2}{3} \delta_{ik} \frac{\partial u_s}{\partial x_s} \right)^2 + \mu_\vartheta \left(\frac{\partial u_s}{\partial x_s} \right)^2, \quad [\text{J m}^{-3} \text{s}^{-1}], \quad (49)$$

where summation over repeated indices is implied here. Note that since ϕ_ν is positive definite, its presence on the RHS of eq. (48) can only increase the fluid element's entropy, which makes perfect sense since friction, which is what viscosity is for fluids, is an irreversible process.

For more on classical hydrodynamics, see the references listed in the bibliography at the end of this chapter.

0.5 The magnetohydrodynamical induction equation

{pch_sec:MHD}

Our task is now to generalize the governing equations of hydrodynamics to include the effects of the electric and magnetic fields, and to obtain evolution equations for these two physical quantities. Keep in mind that electrical charge neutrality, as required by MHD, does not imply that the fluid's microscopic constituents are themselves neutral, but rather that positive and negative electrical charges are present in equal numbers in any fluid element.

The starting point, you guess it I hope, is Maxwell's celebrated equations:

$$\nabla \cdot \mathbf{E} = \frac{\rho_e}{\varepsilon_0}, \quad [\text{Gauss' Law}] \quad (50) \quad \{\text{eq:max1}\}$$

$$\nabla \cdot \mathbf{B} = 0, \quad [\text{Anonymous}] \quad (51) \quad \{\text{eq:max2}\}$$

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}, \quad [\text{Faraday's Law}] \quad (52) \quad \{\text{eq:max3}\}$$

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{J} + \mu_0 \varepsilon_0 \frac{\partial \mathbf{E}}{\partial t}, \quad [\text{Ampere/Maxwell's Law}] \quad (53) \quad \{\text{eq:max4}\}$$

where, in the SI system of units, the electric field is measured in N C^{-1} ($\equiv \text{V m}^{-1}$), and the magnetic field² \mathbf{B} in tesla (T). The quantity ρ_e is the electrical charge density (C m^{-3}), and \mathbf{J} is the electrical current density (A m^{-2}). The permittivity ε_0 ($= 8.85 \times 10^{-12} \text{C}^2 \text{N}^{-1} \text{m}^{-2}$ in vacuum) and magnetic permeability μ_0 ($= 4\pi \times 10^{-7} \text{N A}^{-2}$ in vacuum) can be considered

² Strictly speaking, \mathbf{B} should be called the magnetic flux density or somesuch, but on this one we'll stick to common astrophysical usage.

as constants in what follows, since we will not be dealing with polarisable or ferromagnetic substances.

The first step is (with all due respect to the man) to do away altogether with Maxwell's displacement current in eq. (53). This can be justified if the fluid flow is non-relativistic and there are no batteries around being turned on or off, two rather sweeping statements that will be substantiated in §0.7. For the time being we just revert to the original form of Ampère's Law:

$$\{\text{eq:Ampere}\} \quad \nabla \times \mathbf{B} = \mu_0 \mathbf{J} . \quad (54)$$

In general, the application of an electric field \mathbf{E} across an electrically conducting substance will generate an electrical current density \mathbf{J} . Ohm's Law postulates that the relationship between \mathbf{J} and \mathbf{E} is linear:

$$\{\text{eq:Ohmprime}\} \quad \mathbf{J}' = \sigma \mathbf{E}' , \quad (55)$$

where σ is the electrical conductivity (units: $\text{C}^2\text{s}^{-1}\text{m}^{-3}\text{kg}^{-1} \equiv \Omega^{-1}\text{m}^{-1}$, $\Omega \equiv \text{Ohm}$). Here the primes ("prime") are added to emphasize that Ohm's Law is expected to hold in a conducting substance *at rest*. In the context of a fluid moving with velocity \mathbf{u} (relativistic or not), eq. (55) can only be expected to hold in a reference frame comoving with the fluid. So we need to transform eq. (55) to the laboratory (rest) frame. In the non-relativistic limit ($u/c \ll 1$, implying $\gamma \rightarrow 1$), the usual Lorentz transformation for the electrical current density simplifies to $\mathbf{J}' = \mathbf{J}$, and that for the electric field to $\mathbf{E}' = \mathbf{E} + \mathbf{u} \times \mathbf{B}$, so that Ohm's Law takes on the generalized form

$$\{\text{eq:Ohm}\} \quad \boxed{\mathbf{J} = \sigma(\mathbf{E} + \mathbf{u} \times \mathbf{B})} , \quad (56)$$

or, making use of the pre-Maxwellian form of Ampère's Law and reorganizing the terms:

$$\{\text{eq:fullE}\} \quad \mathbf{E} = -\mathbf{u} \times \mathbf{B} + \frac{1}{\mu_0 \sigma} (\nabla \times \mathbf{B}) . \quad (57)$$

We now insert this expression for the electric field into Faraday's Law (52) to obtain the justly famous *magnetohydrodynamical induction equation*:

$$\{\text{eq:induction}\} \quad \boxed{\frac{\partial \mathbf{B}}{\partial t} = \nabla \times (\mathbf{u} \times \mathbf{B} - \eta \nabla \times \mathbf{B})} . \quad (58)$$

where

$$\{\text{eq:magdiff}\} \quad \eta = \frac{1}{\mu_0 \sigma} \quad [\text{m}^2\text{s}^{-1}] \quad (59)$$

is the *magnetic diffusivity*³. The first term on the RHS of eq. (58) represents the inductive action of fluid flowing across a magnetic field, while the second term represents dissipation of the electrical currents sustaining the field.

Keep in mind that any solution of eq. (58) must also satisfy eq. (51) at all times. It can be easily shown (try it!) that if $\nabla \cdot \mathbf{B} = 0$ at some initial time, the form of eq. (58) guarantees that zero divergence will be maintained at all subsequent times⁴.

0.6 Scaling analysis

{pch_sec:scaling}

The evolution of a magnetic field under the action of a prescribed flow \mathbf{u} will depend greatly on whether or not the inductive term on the RHS of eq. (58) dominates the diffusive term. Under what conditions will this be the case? We seek a first (tentative) answer to this question by performing a dimensional analysis of eq. (58); this involves replacing the temporal derivative by $1/\tau$ and the spatial derivatives by $1/\ell$, where τ and ℓ are time and length scales that suitably characterizes the variations of *both* \mathbf{u} and \mathbf{B} :

$$\frac{B}{\tau} = \frac{u_0 B}{\ell} + \frac{\eta B}{\ell^2}, \quad (60) \quad \{\text{eq:scalanal}\}$$

where B and u_0 are a “typical” values for the flow velocity and magnetic field strength over the domain of interest. The ratio of the first to second term on the RHS of eq. (60) is a dimensionless quantity known as the *magnetic Reynolds number*⁵:

$$\boxed{R_m = \frac{u_0 \ell}{\eta}}, \quad (61) \quad \{\text{eq:Rm}\}$$

which measures the relative importance of induction versus dissipation *over length scales of order* ℓ . Note that R_m does not depend on the magnetic field strength, a direct consequence of the linearity (in \mathbf{B}) of the MHD induction equation. Our scaling analysis simply says that in the limit $R_m \gg 1$, induction by the flow dominates the evolution of \mathbf{B} , while in the opposite limit of

³ A note of warning: some MHD textbooks use the symbol “ η ” for the inverse conductivity (units $\Omega \text{ m}$), so that the dissipative term on the RHS of the induction equation retains a μ_0^{-1} prefactor.

⁴ This is true under exact arithmetic; if numerical solutions to eq. (58) are sought, care must be taken to ensure $\nabla \cdot \mathbf{B} = 0$ as the solution is advanced in time.

⁵ Not the structural similarity with the usual viscous Reynolds number defined in §0.4.3, with the magnetic diffusivity η replacing the kinematic viscosity ν in the denominator. Had we not absorbed μ_0 in our definition of η , the magnetic permeability μ_0 would appear in the numerator of the magnetic Reynolds number, which I personally find objectionable.

$R_m \ll 1$, induction makes a negligible contribution and \mathbf{B} simply decays away under the influence of Ohmic dissipation.

One may anticipate great simplifications of magnetohydrodynamics if we operate in either of these limits. If $R_m \ll 1$, only the second term is retained on the RHS of eq. (60), which leads immediately to

$$\{\text{eq:tdiff}\} \quad \tau = \frac{\ell^2}{\eta}, \quad (62)$$

a quantity known as the *magnetic diffusion time*. It measures the time taken for a magnetic field contained in a volume of typical linear dimension ℓ to dissipate and/or diffusively leak out of the volume. Now, for most astrophysical objects, this timescale turns out to be quite large, indeed often larger than the age of the universe! (see Table 2). This is not so much because astrophysical plasmas are such incredibly good electrical conductors, but rather because astrophysical objects tend to be very, very large. The *existence* of solar and stellar magnetic fields is then not really surprising; any large-scale fossil field present in a star's interior upon its arrival on the ZAMS would still be there today at almost its initial strength. The challenge in modeling the solar and stellar magnetic fields is to reproduce the peculiarities of their spatiotemporal variations, most notably the decadal cyclic variations observed in the Sun and solar-type stars.

The opposite limit $R_m \gg 1$, defines the *ideal MHD* limit. Then it is the first term that is retained on the RHS of eq. (60), so that

$$\{\text{eq:ttturnover}\} \quad \tau = \ell/u_0, \quad (63)$$

corresponding to the *turnover time* associated with the flow \mathbf{u} . Note already that under ideal MHD, the only non-trivial (i.e., $\mathbf{u} \neq 0$ and $\mathbf{B} \neq 0$) steady-state ($\partial/\partial t = 0$) solutions of the MHD equation are only possible for field-aligned flows.

Table 2 below lists estimates of the magnetic Reynolds number (and related physical quantities) for the various astrophysical systems considered earlier in Table 1⁶:

The magnetic Reynolds number is clearly huge in all cases, which would suggest that the ideal MHD limit is the one most applicable to all these astrophysical systems. But things are not so simple. From a purely mathematical point of view, taking the limit $R_m \rightarrow \infty$ of the MHD induction equation is

⁶ Choices for length scale ℓ ($\equiv L$) as in Table 1. Velocity estimates correspond to large convective cells (solar interior), granulation (photosphere), solar wind speed (corona and solar wind), and turbulence (molecular clouds and interstellar medium). All these numbers (especially the turbulent velocity estimates) are again very rough, and rounded to the nearest factor of ten. The magnetic diffusivity estimates given for molecular clouds and interstellar medium depend critically on the assumed degree of ionization, and so are also very rough.

{pch_tab2}

Table 2 Properties of some astrophysical objects and flows

System/flow	L [km]	σ [$\Omega^{-1}\text{m}^{-1}$]	η [m^2s^{-1}]	τ [yr]	u [km/s]	R_m
Solar interior	10^6	10^4	100	10^9	0.1	10^9
Solar atmosphere	10^3	10^3	1000	10^2	1	10^6
Solar corona	10^5	10^6	1	10^8	10	10^{12}
Solar wind (1 AU)	10^5	10^4	100	10^8	300	10^{11}
Molecular cloud	10^{14}	10^2	10^4	10^{17}	100	10^{18}
Interstellar medium	10^{16}	10^3	1000	10^{22}	100	10^{21}
Sphere of copper	10^{-3}	10^8	10^{-1}	10^{-7}	—	—

problematic, because the order of the highest spatial derivatives decreases by one. This situation is similar to the behavior of viscous flows at very high Reynolds number: solutions to eq. (58) with $\eta \rightarrow 0$ in general *do not* smoothly tend towards solutions obtained for $\eta = 0$. Moreover, the distinction between the two physical regimes $R_m \ll 1$ and $R_m \gg 1$ is meaningful as long as one can define a suitable R_m for the flow as a whole, which, in turn, requires one to estimate, *a priori*, a length scale ℓ that adequately characterizes the flow and magnetic field at all time and throughout the spatial domain of interest. As we proceed it will become clear that this is not always straightforward, or even possible. Finally, the scaling analysis does away entirely with the geometrical aspects of the problem, by substituting $u_0 B$ for $\mathbf{u} \times \mathbf{B}$; yet there are situations (e.g. a field-aligned flow) where even a very large \mathbf{u} has no inductive effect whatsoever.

0.7 The Lorentz force

{pch_sec:magF}

Getting to eq. (58) was pretty easy because we summarily swept the displacement current under the rug, but it represents only half (in fact the easy half) of our task; we must now investigate the effect of the magnetic field on the flow \mathbf{u} ; and this, it turns out, is the tricky part of the MHD approximation.

You will certainly recall that the *Lorentz force* acting on an electrically charge particle moving at velocity \mathbf{u} in a region of space permeated by electric and magnetic fields is given by

$$\mathbf{f} = q(\mathbf{E} + \mathbf{u} \times \mathbf{B}), \quad [\text{N}]. \quad (64) \quad \{\text{eq:lorentz1}\}$$

where q is the electrical charge. Consider now a volume element ΔV containing many such particles; in the continuum limit, the total force per unit volume (\mathbf{F}) acting on the volume element will be the sum of the forces acting on each individual charged constituents divided by the volume element:

$$\begin{aligned}
\mathbf{F} &= \frac{1}{\Delta V} \sum_k \mathbf{f}_k = \frac{1}{\Delta V} \sum_k q_k (\mathbf{E} + \mathbf{u}_k \times \mathbf{B}) \\
&= \left(\frac{1}{\Delta V} \sum_k q_k \right) \mathbf{E} + \left(\frac{1}{\Delta V} \sum_k q_k \mathbf{u}_k \right) \times \mathbf{B} \\
&= \rho_e \mathbf{E} + \mathbf{J} \times \mathbf{B}, \quad [\text{N m}^{-3}]. \tag{65} \quad \{\text{eq:lorentz2}\}
\end{aligned}$$

where the last equality follows from the usual definition of charge density and electrical current density. At this point you might be tempted to eliminate the term proportional to \mathbf{E} , on the grounds that in MHD we are dealing with a globally neutral plasma, meaning $\rho_e = 0$, therefore $\rho_e \mathbf{E} \equiv 0$ and that's the end of it. That would be way too easy...

Let's begin by taking the divergence on both side of the generalized form of Ohm's Law (eq. (56)). We then make use of Gauss's Law (eq. (51)) to get rid of the $\nabla \cdot \mathbf{E}$ term, and of the charge conservation Law

$$\{\text{eq:qcons}\} \quad \frac{\partial \rho_e}{\partial t} + \nabla \cdot \mathbf{J} = 0 \tag{66}$$

to get rid of the $\nabla \cdot \mathbf{J}$ term. The end result of all this physico-algebraical juggling is the following expression:

$$\{\text{eq:indE}\} \quad \frac{\partial \rho_e}{\partial t} + \frac{\rho_e}{(\varepsilon_0/\sigma)} + \sigma \nabla \cdot (\mathbf{u} \times \mathbf{B}) = 0. \tag{67}$$

The combination ε_0/σ has units of time, and is called the *charge relaxation time*, henceforth denoted τ_e . It is the timescale on which charge separation takes place in a conductor if an electric field is suddenly turned on. For most conductors, this a very small number, of order 10^{-18} s !! This is because the electrical field reacts to the motion of electric charges at the speed of light (in the substance under consideration, which is slower than in a vacuum but still mighty fast). Indeed, in a conducting fluid at rest ($\mathbf{u} = 0$) the above expression integrates readily to

$$\{\text{eq:electrorelax}\} \quad \rho_e(t) = \rho_e(0) \exp(-t/\tau_e), \tag{68}$$

thus the name "relaxation time" for τ_e .

Now let us consider the case of a slowly moving fluid, in the sense that it is moving on a timescale much larger than τ_e ; this means that the induced electrical field will vary on a similar timescale (at best), and therefore the time derivative of ρ_e can be neglected in comparison to the ρ_e/τ_e term in eq. (67), leading to

$$\{\text{eq:rhoemhd}\} \quad \rho_e = \varepsilon_0 \nabla \cdot (\mathbf{u} \times \mathbf{B}). \tag{69}$$

This indicates that a finite charge density can be sustained inside a *moving* conducting fluid. The associated electrostatic force per unit volume, $\rho_e \mathbf{E}$, is

definitely non-zero but turns out to be much smaller than the magnetic force. Indeed, a dimensional analysis of eq. (65), using eq. (69) to estimate ρ_e , gives:

$$\rho_e \mathbf{E} \sim \left(\frac{\varepsilon_0 u B}{\ell} \right) \left(\frac{J}{\sigma} \right) \sim \left(\frac{u \tau_e}{\ell} \right) JB, \quad (70)$$

$$\mathbf{J} \times \mathbf{B} \sim JB, \quad (71)$$

where Ohm's Law was used to express \mathbf{E} in terms of \mathbf{J} , and once again ℓ is a typical length scale characterizing the variations of the flow and magnetic field. The ratio of electrostatic to magnetic forces is thus of order $u \tau_e / \ell$. Now $\tau_e \ll 1$ to start with, and for non-relativistic fluid motion we can expect that the flow's turnover time ℓ / u is much larger than the crossing time for an electromagnetic disturbance $\sim \ell / c \sim \tau_e$; both effects conspire to render the electrostatic force absolutely minuscule compared to the magnetic force, so that eq. (65) becomes

$$\boxed{\mathbf{F} = \mathbf{J} \times \mathbf{B}, \quad [\text{MHD approximation}]}. \quad (72) \quad \{\text{eq:FMHD}\}$$

This must be added to the RHS of the Navier-Stokes equation (22)... with a $1/\rho$ prefactor so we get a force per unit mass, rather than per unit volume.

Now, getting back to this business of having dropped the displacement current in the full Maxwellian form of Ampère's Law (eq. (53)); it can now be all justified on the grounds that the time derivative of the charge density can be neglected in the non-relativistic limit. Indeed, to be consistent the charge conservation equation (66) now reduces to

$$\nabla \cdot \mathbf{J} = 0; \quad (73)$$

taking the divergence on both sides of eq. (53) then leads to

$$\nabla \cdot \mathbf{J} = -\varepsilon_0 \nabla \cdot \left(\frac{\partial \mathbf{E}}{\partial t} \right) = \varepsilon_0 \frac{\partial}{\partial t} (\nabla \cdot \mathbf{E}) = \frac{\partial \rho_e}{\partial t}; \quad (74)$$

this demonstrates that dropping the time derivative of the charge density is equivalent to neglecting Maxwell's displacement current in eq. (53). To sum up, provided we exclude very rapid transient events (such as turning a battery on or off, or any such process which would generate a large $\partial \rho_e / \partial t$), under the MHD approximation the following statements are all equivalent:

- The fluid motions are non-relativistic;
- The electrostatic force can be neglected as compared to the magnetic force;
- Maxwell's displacement current can be neglected.

0.8 Joule heating

{pch_sec:Joule}

In the presence of finite electrical conductivity, the volumetric heating associated with the dissipation of electric currents must be included on the RHS of the energy equation, in the form of the so-called *Joule heating function*:

$$\{\text{eq:Joule}\} \quad \phi_B = \frac{\eta}{\mu_0} (\nabla \times \mathbf{B})^2, \quad [\text{J m}^{-3} \text{s}^{-1}]. \quad (75)$$

Note however that in very nearly all astrophysical circumstances, Joule heating makes an insignificant contribution to the energy budget. When it occurs, heating by magnetic energy dissipation, such as in flares, involves dynamical mechanisms that lead to effective dissipation far more rapid and efficient than Joule heating.

{pch_sec:fullMHD}

0.9 The full set of MHD equations

For the record, we now collect the set of partial differential equations governing the behavior of magnetized fluids in the MHD limit. In anticipation of developments to follow, we write these equations in a frame of reference rotating with angular velocity $\boldsymbol{\Omega}$, with the centrifugal force absorbed within the pressure gradient term:

$$\{\text{eq:fullMHD1}\} \quad \frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{u}) = 0, \quad (76)$$

$$\{\text{eq:fullMHD2}\} \quad \frac{D\mathbf{u}}{Dt} = -\frac{1}{\rho} \nabla p - 2\boldsymbol{\Omega} \times \mathbf{u} + \mathbf{g} + \frac{1}{\mu_0 \rho} (\nabla \times \mathbf{B}) \times \mathbf{B} + \frac{1}{\rho} \nabla \cdot \boldsymbol{\tau}, \quad (77)$$

$$\{\text{eq:fullMHD3}\} \quad \frac{De}{Dt} + (\gamma - 1)e \nabla \cdot \mathbf{u} = \frac{1}{\rho} \left[\nabla \cdot \left((\chi + \chi_r) \nabla T \right) + \phi_\nu + \phi_B \right], \quad (78)$$

$$\{\text{eq:fullMHD4}\} \quad \frac{\partial \mathbf{B}}{\partial t} = \nabla \times (\mathbf{u} \times \mathbf{B} - \eta \nabla \times \mathbf{B}). \quad (79)$$

Equations (76)–(79) are further complemented by the two constraint equations:

$$\{\text{eq:fullMHD5}\} \quad \nabla \cdot \mathbf{B} = 0, \quad (80)$$

$$\{\text{eq:fullMHD6}\} \quad p = f(\rho, T, \dots), \quad (81)$$

and suitable expressions for the viscous stress tensor and for the physical coefficient ν , χ , η , etc. Note that gravity \mathbf{g} is explicitly included on the RHS of (77), that e is the specific internal energy of the plasma (magnetic energy will be dealt with separately shortly), and that eq. (81) is just some generic form for an equation of state linking the pressure to the properties of the plasma such as density, temperature, chemical composition, etc.

This is it in principle, but in what follows we shall seldom solve these equations in this complete form. In the parameter regime characterizing most astrophysical fluids, we usually have $\text{Re} \gg 1$, which means that the $(\mathbf{u} \cdot \nabla \mathbf{u})$ term in eq. (77) will play important role; this, in turn, means turbulence, already in itself an unsolved problem even for unmagnetized fluids. There is also a strong nonlinear coupling between eqs. (77) and eqs. (79), so that the turbulent cascade involves both the flow and magnetic field. Finally, with both $\text{Re} \gg 1$ and $\text{R}_m \gg 1$, astrophysical flows will in general develop structures on length scales very much smaller than that characterizing the system under study, so that even fully numerical solutions of the above set of MHD equations will tax the power of the largest extant massively parallel computers, and will continue to do so in the foreseeable future; which is why judicious geometrical and/or physical simplification remains a key issue in the art of astrophysical magnetohydrodynamics... and will also continue to remain so in the same foreseeable future!

0.10 MHD waves

{pch_sec:MHDwaves}

Although it looks innocuous enough, the magnetic force in the MHD approximation has some rather complex consequences for fluid flows, as we will have ample occasions to verify throughout this course. One particularly intricate aspects relates to the types of *waves* that can be supported in a magnetized fluid; in a classical unmagnetized fluid, one deals primarily with sound waves (pressure acting as a restoring force), gravity waves (gravity acting as restoring force), or Rossby waves (Coriolis as a restoring force). It turns out that the Lorentz force introduces not one, but really two additional restoring forces.

Making judicious use of eqs. (51) and (54), together with some classical vector identities, eq. (72) can be rewritten as

$$\mathbf{F} = \frac{1}{\mu_0} \left[(\mathbf{B} \cdot \nabla) \mathbf{B} - \frac{1}{2} \nabla (\mathbf{B}^2) \right], \quad (82) \quad \{\text{eq:FMHD2}\}$$

where $B^2 \equiv \mathbf{B} \cdot \mathbf{B}$. The first term on the RHS is the *magnetic tension*, and the second the *magnetic pressure*. Fluctuations in magnetic pressure can propagate as a longitudinal wave, much as a sound wave, as depicted on Fig. 4A. In fact, two such *magnetosonic* waves modes actually exist, according to whether

the magnetic pressure fluctuation is in phase with the gas pressure fluctuation (the so-called fast mode), or in antiphase (the slow mode). In addition, magnetic tension can produce a restoring force that allows the propagation of wave-on-a-string-like transverse waves, known as *Alfvén waves*, as illustrated on Fig. 4B. Small-amplitude Alfvén waves travel with a speed u_A given by

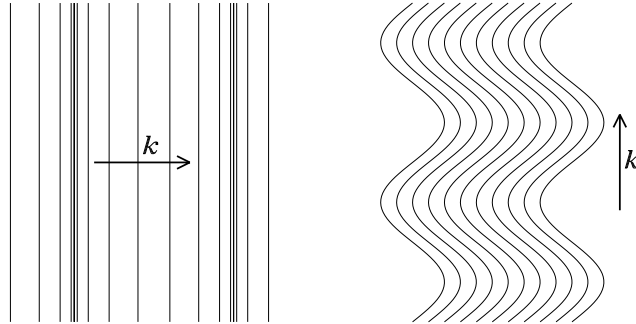


Fig. 4 ^{fig:MHDwaves} The two fundamental MHD wave modes in a uniform background magnetic field: (A) magnetosonic mode, and (B) Alfvén mode. The wave vector \mathbf{k} is indicated as a thick arrow, and highlights the fact that the magnetosonic mode is a longitudinal wave, while the Alfvén mode is a transverse wave. In the presence of plasma, the magnetosonic mode breaks into two submodes, according to the phasing between the magnetic pressure and gas pressure perturbations (see text).

$$\{\text{eq:Alfspeed}\} \quad u_A = \frac{B_0}{\sqrt{\mu_0 \rho}}, \quad (83)$$

where B_0 is the magnitude of the (uniform) magnetic field along which the wave is propagating, and ρ is the (constant) fluid density. Mechanical forcing of a magnetic field permeating a compressible fluid will in general excite all three wave modes. This is an essential aspect of the development and propagation of perturbations in the solar corona and wind, but plays a minor role in solar interior dynamics... unless one happens to be interested in subsurface magnetohelioseismology.

^{pch_sec:magE} 0.11 Magnetic energy

Consider the expression resulting from dotting \mathbf{B} into the induction equation (58), integrating over the spatial domain (V) under consideration, and making judicious use of various well-known vector identities and of Gauss' theorem:

$$\{\text{eq:emag}\} \quad \frac{d}{dt} \int_V \frac{\mathbf{B}^2}{2\mu_0} dV = - \oint_S (\mathbf{S} \cdot \hat{\mathbf{n}}) dS - \int_V (\mathbf{u} \cdot \mathbf{F}) dV - \int_V \sigma^{-1} \mathbf{J}^2 dV, \quad (84)$$

where $\hat{\mathbf{n}}$ is a outward-directed unit vector normal to the boundary surface, and the vector quantity \mathbf{S} is the Poynting flux:

$$\mathbf{S} = \frac{1}{\mu_0} \mathbf{E} \times \mathbf{B}. \quad (85) \quad \{\text{E2.111}\}$$

Examine now the three terms on the RHS of eq. (84); the first is the Poynting flux component into the domain, integrated over the domain boundaries, i.e., the flux of electromagnetic energy in (integrand < 0) or out (integrand > 0) of the domain. This term evidently vanishes in the absence of applied magnetic or electric fields on the boundaries. The second is the work done by the Lorentz force (\mathbf{F}) on the flow. In general this term can be either positive or negative; in the dynamo context we are interested in the $\mathbf{u} \cdot \mathbf{F} < 0$ situation, where the flow transfers energy to the magnetic field. The third term is evidently always negative, and represents the rate of energy loss due to Ohmic dissipation. Equations (84) then naturally leads to interpret the quantity $\mathbf{B}^2/2\mu_0$ as the magnetic energy density, and the total *magnetic energy* (\mathcal{E}_B) within the domain is:

$$\boxed{\mathcal{E}_B = \frac{1}{2\mu_0} \int_V \mathbf{B}^2 dV}. \quad (86) \quad \{\text{E2.217}\}$$

The MHD dynamos that will be the focus of much of what follows are fluid systems that convert mechanical energy into magnetic energy, through the agency of the $\mathbf{u} \cdot \mathbf{F}$ term on the RHS of eq. (84).

0.12 Magnetic flux freezing and Alfvén's theorem

{pch_sec:freeze}

Let us return to the differential form of Faraday's Law:

$$\nabla \times \mathbf{E} = - \frac{\partial \mathbf{B}}{\partial t}. \quad (87)$$

Project now each side of this expression onto a unit vector normal to some surface S fixed in space and bounded by a closed contour γ , integrate over S , and apply Stokes' theorem to the LHS:

$$\int_S (\nabla \times \mathbf{E}) \cdot \hat{\mathbf{n}} dS = \oint_\gamma \mathbf{E} \cdot d\boldsymbol{\ell} = - \int_S \left(\frac{\partial \mathbf{B}}{\partial t} \right) \cdot \hat{\mathbf{n}} dS. \quad (88)$$

So far the surface S remains completely arbitrary. If it is fixed in space, then we get the usual integral form of Faraday's Law:

$$\oint_{\gamma} \mathbf{E} \cdot d\boldsymbol{\ell} = -\frac{\partial}{\partial t} \int_S \mathbf{B} \cdot \hat{\mathbf{n}} dS, \quad (89) \quad \{\text{eq:almost}\}$$

with the LHS corresponding to the electromotive force, and the RHS to the time variation of the magnetic flux (Φ_B). If we now assume instead that the surface S is a material surface moving with the fluid, then (1) we must substitute the Lagrangian operator D/Dt for the partial derivative on the RHS of eq. (89); and (2) we are allowed to invoke Ohm's Law to substitute \mathbf{J}/σ for \mathbf{E} on the RHS since any point of the (material) contour is by definition co-moving with the fluid:

$$\frac{1}{\sigma} \oint_{\gamma} \mathbf{J} \cdot d\boldsymbol{\ell} = -\frac{D}{Dt} \int_S \mathbf{B} \cdot \hat{\mathbf{n}} dS. \quad (90)$$

Now, obviously, in the limit of infinite conductivity we have

$$\{\text{eq:alfven}\} \quad \frac{D}{Dt} \int_S \mathbf{B} \cdot \hat{\mathbf{n}} dS = 0. \quad (91)$$

This states that in the ideal MHD limit $\sigma \rightarrow \infty$, the magnetic flux threading any (open) surface is a conserved quantity as the surface is advected (and possibly deformed) by the flow. This result is known as *Alfvén's theorem*.

Note in particular that in the limit of an infinitesimal surface pierced by “only one” fieldline, Alfvén's theorem is equivalent to saying that magnetic fieldlines must move in the same way as fluid elements; it is customary to say that the magnetic flux is *frozen* into the fluid. In this manner it behaves just like vorticity in the inviscid limit $\nu \rightarrow 0$. And like in the case of vorticity, sheared flows can amplify magnetic fields by stretching, a subject we will investigate in detail in the following chapter.

Alfvén's theorem can be arrived at in a different way, upon noting that the magnetic field is a solenoidal vector, in that $\nabla \cdot \mathbf{B} = 0$; any such vector transported by a flow \mathbf{u} is subjected to the so-called kinematic theorem, stating that:

$$\{\text{eq:kinth2}\} \quad \frac{D}{Dt} \int_{S_m} \mathbf{B} \cdot \hat{\mathbf{n}} dS = \int_{S_m} \left[\frac{\partial \mathbf{B}}{\partial t} - \nabla \times (\mathbf{u} \times \mathbf{B}) \right] \cdot \hat{\mathbf{n}} dS. \quad (92)$$

Now in the ideal limit, the RHS is zero as per our MHD induction equation (58) with $\eta = 0$, and the LHS is just the magnetic flux threading the material surface S_m , so there we have it.

0.13 The magnetic vector potential

It will often prove useful to work with the MHD induction equation written in terms of a *vector potential*, \mathbf{A} (units T m), such that $\mathbf{B} = \nabla \times \mathbf{A}$. Equation

(58) is then readily integrated to

$$\{E2.02\} \quad \frac{\partial \mathbf{A}}{\partial t} = \mathbf{u} \times (\nabla \times \mathbf{A}) - \eta \nabla \times (\nabla \times \mathbf{A}) + \nabla \varphi, \quad (93)$$

where, in “uncurling” the induction equation we may elect to append the gradient of a scalar function to the RHS, with no effect on \mathbf{B} . This additional term may contribute to the electric field \mathbf{E} , however, and so φ is conveniently regarded as the electrostatic potential. We will often pick the Coulomb gauge $\varphi = \nabla \cdot \mathbf{A}$ in order to express the Ohmic dissipation term simply as $\eta \nabla^2 \mathbf{A}$.

0.14 Magnetic helicity

{pch_sec:helicity}

In analogy with fluid helicity, one can define the *magnetic helicity* as

$$h_B = \mathbf{A} \cdot \mathbf{B} . \quad (94)$$

Consider now the variation of the total magnetic helicity (\mathcal{H}_B) in a co-moving fluid volume V ; making judicious use of eqs. (58), (93) and (54), a good deal of vector algebra eventually leads to the following evolution equation for (\mathcal{H}_B):

$$\frac{D}{Dt} \underbrace{\int_V \mathbf{A} \cdot \mathbf{B} dV}_{\mathcal{H}_B} = -2\mu_0 \eta \underbrace{\int_V \mathbf{J} \cdot \mathbf{B} dV}_{\mathcal{H}_J}, \quad (95) \quad \{eq:maghel\}$$

where the integral on the RHS defines the total **current helicity** \mathcal{H}_J , which measure the topological linkage between magnetic fieldlines and electrical currents within the volume, much like the way in which the total magnetic helicity \mathcal{H}_B measures the linkage of magnetic flux systems within V .

Equation (95) indicates that in the ideal MHD limit, magnetic helicity becomes a conserved quantity. This will turn out to pose a severe constraint on magnetic field amplification in astrophysical dynamos, an issue to which we will return in due time.

0.15 Force-free magnetic fields

{pch_sec:ffB}

In many astrophysical systems, the magnetic field dominates the dynamics and energetics of the system. Left to itself, such a system would tend to evolve to a force-free state described by

$$\mathbf{F} = \mathbf{J} \times \mathbf{B} = 0 . \quad (96)$$

Broadly speaking, this can be achieved in two physically distinct ways (excluding the trivial solution $\mathbf{B} = 0$). The first is $\mathbf{J} = 0$ throughout the system. Then Ampère’s Law becomes $\nabla \times \mathbf{B} = 0$, which means that, as with the electric field in electrostatic, \mathbf{B} can be expressed as the gradient of a potential. Such a magnetic field is called a *potential field*. Substitution into $\nabla \cdot \mathbf{B} = 0$ then yields a Laplace-type problem:

$$\{\text{eq:potB}\} \quad \mathbf{B} = \nabla\varphi, \quad \nabla^2\varphi = 0, \quad [\text{Potential field}] . \quad (97)$$

Alternately, a system including a non-zero current density can still be force free, provided the currents flow everywhere parallel to the magnetic field, i.e.,

$$\{\text{eq:alphaB}\} \quad \nabla \times \mathbf{B} = \alpha\mathbf{B}, \quad (98)$$

where α need not necessarily be a constant, i.e., it can vary from one fieldline to another, vary in space, and even depend on the (local) value of \mathbf{B} . Imagine now a situation where, in some domain (for example, the exterior of a star), we are provided with a boundary condition on \mathbf{B} and the task is to construct a force-free field. Adopting the potential field *Ansatz* can lead to very different reconstructions than if we adopt instead eq. (98), given that in the latter case one is free to specify any electric current distribution within the domain, as long as \mathbf{J} remains parallel to \mathbf{B} .

A very important result in this context is known as *Aly’s Theorem*; it states that in a semi-infinite domain with \mathbf{B}_\perp imposed at the boundary and $\mathbf{B} \rightarrow 0$ as $\mathbf{x} \rightarrow \infty$, the (unique) potential field solution satisfying the boundary conditions has a magnetic energy that is *lower* than any of the (multiple) solutions of eq. (98) that satisfy the same boundary conditions, even with complete freedom to specify $\alpha(\mathbf{x})$ within the domain. This poses a strict limit to the amount of magnetic energy stored into a system that can actually be tapped into to power astrophysically interesting phenomena.

0.16 The ultimate origin of astrophysical magnetic fields

`{pch_sec:origB}`

`{pch_ssec:BvE}` 0.16.1 Why \mathbf{B} and not \mathbf{E} ?

Pretty much anywhere we look in the known universe, there are magnetic fields of all strengths and shapes everywhere; but electric fields are conspicuously absent. Why is that? You might think, looking at Maxwell’s equations (50)–(53) that \mathbf{E} and \mathbf{B} appear therein on apparently equal footing, leaving nothing to allow us to anticipate the observed astrophysical preponderance of magnetic fields over electrical fields. Moreover, one observer’s magnetic field

can be turned into another's electric field by a simple change of reference frame. So what's the deal here?

Well, for one thing if you use any sort of sensible “rest frame” for astronomical observation (Earth at rest; solar system at rest; Milky Way at rest; local group at rest; etc *ad infinitum*) there is a lot of \mathbf{B} around and precious little \mathbf{E} . The crucial difference between \mathbf{E} and \mathbf{B} in Maxwell's equations is not the fields themselves, nor the reference frame in which they are measured, but their *sources*. The Universe may be largely empty, but the fact is that it contains a whopping number of electrically charged particles of various sorts (free electrons, ionized atoms or molecules, photoelectrically charged dust grains, etc). If a large-scale electric field were suddenly to be turned on, all these charges will do the honorable thing, which is to separate along the electric field direction until the secondary electric field so produced cancels the externally applied electric field, at which point charge separation ceases.

Moreover, the low densities of most astrophysical plasmas lead to very large mean-free paths for microscopic constituents, leading in turn to fairly good electrical conductivities and very short electrostatic relaxation times τ_e (see eq. (68)), even when the ionisation fraction is quite low (such as in molecular clouds). In other words, astrophysical electric fields, if and whenever they appear, get shortcircuited mighty fast.

Not so with magnetic fields. For starters, as far as anyone can tell there are no magnetic monopoles out there (well, maybe just one, of primordial origin... more on this shortly), so shortcircuiting the magnetic field by monopole separation is out of the question. Magnetic fields, left to themselves, will simply decay as the electrical currents that support them (remember Ampère's Law) suffer Ohmic dissipation. We already obtained a timescale for this process given by eq. (62), and we already noted, on the basis of the compilation presented in Table 2, that this timescale is extremely large, often exceeding the age of the universe. Once magnetic fields are produced, by whatever means, they stick around for a long, long time. But when and how do they first appear? If we remain within the realm of MHD, then we immediately hit a Big Problem, arising from the linearity of the MHD induction equation (58): if $\mathbf{B} = 0$ at some time t_0 then $\mathbf{B} = 0$ at all subsequent times $t > t_0$, a problem that persists unabated as t_0 is pushed all the way back to the Big Bang. We need something else.

0.16.2 Monopoles and batteries

{pch_ssec:origB}

In subsequent chapters we will see that astrophysical flows are actually quite apt at amplifying magnetic fields, so what we are after here is a very small *seed field* to start up the process. Cheap and easy explanations along the line of an original seed magnetic field being a primordial relic of the Big Bang need not concern us here. Nor is early-universe ferromagnetism a viable option, since permanent magnets require an externally-applied magnetic field

to become magnetized in the first place. Interestingly, the two options that are currently deemed viable stand at the opposite ends of the physical exotism scale: magnetic monopoles... and batteries⁷.

Already back in 1931 Paul Dirac (1902-1984) pointed out that there is nothing to prevent there being magnetic monopoles so long as the magnetic charge on a particle is some integer multiple of $g \equiv hc/(4\pi e) \approx 69e$, where h is Planck's constant, and e is the fundamental electric charge. With just *one* magnetic monopole in the universe we have our basic seed field. Some Grand Unified (field) Theories “predict” that very early in the formation of the universe a lot of $m_g \approx 10^{16}$ GeV/ c^2 magnetic monopoles should be produced, to the point that an inflationary cosmological scenarios are needed to ensure that only a few such massive monopoles end up within each inflated subdomain.

In light of the fact that no one has yet seen a magnetic monopole, it would be wise to find a more pedestrian means to create seed magnetic fields, relying on basic physics that we know functions sensibly at least in our part of the universe. To this end, we return to our derivation of the induction equation (§0.5). Recall that one essential step toward MHD from Maxwell required stipulating Ohm's law, in the form of eq. (56) for the laboratory frame of reference. Consider now the possibility of a “mechanically-driven” process of charge separation (i.e., not related to the presence of an electric field in any reference frame); Ohm's law then picks up an extra term:

$$\{\text{eq:bier1}\} \quad \mathbf{J} = \sigma \left[\mathbf{E} + \mathbf{u} \times \mathbf{B} \right] + \mathbf{J}_{\text{mech}} . \quad (99)$$

If we keep only the very first term on the RHS of equation (99), *and* drop the displacement current in equation (53), then we get back to the induction equation (58). If we avail ourselves of neither of these opportunities then we obtain instead:

$$\{\text{eq:bier2}\} \quad \left\{ 1 + \frac{\eta}{c^2} \frac{\partial}{\partial t} \right\} \frac{\partial \mathbf{B}}{\partial t} = \nabla \times \left(\mathbf{u} \times \mathbf{B} - \eta \nabla \times \mathbf{B} + \mu_0 \mathbf{J}_{\text{mech}} \right) . \quad (100)$$

Notice that our only hope for creating \mathbf{B} out of nothing (so to speak) is the \mathbf{J}_{mech} term; retaining the displacement current gives us no advantage.

The \mathbf{J}_{mech} term represents our ability to mechanically grab a hold of electric charges and force currents to flow; in other words, an electromotive force. In the dense interior of a conducting star, plasma kinetic theory permits one to write down a prescription for this “battery” contribution to the total electric current density as:

$$\{\text{eq:bier3}\} \quad \mathbf{J}_{\text{mech}} = \frac{\sigma}{en_e} \left[\nabla p_e - \frac{1}{c} \mathbf{J} \times \mathbf{B} \right] , \quad (101)$$

⁷ This section is adapted from class notes written by Thomas J. Bogdan for the graduate class APAS7500 we co-taught in 1997 at CU Boulder.

where p_e is the contribution of the electrons alone to the thermal pressure (see references in bibliography). For a completely ionized pure hydrogen plasma, p_e is just half of the total gas pressure, and $n_e = \rho/m_p$, and so,

$$\mathbf{J}_{\text{mech}} = \frac{\sigma m_p}{2e\rho} \left[\nabla p - \frac{2}{c} \mathbf{J} \times \mathbf{B} \right]. \quad (102) \quad \{\text{eq:bier4}\}$$

Now, the second term on the RHS of equation (102) does not do us any good since it carries a factor of \mathbf{B} , so the whole plan rests upon the first term generating a seed magnetic field. For a spherically symmetric star, we know from hydrostatic equilibrium that $\nabla\Phi = (\nabla p)/\rho$, and so the product $\eta\mathbf{J}_{\text{mech}} \propto \nabla\Phi$. This will not work because of the curl operator on the RHS of equation (100) will yield zero upon acting on \mathbf{J}_{mech} since (∇p) is a gradient of a scalar function. How can we get around this constraint? A viable possibility is rotation. If a star is rotating, then there is a centrifugal force per unit density of $s\Omega^2\hat{\mathbf{e}}_s$ which adds to $\nabla\Phi$ and which leads to the generation of a seed magnetic field. This process of the centrifugal force driving a flow of electrons relative to the ions was first pointed out by Ludwig Biermann (1907–1986) and is now called the *Biermann battery*.

In fact *any* process that can produce a relative motion between the ions and electrons is a potential battery mechanism, and a possible candidate for creating seed magnetic fields. For example, consider a rotating proto-galaxy, where the outer portions of the proto-galaxy move at a speed $U = R\Omega$ relative to the frame in which the microwave background is isotropic. The Thomson scattering of the microwave photons by the electrons results in the so-called *Compton drag effect*, which causes the electrons to counter-rotate with respect to the ions. The net result is an azimuthal current which generates a dipole-like magnetic field.

Of course, if you bother to put typical numbers in these various examples you will find that you don't really generate very *much* magnetic field. But generating a lot of field is not the point, that can be done via magnetic flux conservation in a collapsing protostellar cloud, or, as we shall see in due time, via the $\mathbf{u} \times \mathbf{B}$ term in our MHD induction equation. The basic idea to take away from this section is that invoking weird, unproven physics to get away from $\mathbf{B} = 0$ is not necessary.

0.17 The astrophysical dynamo problem(s)

{pch_sec:dynprob}

Before moving on with astrophysical dynamos, it will prove instructive to first consider the following example of a simple laboratory dynamo, which illustrates nicely how the idea of amplifying magnetic field by bodily moving electrical charges across a magnetic field is not so mysterious as one may initially think.

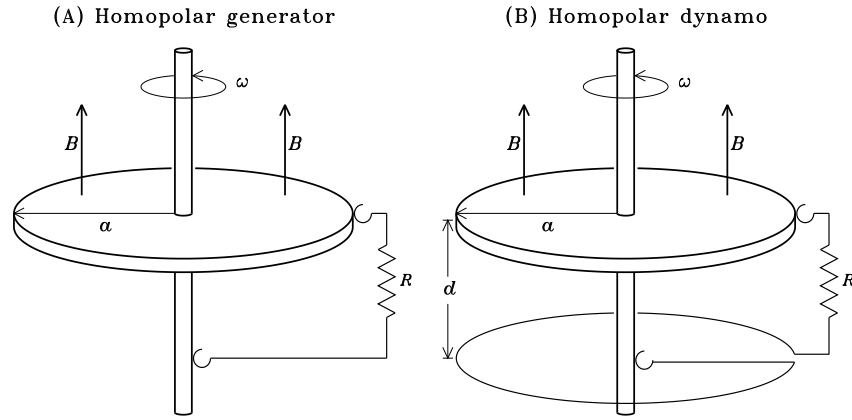


Fig. 5 ^{fig:homopolar} A homopolar generator (A) versus a homopolar dynamo (B). An external magnetic field B is applied across a rotating conducting disk, producing an electromotive force that drives a radial current, a wire connecting the edge of the disk to the axle, forming a circuit of resistance R . The only difference between the two electro-mechanical devices illustrated here is that in the latter case, the wire completing the circuit by connecting on the axle is wrapped into a loop in a plane parallel to the disk, so that a secondary vertical magnetic field is produced (see text).

0.17.1 A simple dynamo

{pch_ssec:homopolar}

One of the many practical inventions of Michael Faraday (1791–1867) was a DC electric current generator based on the rotation of a conducting metallic disk threaded by an external magnetic field. Figure 5(A) illustrates the basic design: a circular disk of radius a mounted on an axle, rotating at angular velocity ω through the agency of some external mechanical force (e.g., Faraday turning a crank). A vertical magnetic field is imposed across the disk. Electrical charges in the disk will feel the usual Lorentz force $\mathbf{F} = q\mathbf{v} \times \mathbf{B}$ where, (initially) \mathbf{v} is just the motion imposed by the rotation of the disk. Working in cylindrical coordinates (s, ϕ, z) one can write

$$\{\text{eq:homo2a}\} \quad \mathbf{v} = (\omega s)\hat{e}_\phi, \quad \mathbf{B} = B_0\hat{e}_z, \quad (103)$$

so that

$$\{\text{eq:homo3}\} \quad \mathbf{F} = (q\omega s B_0)\hat{e}_s. \quad (104)$$

Now consider the circuit formed by connecting the edge of the disk to the base of the axle via frictionless sliding contacts. With the lower part of the circuit away from the imposed magnetic field, the only portion of the circuit where the magnetic force acts on the charges is within the disk, amounting to an electromotive force

$$\{\text{eq:homo4}\} \quad \mathcal{E} = \oint_{\text{circuit}} \left(\frac{\mathbf{F}}{q} \right) \cdot d\boldsymbol{\ell} = \int_0^a \omega B_0 s ds = \frac{\omega B_0 a^2}{2} . \quad (105)$$

Neglecting for the time being the self-inductance of the circuit, the current flowing through the resistor is simply given by $I = \mathcal{E}/R$. This device is called a *homopolar generator*.

There is a subtle modification to this setup that can turn this generator into a *homopolar dynamo*, namely a device that converts mechanical energy into self-amplifying electrical currents and magnetic fields. Instead of simply connecting the resistor straight to the axle as on 5(A), the wire is wrapped around the axle in a loop lying in a plane parallel to the disk, and then connected to the axle, as shown on 5(B). Use your right-hand rule to convince yourself that this current loop will now produce a secondary magnetic field B_* that will superpose itself on the external field B_0 . The magnetic flux through the disk associated with this secondary field will be proportional to the current flowing in the wire loop, the proportionality constant being defined as the inductance (M):

$$MI = \Phi = \pi a^2 B_* , \quad (106) \quad \{\text{eq:homo5}\}$$

where the second equality comes from assuming that the secondary field is vertical and constant across the disk; but what really matters here is that $B_* \propto I$ since the geometry is fixed. We now write an equation for the electrical current, this time taking into consideration the counter-electromotive force associated with self-inductance of the circuit since \mathbf{B} , and therefore also I , are time-varying:

$$\mathcal{E} - L \frac{dI}{dt} = RI \quad (107) \quad \{\text{eq:homo6}\}$$

where L is the coefficient of self-inductance. Substituting eqs. (105) and (106) into this expression, leads to

$$L \frac{dI}{dt} = \frac{\omega a^2}{2} \left(B_0 + \frac{MI}{\pi a^2} \right) - RI \quad (108) \quad \{\text{eq:homo7}\}$$

indicating that the current –and thus the magnetic field– will grow provided that initially,

$$\frac{\omega a^2 B_0}{2} > RI . \quad (109) \quad \{\text{eq:homo8}\}$$

which it certainly will at first since $I = 0$ at $t = 0$. There will eventually come a time (t_*) when the secondary magnetic field will be comparable in strength to the externally applied field B_0 , at which point we may as well “disconnect” B_0 ; eq. (108) then becomes

$$L \frac{dI}{dt} = \left(\frac{\omega M}{2\pi} - R \right) I, \quad (110) \quad \{\text{eq:homo9}\}$$

which integrates to

$$\{\text{eq:homo10}\} \quad I(t) = I(t_*) \exp \left[\frac{1}{L} \left(\frac{\omega M}{2\pi} - R \right) t \right]. \quad (111)$$

indicating that the current —and magnetic field— will grow provided the externally-imposed angular velocity exceeds a critical value:

$$\{\text{eq:homo11}\} \quad \omega > \omega_c = \frac{2\pi R}{M}. \quad (112)$$

This is not a (dreaded) case of perpetual motion, or creating energy out of nothing, or anything like that. The energy content of the growing magnetic field ultimately comes from the biceps of the poor experimenter working ever harder and harder to turn the crank and keep the angular velocity ω at a constant value.

There are many features of this dynamo system worth noting, and which all find their equivalent in the MHD dynamos to be studied in chapters to follow:

1. There exist a critical angular velocity that must be reached for the self-inductance to beat Ohmic dissipation in the resistor, leading to an exponential growth of the magnetic field; below this critical value, the field decays away exponentially once the initial field B_0 is removed.
2. Not all circuits connecting the edge of the disk to the axle will operate in this way; if we suddenly reverse the rotation of the disk, or wrap the wire the other way around the axle, the magnetic field produced by the loop will *oppose* the applied field;
3. The externally applied magnetic field B_0 is only needed as a *seed field* to initiate the amplification process.
4. The homopolar dynamo is really nothing more than a device turning mechanical energy into electromagnetic energy, more specifically magnetic energy.

{pch_ssec:challenge} 0.17.2 The challenges

Copper wires and sliding contacts being a rather sparse commodity in the universe, we must now figure out to apply the general idea of a dynamo to astrophysical fluids. In the MHD limit, our hope lies evidently with the induction term $\nabla \times (\mathbf{u} \times \mathbf{B})$ in the induction equation (58).

In its simplest form, the *dynamo problem* consists in finding a flow field \mathbf{u} that can sustain a magnetic field against Ohmic dissipation. We will encounter in the following chapter flows that can amplify a magnetic field during

a transient time interval, after which \mathbf{B} decays again. So we tighten our definition of the dynamo problem by demanding that a flow be a dynamo if it can lead to $\mathcal{E}_B > 0$ for times much larger than all relevant advective and diffusive timescales of the problem. To make things even harder, we'll add the additional condition that no electromagnetic energy be supplied across the domain boundaries i.e., $\mathbf{S} \cdot \mathbf{n} = 0$ in eq. (84).

We must distinguish the *kinematic dynamo problem*, where the flow field \mathbf{u} is considered given *a priori* and constructed without any regards for its underlying dynamics, from what can only be called (for lack of a generally agreed-upon terminology) the *full dynamo problem*, in which the flow \mathbf{u} results from a solution of the full set of MHD equations (§0.9), including the backreaction of the magnetic field on the flow via the Lorentz force term $\mathbf{J} \times \mathbf{B}$ on the RHS of the Navier-Stokes equation. The kinematic regime carries the immense practical advantage that the induction equation then becomes truly linear in \mathbf{B} , and the dynamo problem reduces to finding a (smooth) flow field \mathbf{u} that has the requisite topological properties to lead to field amplification. In the following chapters we will concentrate mostly on this kinematic regime, but will occasionally touch upon the much more difficult dynamical problem.

The *solar dynamo problem* can be tackled either in kinematic or fully dynamical form. The aim there is to reproduce observed spatiotemporal patterns of solar (and stellar) magnetic field evolution, including things like cyclic polarity reversals, equatorward migration of activity belts, relative strengths and phase relationships between poloidal and toroidal component, etc. This will prove to be a very tall order. Yet, from solar irradiance variations and their possible influence on Earth's climate, space weather prediction, and the understanding of stellar magnetic fields, it all begins with the solar cycle. Keep this in mind as we now start to dig into the mathematical and physical intricacies of magnetic field generation in electrically conducting fluids. We'll seem to venture pretty far away from the sun and stars at times, but stick to it and you'll see it all fitting together at the end. And now, into the abyss...

Bibliography:

There are a great many textbooks available on classical hydrodynamics. The following are my own top-three personal favorites:

- Tritton, D.J., *Physical Fluid Dynamics*, 2nd ed., Oxford University Press (1988),
- Acheson, D.J., *Elementary Fluid Dynamics*, Clarendon Press (1990)
- Landau, L., et Lifschitz, E., *Fluid Mechanics*, Oxford: Pergamon Press (1959).

If you need a refresher on undergraduate electromagnetism, I would recommend

Griffith, D.J., *Introduction to Electrodynamics*, 3rd ed., Prentice Hall (1999).

At the graduate level, the standard reference remains

Jackson, J.D., *Classical Electrodynamics*, 2nd ed., John Wiley & Sons (1975),

who does devote a chapter to magnetohydrodynamics, including a discussion of magnetic wave modes. My personal favorite on magnetohydrodynamics is:

Davidson, P.A., *An Introduction to Magnetohydrodynamics*, Cambridge University Press (2001).

Sections 0.7 and 0.12 are strongly inspired by Davidson's own presentation of the subject. He also presents an illuminating proof of the kinematic theorem embodied in eq. (92). The following textbook is also well worth consulting:

Goedbloed, H., & Poedts, S., *Principles of Magnetohydrodynamics*, Cambridge University Press (2004).

These authors put greater emphasis on MHD waves, shocks, and on the intersection of MHD and plasma physics. For those seeking even more focus on plasma physics aspects, I would recommend:

Kulsrud, R.M., *Plasma Physics for Astrophysics*, Princeton University Press (2005).

Also noteworthy in the general astrophysical context:

Choudhuri, A.R., *The Physics of Fluids and Plasmas*, Cambridge University Press (1998).

On Aly's theorem, see

Aly, J.-J., *Astrophys. J. Lett.*, **375**, L61-L64 (1991),
Smith, D.F., & Low, B.C., *Astrophys. J.*, **410**, 412-425 (1993),

but brace yourself for some serious math. Many ambitious monographs have been written on the general topic of astrophysical magnetic fields. My personal "top-three" selection is:

Parker, E.N., *Cosmic magnetic fields*, Oxford: Clarendon (1979),
Mestel, L., *Stellar magnetism*, Oxford: Clarendon (1999),
Rüdiger, G., & Hollerbach, R., *The magnetic universe*, New York: John Wiley (2004).

Parker's book is unfortunately out of print, and Rüdiger & Hollerbach's outrageously priced. The Mestel book was issued in paperback in 2003, but is still quite pricey.

For some "light" reading on magnetic monopoles in field theory and astrophysics, try,

- Dirac, P.A.M., *Proc. R. Soc. Lond. A*, **133**, 60-72 (1931)
Parker, E.N., *Astrophys. J.*, **160**, 383-404 (1970)
Cabrerera, B., *Phys. Rev. Lett.*, **48**, 1378-1384 (1982)
Kolb, E.W., & Turner, M.S., *The Early Universe*, (New York: Addison-Wesley), §7.6 (1990).

and references therein. For more on Biermann's battery, see

- Biermann, L., *Zeits. f. Naturforsch. A*, **5**, 65 (1950),
Roxburgh, I.W., *Mon. Not. Roy. Astron. Soc.*, **132**, 201-215 (1966),
Chakrabarti, S.K., Rosner, R., & Vainshtein, S.I., *Nature*, **368**, 434-436 (1994),

as well as chap. 13 in the Kulsrud book cited earlier.

Chapter 2

Decay and Amplification of Magnetic Fields

{pch_chap:evolB}

*It's not whether a thing is hard to understand.
It's whether, once understood, it makes any sense.*

*Hans Zinsser
Rats, Lice and History (1934)*

We now begin our long modelling journey towards astrophysical dynamos. This chapter concentrate for the most part on a series of (relatively) simple model problems illustrating the myriad of manners in which a flow and a magnetic field can interact. We first consider the purely resistive decay of magnetic fields (§0.18), then examine various circumstances under which stretching and shearing by a flow can amplify a magnetic field (§0.19). This is followed by a deeper look at some important subtleties of these processes in the context of some (relatively) simple 2D flows (§0.20). We then move on to the so-called anti-dynamo theorems (§0.21), which will shed light on results from previous sections and indicate the way towards dynamo action, which we will finally encounter in §§0.22 and 0.23.

Some of the material contained in this chapter may feel pretty far remote from the realm of astrophysics at times, but please do stick to it because the physical insight (hopefully) developed in the following sections will prove essential to pretty much everything that will come next.

0.18 Resistive decays of magnetic fields

{pch_sec:bdifuz}

Before we try to come up with flows leading to field amplification and dynamo action, we better understand the enemy, namely magnetic field decay by Ohmic dissipation. Consequently, and with the sun and stars in mind, we first consider the evolution of magnetic fields in a sphere (radius R) of electrically conducting fluid, in the absence of any fluid motion (or, more generally, in

the $R_m \ll 1$ limit). The induction equation then reduces to

$$\frac{\partial \mathbf{B}}{\partial t} = -\nabla \times (\eta \nabla \times \mathbf{B}) = \eta \nabla^2 \mathbf{B} - (\nabla \eta) \times (\nabla \times \mathbf{B}). \quad (113) \quad \{\text{E2.12}\}$$

Were it not that we are dealing here with a vector—as opposed to scalar—quantity, for constant η this would look just like a simple heat diffusion equation, with η playing the role of thermal diffusivity. Our derivation of the magnetic energy equation (84) already indicates that under such circumstances, the field can only decay. Back in chapter 1 we already obtained an order-of-magnitude estimate for the timescale $\tau_\eta \sim \ell^2/\eta$ over which a magnetic field \mathbf{B} with typical length scale ℓ can be expected to resistively decay, which in the case of the stellar interiors ended up at $\sim 10^{10}$ yr, i.e., about the main-sequence lifetime of the Sun. Let’s now validate this estimate by securing formal solutions to the diffusive decay problem.

0.18.1 Axisymmetric magnetic fields

Without any significant loss of generality, we can focus on *axisymmetric* magnetic fields, i.e., fields showing symmetry with respect to an axis, usually rotational. Working in spherical polar coordinates (r, θ, ϕ) with the polar axis coinciding with the field’s symmetry axis, the most general axisymmetric (now meaning $\partial/\partial\phi = 0$) magnetic field can be written as:

$$\{\text{E2.70b}\} \quad \boxed{\mathbf{B}(r, \theta, t) = \nabla \times (A(r, \theta, t) \hat{\mathbf{e}}_\phi) + B(r, \theta, t) \hat{\mathbf{e}}_\phi.} \quad (114)$$

Here the vector potential component A defines the *poloidal* components of the magnetic field, i.e., the component contained in meridional (r, θ) planes. The azimuthal component B is often called the *toroidal field*. Equation (114) satisfies the constraint $\nabla \cdot \mathbf{B} = 0$ by construction, and another great advantage of this mixed representation is that the MHD induction equation for the vector \mathbf{B} can be separated in two equations for the scalar components A and B . In the case of pure diffusive decay, and for a magnetic diffusivity η depending at worst only on r , substitution of eq. (114) into (113) leads to:

$$\{\text{E2.10a}\} \quad \frac{\partial A}{\partial t} = \eta \left(\nabla^2 - \frac{1}{\varpi^2} \right) A, \quad (115)$$

$$\{\text{E2.10b}\} \quad \frac{\partial B}{\partial t} = \eta \left(\nabla^2 - \frac{1}{\varpi^2} \right) B + \frac{1}{\varpi} \frac{\partial \eta}{\partial r} \frac{\partial(\varpi B)}{\partial r}, \quad (116)$$

where $\varpi = r \sin \theta$. These are still diffusion-like PDEs, now fully decoupled from one another. In the “exterior” $r > R$ there is only vacuum, which implies vanishing electric currents. In practice we will need to match whatever

solution we compute in $r < R$ to a current-free solution in $r > R$; such a solution must satisfy

$$\mu_0 \mathbf{J} = \nabla \times \mathbf{B} = 0 . \quad (117) \quad \{\text{E2.05}\}$$

For an axisymmetric system, eq. (117) translates into the requirement that

$$\left(\nabla^2 - \frac{1}{\varpi^2} \right) A(r, \theta, t) = 0 , \quad r > R , \quad (118) \quad \{\text{E2.9a}\}$$

$$B(r, \theta, t) = 0 , \quad r > R . \quad (119) \quad \{\text{E2.9b}\}$$

Solutions to eq. (118) have the general form

$$A(r, \theta, t) = \sum_{l=1}^{\infty} a_l \left(\frac{R}{r} \right)^{l+1} Y_{l0}(\cos \theta) \quad r > R , \quad (120) \quad \{\text{E2.9c}\}$$

where the Y_{l0} are the usual spherical harmonics of $m = 0$ azimuthal order, and l is a positive integer, modes with negative l being discarded to ensure proper behavior as $r \rightarrow \infty$.

0.18.2 Reformulation as an eigenvalue problem

Let us now seek specific solutions for a few situations of solar/stellar interest⁸. The first point to note is that the coefficients that appear in eqs. (115)–(116) have no explicit dependence on time; provided that the magnetic diffusivity η is at worst only a function of r , it is then profitable to seek a separable solution of the form:

$$e^{-\lambda t} f_{\lambda}(r) Y_{lm}(\theta, \phi) , \quad (121)$$

where the Y_{lm} are again the spherical harmonic, the natural functional basis for modal development on a spherical surface. Substitution of this *Ansatz* into eqs. (115) or (116), with $m = 0$ in view of axisymmetry, yields the ODE:

$$\left[\frac{1}{r^2} \frac{d}{dr} r^2 \frac{d}{dr} - \frac{l(l+1)}{r^2} + \frac{\lambda}{\eta(r)} \right] f_{\lambda}(r) = 0 . \quad (122) \quad \{\text{E2.tom2}\}$$

Assume now that the magnetic diffusivity η is constant; the spherical Bessel functions $j(kr)$, with $k^2 = \lambda/\eta$, are then the appropriate solution. The decay rate, λ , is then determined by the above 1D eigenvalue problem, along with

⁸ This and the following two subsections are to a large extent adapted from class notes written by Thomas J. Bogdan for the graduate class APAS7500 we co-taught in 1997 at CU Boulder.

some boundary conditions at the surface of the sphere, which turns out to depend on the vector character of the decaying magnetic field.

0.18.3 Poloidal field decay

We first consider the decay of a purely poloidal field, i.e., $f_\lambda(r)$ is taken to describe the radial dependency of the toroidal vector potential component $A(r, \theta, t)$. Both the interior solution and outer potential field solution carry the Y_{l0} angular dependency, so continuity of A at $r = R$ demands that

$$f_\lambda(r) = \begin{cases} j_l(kr) & r < R, \\ j_l(kR) \left(\frac{R}{r}\right)^{l+1} & r > R. \end{cases} \quad (123)$$

The continuity of the radial derivative at $r = R$, necessary for the continuity of the latitudinal component of the magnetic field, then requires

$$kRj_l'(kR) + (l+1)j_l(kR) = kRj_{l-1}(kR) = 0, \quad (124)$$

which means that the decay rate of a poloidal magnetic field is determined by the zeros of a spherical Bessel function. An $l = 1$ dipole calls for the positive zeros of $j_0(x) = \sin x/x$:

$$\lambda_n = \frac{\eta\pi^2 n^2}{R^2}, \quad \text{for } l = 1, n = 1, 2, 3, \dots \quad (125)$$

Notice the many possible *overtone*s associated with $n \geq 2$. These decay more rapidly than the fundamental ($n = 1$), since the radial eigenfunctions possess $n - 1$ field reversals. For such overtones, the effective length scale to be used in the decay-time estimate is roughly the radial distance between the field reversals, or $\approx R/n$.

Figure 6 (top row) shows the first three fundamental ($n = 1$) modes of angular degrees $l = 1, 2, 3$, corresponding to dipolar, quadrupolar, an hexapolar magnetic fields, as well as a few higher overtones for $l = 1, 2$ (bottom row). The decay time estimate provided by eq. (62) turns out to be too large by a factor $\pi^2 \approx 10$, for a sun with constant diffusivity. Still not so bad for a pure order-of-magnitude estimate!

0.18.4 Toroidal field decay

Computing the decay rate of a purely toroidal magnetic field follows the same basic logic. We now require $B = 0$ at $r = R$, but we must further demand that its radial derivative also be continuous, to avoid a blowup of the current density at the surface. This second requirement is what ends up determining the decay rate, which again ends up related to the zero of a spherical Bessel

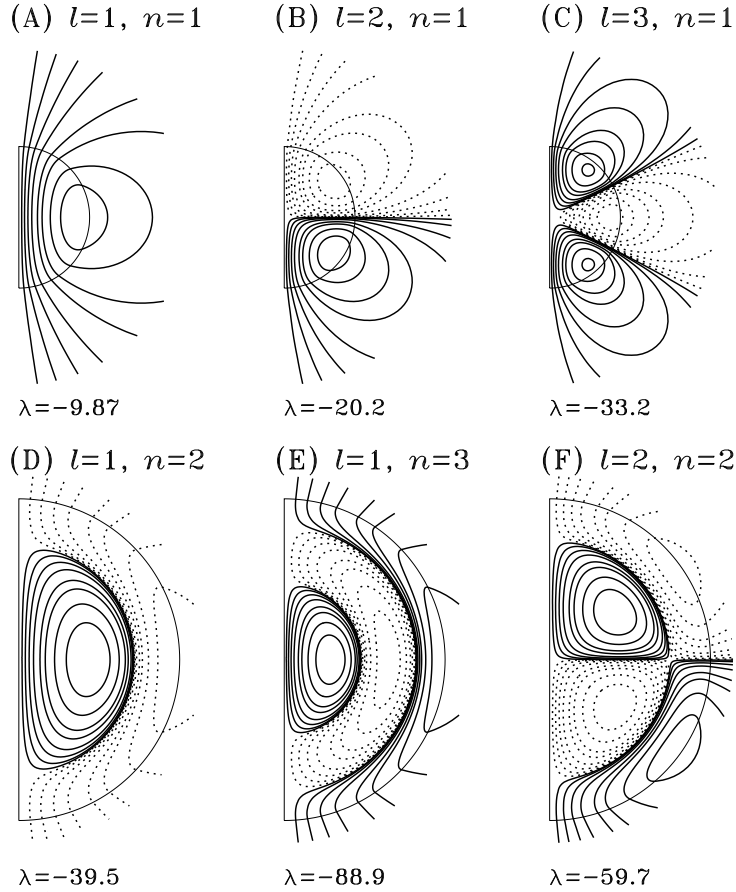


Fig. 6 Six diffusive eigenmodes for a purely poloidal field pervading a sphere of constant magnetic diffusivity embedded in vacuum. The top row shows the three fundamental ($n = 1$) diffusive eigenmodes with smallest eigenvalues, i.e., largest decay times. They correspond to the well-known dipolar, quadrupolar, and hexapolar modes ($l = 1, 2$ and 3). The bottom row shows a few eigenmodes of higher radial overtones. Poloidal fieldlines are shown in a meridional plane, and the eigenvalues are given in units of the inverse diffusion time ($\tau^{-1} \sim \eta/R^2$).

function—only of index l rather than $l - 1$ as was found for the decay of the poloidal field. Hence, a dipole ($l = 1$) toroidal magnetic field decays at precisely the same rate as a quadrupole ($l = 2$) poloidal magnetic field (still for constant diffusivity). Sneaking a peak in a handbook of special functions soon reveals that the decay rate of a dipole toroidal field follows from the transcendental equation:

$$\tan kR = kR . \quad (126)$$

The smallest non-zero solution of this equation gives,

$$\lambda_1 = \frac{\eta(4.493409\dots)^2}{R^2}, \quad l = 1 \text{ toroidal and } l = 2 \text{ poloidal.} \quad (127)$$

{pch_ssec:etavar}

0.18.5 Results for a magnetic diffusivity varying with depth

We end this section by a brief examination of the diffusive decay of large-scale poloidal magnetic fields in the solar interior. The primary complication centers on the magnetic diffusivity, which is no longer constant throughout the domain, and turns out to be rather difficult to compute from first principles. To begin with, the depth variations of the temperature and density in a solar model causes the magnetic diffusivity to increase from about $10^{-2} \text{ m}^2\text{s}^{-1}$ in the central core to $\sim 1 \text{ m}^2\text{s}^{-1}$ at the core-envelope interface. This already substantial variation is however dwarfed by the much larger increase in the net magnetic diffusivity expected in the turbulent environment of the convective envelope. We will look into this in some detail in chapter 3, but for the time being let us simply take for granted that η is much larger in the envelope than in the core.

In order to examine the consequences of a strongly depth-dependent magnetic diffusivity on the diffusive eigenmodes, we consider a simplified situation whereby η assumes a constant value η_c in the core, a constant value η_e ($\gg \eta_c$) in the envelope, the transition occurring smoothly across a thin spherical layer coinciding with the core-envelope interface. Mathematically, such a variation can be expressed as

$$\{\text{E2.127}\} \quad \eta(r) = \eta_c + \frac{\eta_e - \eta_c}{2} \left[1 + \operatorname{erf} \left(\frac{r - r_c}{w} \right) \right], \quad (128)$$

where $\operatorname{erf}(x)$ is the error function, r_c is the radius of the core-envelope interface, and w is the half-width of the transition layer.

We are still facing the 1D eigenvalue problem presented by eq. (122)! Expressing time in units of the diffusion time R^2/η_e based on the envelope diffusivity, we seek numerical solutions, subjected to the boundary conditions $f_\lambda(0) = 0$ and smooth matching to a potential field solution in $r/R > 1$, with the diffusivity ratio $\Delta\eta = \eta_c/\eta_e$ as a parameter of the model. Since we can make a reasonable guess at the first few eigenvalues on the basis of the diffusion time and adopted values of l and η_c ($\sim \pi^2 n l \Delta\eta$, for l and n not too large), a (relatively) simple technique inverse iteration is well-suited to secure both eigenvalues and eigenfunctions for the problem.

Figure 7 shows the radial eigenfunctions for the slowest decaying poloidal eigenmodes ($l = 1$, $n = 1$), with $r_c/R = 0.7$, $w/R = 0.05$ in eq. (128) and diffusivity contrasts $\Delta\eta = 1$ (constant diffusivity), 10^{-1} and 10^{-3} . The corresponding eigenvalues, in units of R^2/η_e , are $\lambda = -9.87$, -2.14 and -0.028 . Clearly, the (global) decay time is regulated by the region of *smallest* diffusiv-

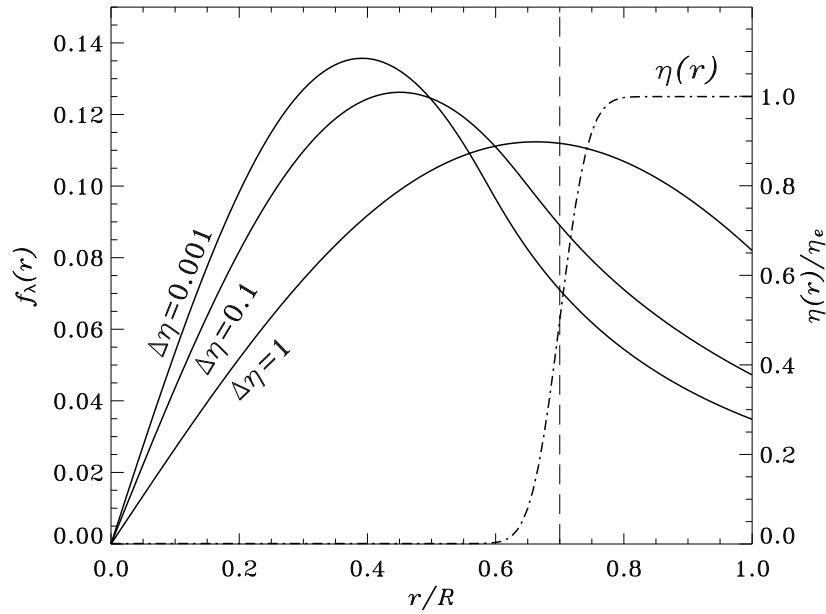


Fig. 7 ^{F2.2} Radial eigenfunctions for the slowest decaying ($\ell = 1$) poloidal eigenmodes ($l = 1, n = 1$) in a sphere embedded in a vacuum. The diffusivity computed using eq. (128) with $r_c/R = 0.7$, $w/R = 0.05$, and for three values of the core-to-envelope diffusivity ratio ($\Delta\eta$). The eigenvalues, in units of η_e/R^2 , are $\lambda = -9.87, -2.14$ and -0.028 for $\Delta\eta = 1, 0.1$, and 10^{-3} , respectively. The diffusivity profile for $\Delta\eta = 10^{-3}$ is also plotted (dash-dotted line). The dashed line indicates the location of the core-envelope interface.

ity, since λ scales approximately as $(\Delta\eta)^{-1}$. Notice also how the eigenmodes are increasingly concentrated in the core region ($r/R \lesssim 0.7$) as $\Delta\eta$ decreases, i.e., they are “expelled” from the convective envelope.

0.18.6 Fossil stellar magnetic fields

The marked decrease of the diffusive decay time with increasing angular and radial degrees of the eigenmodes is a noteworthy result. It means that left to decay long enough, any arbitrarily complex magnetic field in the Sun or stars will eventually end up looking dipolar. Conversely, a fluid flow acting as a dynamo in a sphere and trying to “beat” Ohmic dissipation can be expected to preferentially produce a magnetic field approximating diffusive eigenmodes of low angular and radial degrees (or some combination thereof), since these are the least sensitive to Ohmic dissipation.

There exists classes of early-type main-sequence stars, i.e. stars hotter and more luminous than the Sun and without deep convective envelope, that are

believed to contain strong, large-scale fossil magnetic fields left over from their contraction toward the main-sequence. The chemically peculiar Ap/Bp stars are the best studied class of such objects. Reconstruction of their surface magnetic field distribution suggests almost invariably that the fields are largely dominated by a large-scale dipole-like component, as one would have expected from the preceding discussion if the observed magnetic fields have been diffusively decaying for tens or hundreds of millions of years. It is indeed quite striking that the highest strengths of large-scale magnetic fields in main-sequence stars (a few T in Ap stars), in white dwarfs ($\sim 10^5$ T) and in the most strongly magnetized neutron stars ($\sim 10^{11}$ T) all amount to similar total surface magnetic fluxes, $\sim 10^{19}$ Wb, lending support to the idea that these high field strengths can be understood from simple flux-freezing arguments (§0.12), with field amplification resulting directly from magnetic flux conservation as the star shrinks to form a compact object. We will revisit the origin of A-star magnetic fields in chapter 5.

0.19 Magnetic field amplification by stretching and shearing

{pch_sec:stretch}

Having thus investigated in some details the resistive decay of magnetic field, we turn to the other physical mechanism embodied in eq. (58): growth of the magnetic field in response to the inductive action of a flow \mathbf{u} . We first take a quick look at field amplification in a few idealized model flows, and then move on to a specific example involving a “real” astrophysical flow.

0.19.1 Hydrodynamical stretching and field amplification

Let’s revert for a moment to the ideal MHD case ($\eta = 0$). The induction equation can then expressed as

$$\left\{E2.204b\right\} \quad \left(\frac{\partial}{\partial t} + \mathbf{u} \cdot \nabla\right) \mathbf{B} = \mathbf{B} \cdot \nabla \mathbf{u}, \quad (129)$$

where it was further assumed that the flow is incompressible ($\nabla \cdot \mathbf{u} = 0$). The LHS of eq. (129) is the Lagrangian derivative of \mathbf{B} , expressing the time rate of change of \mathbf{B} in a fluid element moving with the flow. The RHS expresses the fact that this rate of change is proportional to the local *shear* in the flow field. Shearing has the effect of *stretching* magnetic fieldlines, which is what leads to magnetic field amplification.

As a simple example, consider on Figure 8 a cylindrical fluid element of length L_1 , threaded by a magnetic field parallel to the axis of the cylinder, embedded in a perfectly conducting incompressible fluid and subjected to a

stretching motion ($\partial u_z / \partial z > 0$) along its central axis such that its length increases to L_2 . Mass conservation demands that

$$\frac{R_2}{R_1} = \sqrt{\frac{L_1}{L_2}}. \quad (130) \quad \{\text{E3.1rebis}\}$$

Conservation of the magnetic flux ($= \pi R^2 B$), in turn, leads to

$$\frac{B_2}{B_1} = \frac{L_2}{L_1}, \quad (131) \quad \{\text{E3.1bis}\}$$

i.e., the field strength is amplified in direct proportion to the level of stretch-

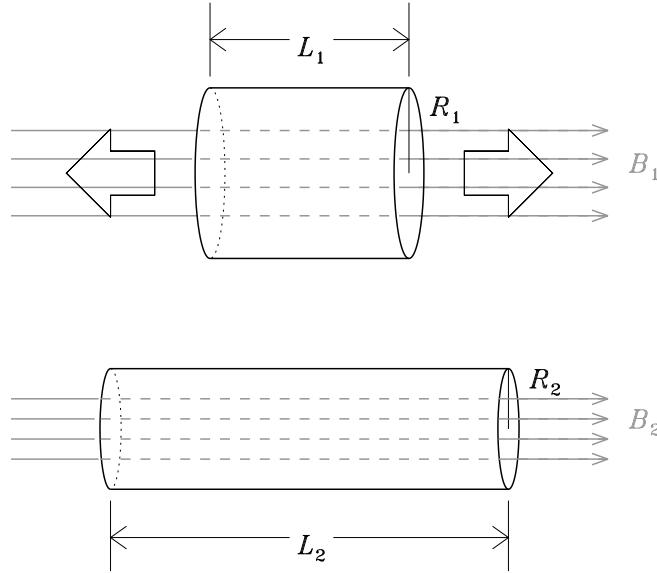


Fig. 8 ^{E2.0} Stretching of a magnetized cylindrical fluid element by a diverging flow. The magnetic field (fieldlines in gray) is horizontal within the tube, has a strength B_1 originally, and B_2 after stretching. In the flux-freezing limit mass conservation within the tube requires its radius to decrease, which in turn leads to field amplification (see text).

ing. This almost trivial result is in fact at the very heart of *any* magnetic field amplification in the magnetohydrodynamical context, and illustrates two crucial aspects of the mechanism: first, this works only if the fieldlines are frozen into the fluid, i.e., in the high- R_m regime. Second, mass conservation plays an essential role here; the stretching motion along the tube axis *must* be accompanied by a compressing fluid motion perpendicular to the axis if mass

conservation is to be satisfied. It is this latter compressive motion, occurring perpendicular to the magnetic fieldlines forming the flux tube, that is ultimately responsible for field amplification; the horizontal fluid motion occurs parallel to the magnetic fieldline, and so cannot in itself have any inductive effect as per eq. (58).

The challenge, of course, is to realize this idealized scenario in practice, i.e., to find a flow which achieves the effect illustrated on Figure 8. This, it turns out, is much simpler than one might expect! Working in cylindrical coordinates (s, ϕ, z) , consider the following incompressible flow:

$$\{\text{eq:cylstag}\} \quad u_s(s) = \frac{\alpha s}{2}, \quad u_\phi = 0, \quad u_z(z) = -\alpha z, \quad (132)$$

with $\alpha > 0$. This describes a flow converging towards the $z = 0$ plane along the z -axis, and diverging radially away from the origin within the $z = 0$ plane. Clearly, $\mathbf{u} = 0$ at the origin $(0, \phi, 0)$ of the cylindrical coordinate system. This is called a *stagnation point*, and its presence is vital to the inductive amplification of the magnetic field. Now place a thin, straight magnetic flux tube in the $z = 0$ plane, and enclosing the stagnation point, as shown on Fig. 9. You can easily verify that you will get exactly the type of stretching effect illustrated in cartoon form on Fig. 8.

0.19.2 The Vainshtein & Zeldovich flux rope dynamo

So, the linear stretching of a flux tube amplifies the magnetic field, but the magnetic flux remains constant by the very nature of the amplification mechanism. Nonetheless, this idea actually forms the basis of a dynamo that can increase both the magnetic field strength and flux. S. Vainshtein and Ya. B. Zeldovich have proposed one of the first and justly celebrated “cartoon” model for this idea, as illustrated on Figure 10. The steps are the following:

1. A circular rope of magnetic field is stretched to twice its length ($a \rightarrow b$). As we just learned, this doubles the magnetic field strength while conserving the flux;
2. The rope is twisted by half a turn ($c \rightarrow d$);
3. One half of the rope is folded over the other half in such a way as to align the magnetic field of each half ($e \rightarrow f$); this now doubles the magnetic flux through any plane crossed by the stacked loops.

This is quite remarkable; the so-called *stretch-twist-fold* sequence (hereafter STF) illustrated on Fig. 10 first doubles the field strength while conserving the total cross-section of the original rope, because the tube’s cross-section ($\propto R^2$) varies as the inverse of its length (as per eq. 130)), then folding doubles the magnetic flux without reducing the field strength since the loop’s cross-section remains unaffected. If the sequence is repeated n times,

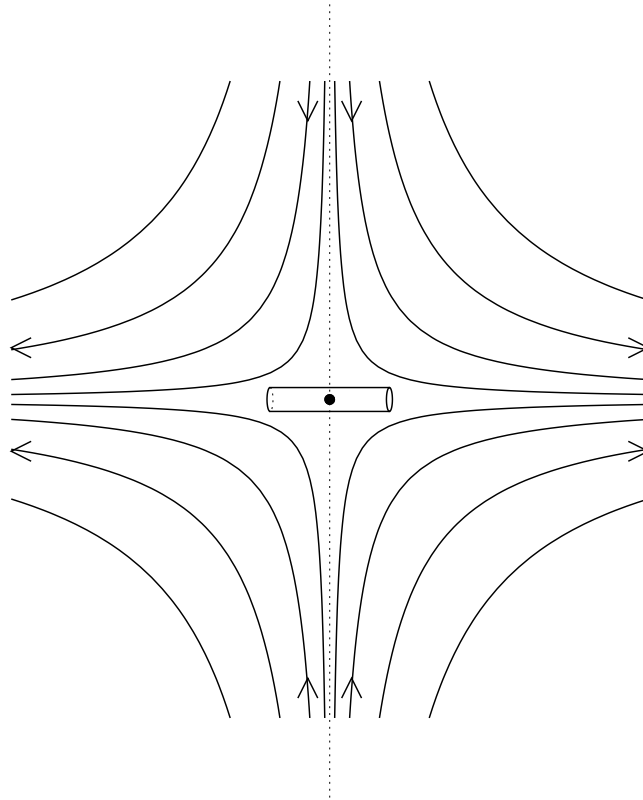


Fig. 9 ^{fig:stagn} Streamlines of the stagnation point flow defined by eq. (132), plotted in a constant- ϕ plane. The flow is rotationally invariant about the symmetry (z) axis, indicated by the dotted line, and the stagnation point (solid dot) is located at the origin of the cylindrical coordinate system. A thin magnetic flux tube located in the $z = 0$ plane and crossing the origin, as shown, will be subjected to the shearing motion illustrated in cartoon form on Fig. 8.

the magnetic field strength (and flux) is then amplified by a factor

$$\{E3.2\} \quad \frac{B^n}{B_0} \propto 2^n = \exp(n \ln 2), \quad (133)$$

with n playing the role of a (discrete) time-like variable, Equation (133) indicates an exponential growth of the magnetic field, with a growth rate $\sigma = \ln 2$. Rejoice! This is our first dynamo!

A concept central to the STF dynamo —and other dynamos to be encountered later— is that of *constructive folding*. Note how essential the twisting

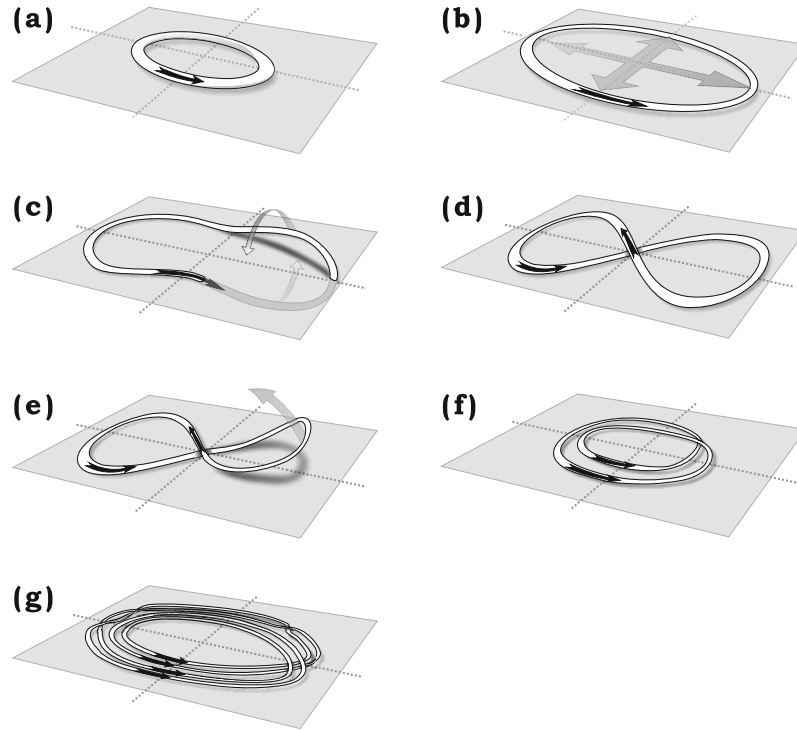


Fig. 10 ^{F3, VZ} Cartoon of the Stretch-Twist-Fold flux rope dynamo of Vainshtein & Zeldovich. A circular flux rope is (a)→(b) stretched, (c)→(d) twisted, and (e)→(f) folded. Diagram (g) shows the resulting structure after another such sequence acting on (f). Diagram produced by D. Passos.

step is to the STF dynamo: without it (or with an even number of twists), the magnetic field in each half of the folded rope would end up pointing in opposite direction, and would then add up to zero net flux, a case of *destructive folding*. We'll have a more to say on the STF dynamo later on; for now we turn to amplification by fluid motions that shear rather than stretch.

{SSshear} 0.19.3 Hydrodynamical shearing and field amplification

Magnetic field amplification by stretching, as illustrated on Figure 8, evidently requires (1) a stagnation point in the flow, and (2) a rather specific positioning and orientation of the flux tube with respect to this stagnation point. We will encounter later more realistic flows that do achieve field amplification through stretching in the vicinity of stagnation points, but there is a different type of fluid motion that can produce a more robust form of magnetic field amplification: *shearing*. The idea is illustrated in cartoon form

on Figure 11. We start with a magnetic flux tube, as before, but this time the flow is everywhere perpendicular to the axis of the tube, and its magnitude varies with height along the length of the tube, e.g., $\mathbf{u} = u_x(z)\hat{\mathbf{e}}_x$, so that $\partial u_x/\partial z \neq 0$; this is called a *planar sheared flow*. In the ideal MHD limit, every small section of the tube is displaced sideways at a rate equal to the flow speed. After a while, the tube will no longer be straight, and for the type of shearing motion illustrated on Figure 11, its length will have increased. By the same logic as before, the field strength within the tube must have increased proportionally to the increase in length of the tube. Note also that here the action of the shear leaves the z -component of the magnetic field unaffected, but produces an x -component where there was initially none.

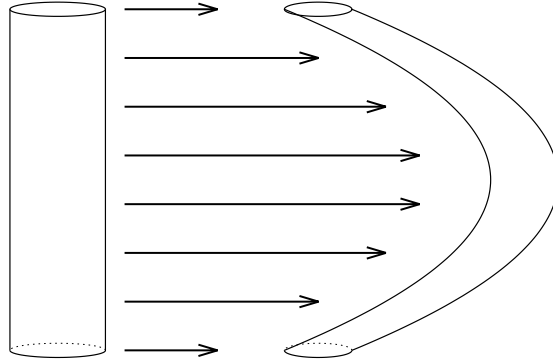


Fig. 11 ^{fig:shear} Stretching of a flux tube by a shearing motion directed perpendicularly to the tube's axis. Unlike on Fig. 8, here the tube does not remain straight, but its length is still increasing as a consequence of the tube's deformation; consequently, in the ideal MHD limit the magnetic field threading the tube will be amplified by a factor given by the ratio of its final to initial length, as per eq. (131).

The beauty of this mechanism is that it does not require stagnation points, and will in general operate for any sheared flow. The latter turn out to be rather common in astrophysical objects, and we now turn to one particularly important example.

0.19.4 Toroidal field production by differential rotation

{SSdr}

A situation of great (astro)physical interest is the induction of a toroidal magnetic field via the shearing of a pre-existing poloidal magnetic field threading

a differentially rotating sphere of electrically conducting fluid. Working now in spherical polar coordinates (r, θ, ϕ) and assuming overall axisymmetry (i.e., the poloidal field and differential rotation share the same symmetry axis), the flow velocity can be written as:

$$\{\text{E2.07}\} \quad \mathbf{u}(r, \theta) = \varpi \Omega(r, \theta) \hat{\mathbf{e}}_\phi, \quad (134)$$

where again $\varpi = r \sin \theta$, and the angular velocity $\Omega(r, \theta)$ is assumed steady ($\partial/\partial t = 0$), corresponding to the kinematic regime introduced earlier. Neglecting once again magnetic dissipation, the induction equation takes on the reduced form

$$\{\text{E2.07a}\} \quad \frac{\partial A}{\partial t} = 0, \quad (135)$$

$$\{\text{E2.07b}\} \quad \frac{\partial B}{\partial t} = \varpi [\nabla \times (A \hat{\mathbf{e}}_\phi)] \cdot \nabla \Omega, \quad (136)$$

where we took again advantage of the poloidal/toroidal decomposition introduced in in §0.18. Equation (136) integrates immediately to

$$\{\text{E2.08}\} \quad B(r, \theta, t) = B(r, \theta, 0) + \left(\varpi [\nabla \times (A \hat{\mathbf{e}}_\phi)] \cdot \nabla \Omega \right). \quad (137)$$

Anywhere in the domain, the toroidal component of the magnetic field grows linearly in time, at a rate proportional to the net local shear and local poloidal field strength. A toroidal magnetic component is being generated by shearing the initially purely poloidal fieldlines in the ϕ -direction, and the magnitude of the poloidal magnetic component remains unaffected, as per eq. (135). Note also that for such an axisymmetric configuration, the only possible steady-state ($\partial/\partial t = 0$) solutions must have

$$\{\text{eq:ferarro}\} \quad [\nabla \times (A \hat{\mathbf{e}}_\phi)] \cdot \nabla \Omega = 0, \quad (138)$$

i.e., the angular velocity must be constant on any given poloidal flux surface. This result is known as *Ferraro's theorem*.

Evidently, computing B via eq. (137) requires a knowledge of the solar internal (differential) rotation profile $\Omega(r, \theta)$. Consider the following parametrization:

$$\{\text{E2.68a}\} \quad \Omega(r, \theta) = \Omega_C + \frac{\Omega_S(\theta) - \Omega_C}{2} \left[1 + \operatorname{erf} \left(\frac{r - r_C}{w} \right) \right], \quad (139)$$

where

$$\{\text{E2.68b}\} \quad \Omega_S(\theta) = \Omega_{Eq} (1 - a_2 \cos^2 \theta - a_4 \cos^4 \theta) \quad (140)$$

is the surface latitudinal differential rotation. We will make repeated use of this parametrization in this and following and chapters, so let's look into it in some detail. Figure 12A shows isocontours of angular velocity (in black) generated by the above parameterization with parameter values $\Omega_C/2\pi = 432.8$ nHz, $\Omega_{Eq}/2\pi = 460.7$ nHz, $a_2 = 0.1264$, $a_4 = 0.1591$, $r_c = 0.713R$, and $w = 0.05R$, as obtained by a best-fit to helioseismic frequency splittings. This properly reproduces the primary features of full helioseismic inversions, namely:

1. A convective envelope ($r \gtrsim r_c$) where the shear is purely latitudinal, with the equatorial region rotating faster than the poles;
2. A core ($r \lesssim r_c$) that rotates rigidly, at a rate equal to that of the surface mid-latitudes;
3. A smooth matching of the core and envelope rotation profiles occurring across a thin spherical layer coinciding with the core-envelope interface ($r = r_c$), known as the *tachocline*.

It should be emphasized already at this juncture that such a solar-like differential rotation profile is quite complex, in that it is characterized by *three* partially overlapping shear regions: a strong positive radial shear in the equatorial regions of the tachocline, an even stronger negative radial shear in its polar regions, and a significant latitudinal shear throughout the convective envelope and extending partway into the tachocline. For a tachocline of half-thickness $w/R_\odot = 0.05$, the mid-latitude latitudinal shear at $r/R_\odot = 0.7$ is comparable in magnitude to the equatorial radial shear; as we will see in the next chapter, its potential contribution to dynamo action should not be casually dismissed.

Figure 12B shows the distribution of toroidal magnetic field resulting from the shearing of the slowest decaying, $n = 1$ dipole-like diffusive eigenmode of §0.18 of strength 10^{-4} T at $r/R = 0.7$, using the diffusivity profile given by eq. (128) with diffusivity contrast $\Delta\eta = 10^{-2}$ (part A, red lines). This is nothing more than eq. (137) evaluated for $t = 10$ yr, with $B(r, \theta, 0) = 0$. Not surprisingly, the toroidal field is concentrated in the regions of large radial shear, at the core-envelope interface (dashed line). Note how the toroidal field distribution is *antisymmetric* about the equatorial plane, precisely what one would expect from the inductive action of a shear flow that is equatorially symmetric on a poloidal magnetic field that is itself antisymmetric about the equator.

Knowing the distributions of toroidal and poloidal fields on Figure 12 allows us to flirt a bit with dynamics, by computing the ϕ -component of the Lorentz force:

$$[\mathbf{F}]_\phi = \frac{1}{\mu_0 \varpi} \mathbf{B}_p \cdot \nabla(\varpi B), \quad (141) \quad \{\text{E5.09}\}$$

The resulting spatial distribution of $[\mathbf{F}]_\phi$ is plotted on Figure 12C. Examine this carefully to convince yourself that the Lorentz force is such as to *op-*

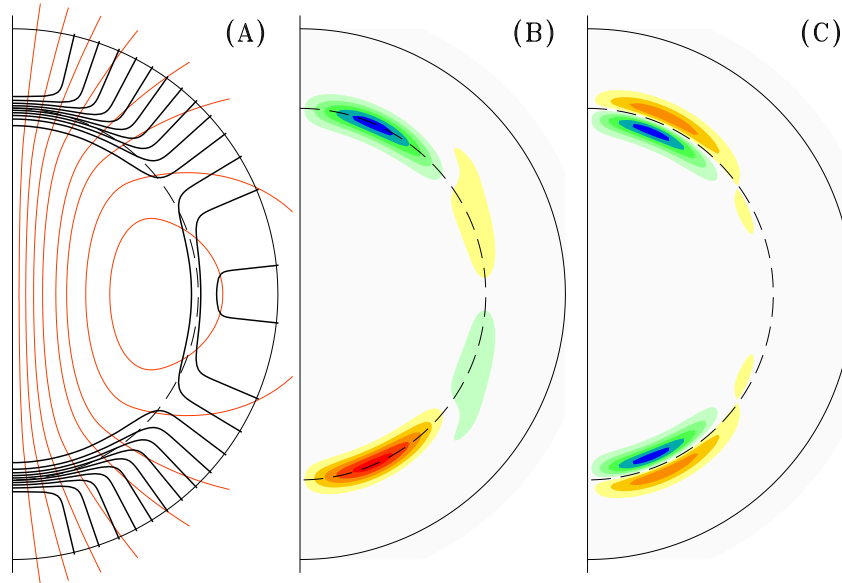


Fig. 12 ^{F2.14} Shearing of a poloidal field into a toroidal component by a solar-like differential rotation profile. Part A shows isocontours of the rotation rate $\Omega(r, \theta)/2\pi$ (solid lines, contour spacing 10 nHz). The red lines are fieldlines for the $n = 1$ dipolar diffusive eigenmode with core-to-envelope diffusivity contrast $\Delta\eta = 10^{-2}$. The dashed line is the core-envelope interface at $r/R = 0.7$. Part B shows isocontours of the toroidal field, with yellow-red (green-blue) corresponding to positive (negative) B . The maximum toroidal field strength is about 0.2T, and contour spacing is 0.02T. Part C shows logarithmically spaced isocontours of the ϕ -component of the Lorentz force associated with the poloidal/toroidal fields of panels A and B.

pose the driving shear. This is an important and totally general property of interacting flows and magnetic fields: the Lorentz force tends to resist the hydrodynamical stretching responsible for field induction. The ultimate fate of the system depends on whether the Lorentz force become dynamically significant before the growth of the toroidal field is mitigated by resistive dissipation; in solar/stellar interiors the former situation is far more likely.

Clearly, the growing magnetic energy of the toroidal field is supplied by the kinetic energy of the rotational shearing motion (this is hidden the second term on the RHS of eq. (84)). In the solar case, this is an attractive field amplification mechanism, because the available supply of rotational kinetic energy is immense⁹. But don't make the mistake of thinking that this is a dynamo! In obtaining eq. (137) we have completely neglected magnetic dissipation, and remember, the dynamos we are seeking are flows that can

⁹ This may no longer be the case, however, if dynamo action takes place in a thin layer below the base of the convective envelope; see the paper by Steiner & Ferris-Mas in the bibliography of the next chapter, for more on this aspect of the problem.

amplify and sustain a magnetic field *against* Ohmic dissipation. In fact, neither flux tube stretching (Fig. 8) or shearing (Fig 11) is a dynamo either, for the same reason¹⁰. Nonetheless, shearing of a poloidal field by differential rotation will turn out to be a central component of *all* solar/stellar dynamo models constructed in later chapters. It is also believed to be an important ingredient of magnetic amplification in stellar accretion disks, and even in galactic disks.

0.20 Magnetic field evolution in a cellular flow

{Sbadvec}

Having examined separately the resistive decay and hydrodynamical induction of magnetic field, we now turn to a situation where both processes operate simultaneously.

0.20.1 A cellular flow solution

{Scell}

Working now in Cartesian geometry, we consider the action of a steady, incompressible ($\nabla \cdot \mathbf{u} = 0$) two-dimensional flow

$$\mathbf{u}(x, y) = u_x(x, y)\hat{\mathbf{e}}_x + u_y(x, y)\hat{\mathbf{e}}_y \quad (142) \quad \{\text{E2.29}\}$$

on a two-dimensional magnetic field

$$\mathbf{B}(x, y, t) = B_x(x, y, t)\hat{\mathbf{e}}_x + B_y(x, y, t)\hat{\mathbf{e}}_y . \quad (143) \quad \{\text{E2.210}\}$$

Note that neither the flow nor the magnetic field have a z -component, and that their x and y -components are both independent of the z -coordinate. The flow is said to be *planar* because $u_z = 0$, and has an ignorable coordinate (i.e., translational symmetry) since $\partial/\partial z \equiv 0$ for all field and flow components. Such a magnetic field can be represented by the vector potential

$$\mathbf{A} = A(x, y, t)\hat{\mathbf{e}}_z , \quad (144) \quad \{\text{E2.211}\}$$

where, as usual, $\mathbf{B} = \nabla \times \mathbf{A}$. Under this representation, lines of constant A in the $[x, y]$ plane coincide with magnetic fieldlines. The only non-trivial component of the induction equation (93) is its z -components, which takes the form

$$\frac{\partial A}{\partial t} + \mathbf{u} \cdot \nabla A = \eta \nabla^2 A . \quad (145) \quad \{\text{E2.212}\}$$

¹⁰ The STF dynamo (Fig. 10) is the lone exception here, it that it remains a dynamo even when magnetic dissipation is brought into the picture; the reasons why are subtle and will be clarified later on.

This is a linear advection-diffusion equation, describing the transport of a passive scalar quantity A by a flow \mathbf{u} , and subject to diffusion, the magnitude of which being measured by η . In view of the symmetry and planar nature of the flow, it is convenient to write the 2-D flow field in terms of a stream function $\Psi(x, y)$:

$$\{\text{E2.213}\} \quad \mathbf{u}(x, y) = u_0 \left(\frac{\partial \Psi}{\partial y} \hat{\mathbf{e}}_x - \frac{\partial \Psi}{\partial x} \hat{\mathbf{e}}_y \right) . \quad (146)$$

It is easily verified that any flow so defined will identically satisfy the condition $\nabla \cdot \mathbf{u} = 0$. As with eq. (144), a given numerical value of Ψ uniquely labels one streamline of the flow. Consider now the stream function

$$\{\text{E2.214}\} \quad \Psi(x, y) = \frac{L}{4\pi} \left(1 - \cos \left(\frac{2\pi x}{L} \right) \right) \left(1 - \cos \left(\frac{2\pi y}{L} \right) \right), \quad x, y \in [0, L] \quad (147)$$

This describes a counterclockwise cellular flow centered on $(x, y) = (L/2, L/2)$, as shown on Figure 13. The maximal velocity amplitude $\max\|\mathbf{u}\| = u_0$ is found along the streamline $\Psi = u_0 L / (2\pi)$, plotted as a thicker line on Figure 13. This streamline is well approximated by a circle of radius $L/4$, and its streamwise circulation period turns out to be $1.065\pi L / 2u_0$, quite close to what one would expect in the case of a perfectly circular streamline. In what follow this timescale is denoted τ_c and referred to as the *turnover time* of the flow. Note that both the normal and tangential components of the flow vanish on the boundaries $x = 0, L$ and $y = 0, L$. This implies that the domain boundary is itself a streamline ($\Psi = 0$, in fact), and that every streamline interior to the boundary closes upon itself within the spatial domain. These (simple) topological properties of the flow defined by eqs. (146) and (147) may seem largely irrelevant at this stage of our inquiries, but later developments will reveal that they are in fact crucial to the dynamo problem.

We now investigate the inductive action of this flow by solving a nondimensional version of eq. (145), by expressing all lengths in units of L , and time in units of L/u_0 , so that

$$\{\text{E2.215}\} \quad \frac{\partial A}{\partial t} = - \frac{\partial \Psi}{\partial y} \frac{\partial A}{\partial x} + \frac{\partial \Psi}{\partial x} \frac{\partial A}{\partial y} + \frac{1}{R_m} \left(\frac{\partial^2 A}{\partial x^2} + \frac{\partial^2 A}{\partial y^2} \right), \quad x, y \in [0, L] \quad (148)$$

where $R_m = u_0 L / \eta$ is the magnetic Reynolds number for this problem, and the corresponding diffusion time is then $\tau_\eta = R_m$ in dimensionless units. Equation (148) is solved as an initial-boundary value problem in two spatial dimensions. All calculations described below start at $t = 0$ with an initially uniform, constant magnetic field $\mathbf{B} = B_0 \hat{\mathbf{e}}_x$, equivalent to:

$$\{\text{E2.216a}\} \quad A(x, y, 0) = B_0 y . \quad (149)$$

We consider a situation where the magnetic field component normal to the boundaries is held fixed, which amounts to holding the vector potential fixed

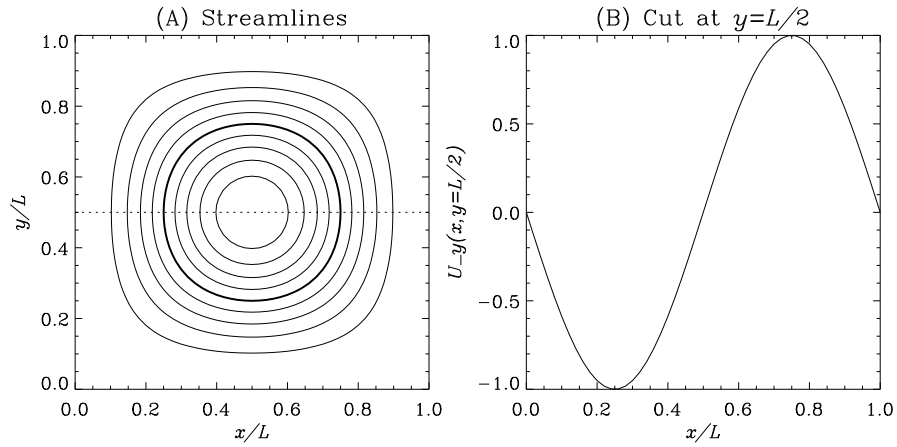


Fig. 13 ^{F2.3} Counterclockwise cellular flow generated by the streamfunction given by eq. (147). Part (A) shows streamlines of the flow, with the thicker streamline corresponding to $\Psi = u_0 L / (2\pi)$, on which the flow attains its maximum speed u_0 . Part (B) shows the profile of $u_y(x)$ along an horizontal cut at $y = 1/2$. A “typical” length scale for the flow is then $\sim L$.

on the boundary. Figure 14 shows the variation with time of the magnetic energy (eq. (86)), for four solutions having $R_m = 10, 10^2, 10^3$ and 10^4 . Figure 15 shows the evolving shape of the magnetic fieldlines in the $R_m = 10^3$ solution at 9 successive epochs. The solid dots are “floaters”, namely Lagrangian markers moving along with the flow. At $t = 0$ all floaters are equidistant and located on the fieldline initially coinciding with the coordinate line $y/L = 0.5$, that (evolving) fieldline being plotted in the same color as the floaters on all panels. Figure 15 covers two turnover times ¹¹.

At first, the magnetic energy increases quadratically in time. This is precisely what one would expect from the shearing action of the flow on the initial B_x -directed magnetic field, which leads to a growth of the B_y -component that is linear in time. However, for $t/\tau_c \gtrsim 2$ the magnetic energy starts to decrease again and eventually ($t/\tau_c \gg 1$) levels off to a constant value. To understand the origin of this behavior we need to turn to Figure 15 and examine the solutions in some detail.

The counterclockwise shearing action of the flow is quite obvious on Fig. 15 in the early phases of the evolution, leading to a rather pretty spiral pattern as magnetic fieldlines get wrapped around one another. Note that the distortion of magnetic fieldlines by the flow implies a great deal of *stretching* in the streamwise direction, as well folding in the cross-stream direction. The former shows up as sharp bends in the fieldlines, while the former is most obvious upon noting that the distance between adjacent floaters increases monoton-

¹¹ An animation of this solution, and related additional solutions, can be viewed on the course Web Page.

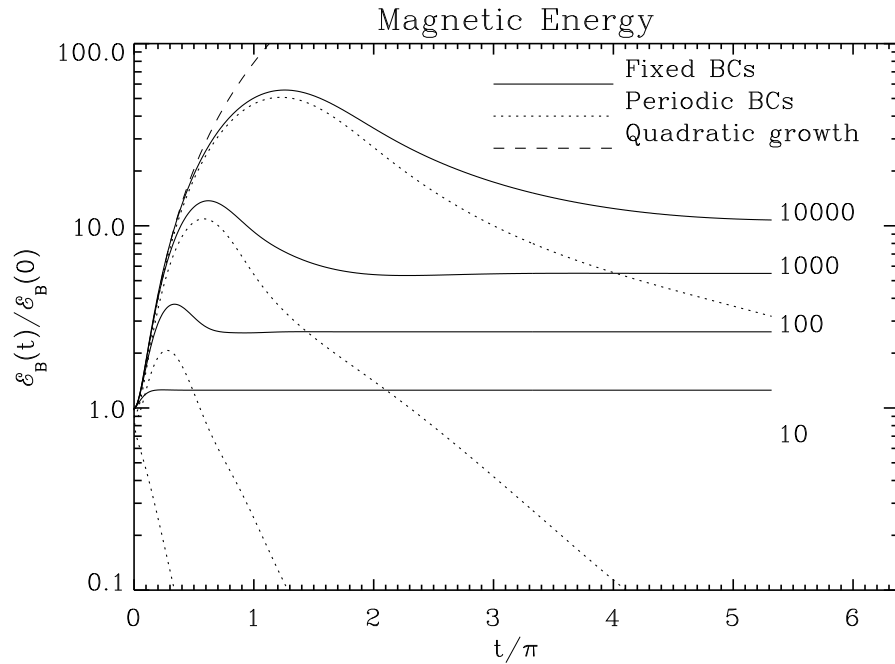


Fig. 14 ^{E2.5} Evolution of the magnetic energy for solutions with increasing R_m . The solutions have been computed over 10 turnover times, at which point they are getting reasonably close to steady-state, at least as far as magnetic energy is concerned. One turnover time corresponds to $t/\pi = 0.532$.

ically in time. It is no accident that the floaters end up in the regions of maximum field amplification on frames 2–5; they are initially positioned on the fieldline coinciding with the line $y = L/2$, everywhere perpendicular to the shearing flow (cf. Figs. 11 and 13), which pretty much ensures maximal inductive effect, as per eq. (145).

The fact that all floaters remain at first “attached” onto their original fieldline is what one would have expected from the fact that this is a relatively high- R_m solution, so that flux-freezing is effectively enforced. As the evolution proceeds, the magnetic field keeps building up in strength (as indicated by the color scale), but is increasingly confined to spiral “sheets” of decreasing thickness. Coincident with these sheets are strong electrical currents perpendicular to the plane of the page, the current density being given here by

$$J_z(x, y) = \frac{1}{\mu_0} \nabla^2 A(x, y) , \quad (150)$$

and a time-evolution resembling that of magnetic energy.

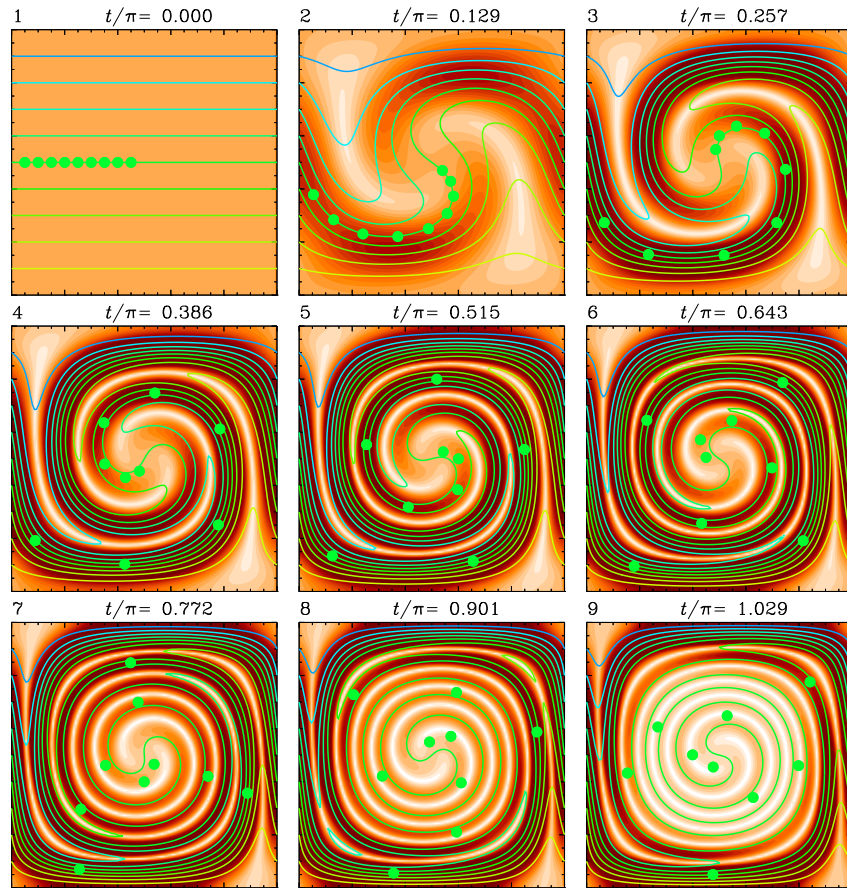


Fig. 15 ^{F2.4} Solution to equation (148) starting from an initially horizontal magnetic field. The panels show the shape of the magnetic fieldlines at successive times. The color scale encodes the absolute strength of the magnetic field, i.e., $\sqrt{B_x^2 + B_y^2}$. The x - and y -axes are horizontal and vertical, respectively, and span the range $x, y \in [0, L]$. Time t is in units of L/u_0 . The solid dots are “floaters”, i.e., Lagrangian marker passively advected by the flow. The magnetic Reynolds number is $R_m = 10^3$.

By the time we hit one turnover time (corresponding approximately to frame 5 on Fig. 15), it seems that we are making progress towards our goal of producing a dynamo; we have a flow field which, upon acting on a preexisting magnetic field, has intensified the strength of that field, at least in some localized regions of the spatial domain. However, beyond $t \sim \tau_c$ the sheets of magnetic fields are gradually disappearing, first near the center of the flow cell (frames 5–7), and later everywhere except close to the domain boundaries (frames 7–9). Notice also how, from frame 5 onward, the floaters are seen to “slip” off their original fieldlines. This means that flux-freezing no longer

holds; in other words, diffusion is taking place. Yet, we evidently still have $t \ll \tau_\eta$ ($\equiv R_m = 10^3$ here), which indicates that diffusion should not yet have had enough time to significantly affect the solution. What is going on here?

{SSf1exp} 0.20.2 Flux expulsion

The solution to this apparent dilemma lies with the realization that we have defined R_m in terms of the global length scale L characterizing the flow. This was a perfectly sensible thing to do on the basis of the flow configuration and initial condition on the magnetic field. However, as the evolution proceeds beyond $\sim \tau_c$ the decreasing thickness of the magnetic field sheets means that the global length scale L is no longer an adequate measure of the “typical” length scale of the magnetic field, which is what is needed to estimate the diffusion time τ_η (see eq. (62)). Figure 16 shows a series of cuts of the vector potential A in a $R_m = 10^4$ solution, plotted along the coordinate line $y = L/2$, at equally spaced successive time intervals covering two turnover times. Clearly the inexorable winding of the fieldline leads to a general decrease of

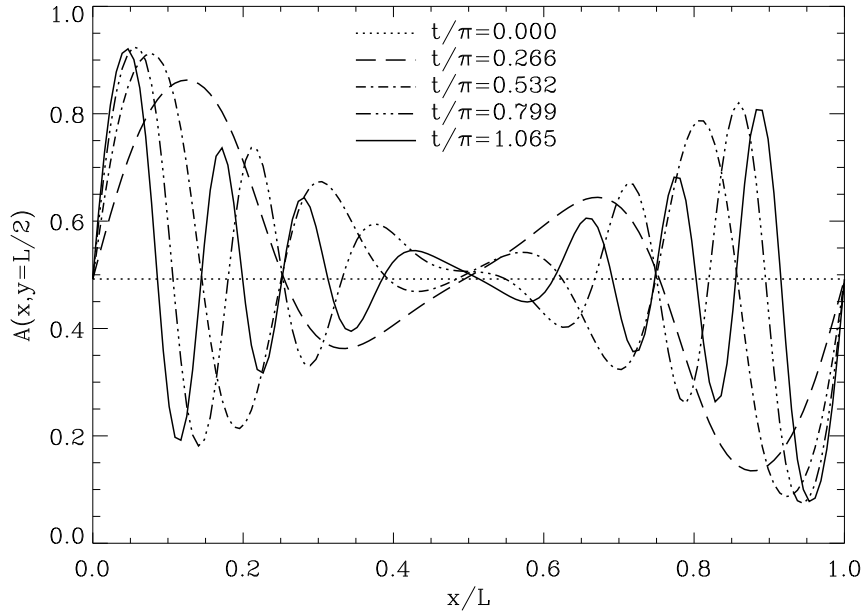


Fig. 16 ^{F2.7} Cuts of a $R_m = 10^4$ solution along the coordinate line $y = 0.5$, at successive times. Note how the “typical” length scale ℓ for the solution decreases with time, from $\ell/L \sim 0.25$ at $t/\pi = 0.266$, down to $\ell/L \sim 0.05$ after two turnover times ($t/\pi = 1.065$).

the length scale characterizing the evolving solution. In fact, each turnover

time adds two new “layers” of alternating magnetic polarity to the spiraling sheet configuration, so that the average length scale ℓ decreases as t^{-1} :

$$\frac{\ell(t)}{L} \propto \frac{L}{u_0 t}, \quad (151) \quad \{\text{E2.218}\}$$

which in turn implies that the *local* dissipation time is also decreasing as t^{-1} . On the other hand, examination of Fig. 15 soon reveals that the (decreasing) length scale characterizes the thickness of elongated magnetic structures that are themselves more or less *aligned* with the streamlines, so that the turnover time τ_c remains the proper timescale measuring field induction. With τ_c fixed and τ_η inexorably decreasing, the solution is bound to reach a point where $\tau_\eta \simeq \tau_c$, *no matter how small dissipation actually is*. To reach that stage just takes longer in the higher R_m solutions, since more winding of the fieldlines is needed. Larger magnetic energy can build up in the transient phase, but the growth of the magnetic field is *always* arrested. Equating τ_c ($\sim L/u_0$) to the *local* dissipation time ℓ^2/η , one readily finds that the length scale ℓ at which both process become comparable can be expressed in terms of the *global* R_m as

$$\frac{\ell}{L} = (R_m)^{-1/2}, \quad R_m = \frac{u_0 L}{\eta}. \quad (152) \quad \{\text{E2.251}\}$$

That such a balance between induction and dissipation materializes means that a steady-state can be attained. Figure 17 shows four such steady state solutions for increasing values of the (global) magnetic Reynolds number R_m . The higher R_m solutions clearly show *flux expulsion* from the central regions of the domain. This is a general feature of steady, high- R_m magnetized flows with closed streamlines: magnetic flux is expelled from the regions of closed streamlines towards the edges of the flow cells, where it ends up concentrated in *boundary layers* which indeed have a thickness of order $R_m^{-1/2}$, as suggested by eq. (152), an within which strong z -directed electrical currents flow — and dissipate! It is important to understand how and why this happens.

To first get an intuitive feel for how flux expulsion operates, go back to Figure 15. As the flow wraps the fieldlines around one another, it does so in a manner that folds fieldlines of opposite polarity closer and closer to each other. When two such fieldlines are squeezed closer together than the dissipative length scale (eq. [152]), resistive decay takes over and destroys the field faster than it is being stretched. This is an instance of *destructive folding*, and can only be avoided along the boundaries, where the normal component of the field is held fixed. For flux expulsion to operate, flux-freezing must be effectively enforced on the spatial scale of the flow. Otherwise the field is largely insensitive to the flow, and fieldlines are hardly deformed with respect to their initial configuration (as on panel [A] of Fig. 17).

Consider now the implication for the total magnetic flux across the domain; flux conservation requires that the normal flux $B_0 L$ imposed at the right

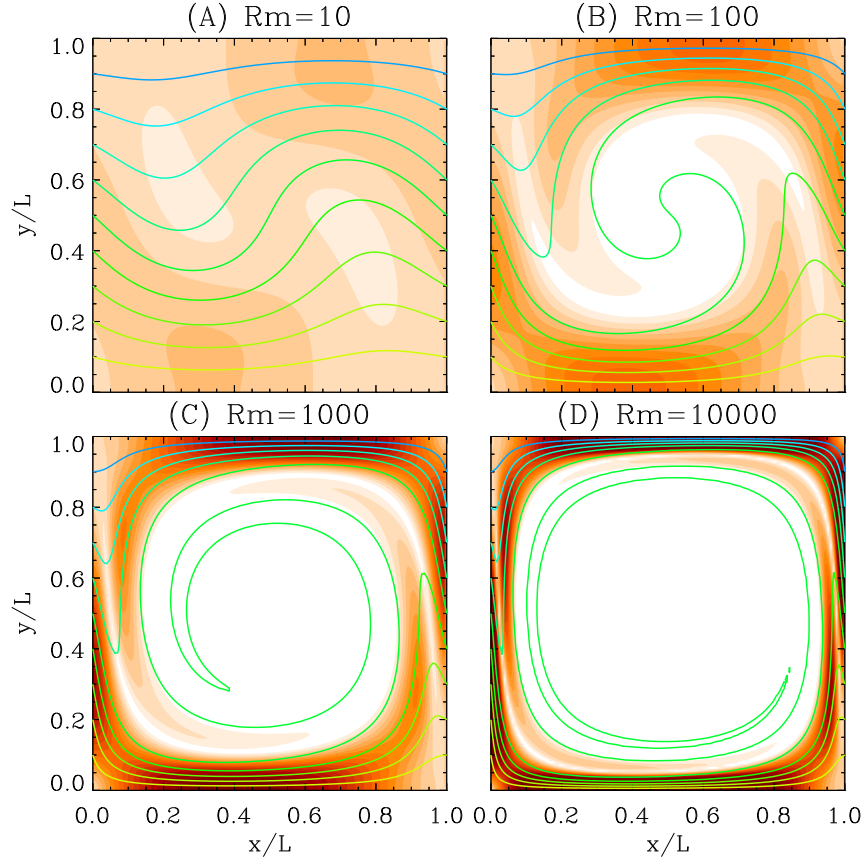


Fig. 17 ^{F2.8} Steady-state solutions to the cellular flow problem, for increasing values of the magnetic Reynolds number R_m . The $R_m = 10^4$ solution is at the resolution limit of the $N_x \times N_y = 128 \times 128$ mesh used to obtain these solutions, as evidenced on part (D) by the presence of small scale irregularities where magnetic fieldlines are sharply bent. The color scale encodes the local magnitude of the magnetic field. Note how, in the higher R_m solutions, magnetic flux is expelled from the center of the flow cell. With $\mathcal{E}_B(0)$ denoting the energy of a purely horizontal field with same normal boundary flux distribution, the magnetic energy for these steady states is $\mathcal{E}_B/\mathcal{E}_B(0) = 1.37, 2.80, 5.81$ and 11.75 , respectively, for panels (A) through (D).

and left boundaries must somehow cross the interior, otherwise Maxwell's equation $\nabla \cdot \mathbf{B} = 0$ would not be satisfied; because of flux expulsion, it can only do so in the thin layers along the bottom and top boundaries. Since the thickness of these layers scales as $R_m^{-1/2}$, it follows that the field strength therein scales as $\sqrt{R_m}$, which in turn implies that the total magnetic energy in the domain also scales as $\sqrt{R_m}$ in the $t \gg \tau_c$ limit.

0.20.3 Digression: the electromagnetic skin depth

You may recall that a sinusoidally oscillating magnetic field imposed at the boundary of a conductor will penetrate the conductor with an amplitude decreasing exponentially away from the boundary and into the conductor, with a length scale called the *electromagnetic skin depth*:

$$\ell = \sqrt{\frac{2\eta}{\omega}}. \quad (153) \quad \{\text{E2.260}\}$$

Now, go back to the cellular flow and imagine that you are an observer located in the center of the flow cell, looking at the domain boundaries while rotating with angular frequency $\propto u_0/L$; what you “see” in front of you is an “oscillating” magnetic field, in the sense that it flips sign with “angular frequency” u_0/L . The corresponding electromagnetic skin depth would then be

$$\frac{\ell}{L} = \sqrt{\frac{2\eta}{u_0 L}} \equiv \sqrt{\frac{2}{R_m}}. \quad (154) \quad \{\text{E2.261}\}$$

which basically corresponds to the thickness of the boundary layer where significant magnetic field is present in the steady-states shown on Figure 17. How about that for a mind flip...

0.20.4 Timescales for field amplification and decay

Back to our cellular flow. Flux expulsion or not, it is clear from Figure 14 (solid lines) that some level of field amplification has occurred in the high R_m solutions, in the sense that $\mathcal{E}_B(t \rightarrow \infty) > \mathcal{E}_B(0)$. But is this a dynamo? The solutions of Fig. 17 have strong electric currents in the direction perpendicular to the plane of the paper, and these currents are subjected to resistive dissipation. Have we then reached our goal, namely to amplify and maintain a weak, preexisting magnetic field against Ohmic dissipation?

In a narrow sense yes, but a bit of reflection will show that the boundary conditions are playing a crucial role. The only reason that the magnetic energy does not asymptotically go to zero is that the normal field component is held fixed at the boundaries, which, in the steady-state, implies a non-zero Poynting flux into the domain across the left and right vertical boundaries. The magnetic field is not avoiding resistive decay because of field induction within the domain, but rather because external energy (and magnetic flux) is being pumped in through the boundaries. This is precisely what is embodied in the first term on the RHS of eq. (84).

What if this were not the case? One way to work around the boundary problem is to replace the fixed flux boundary conditions by periodic boundary conditions on \mathbf{B} , which in terms of A becomes:

$$A(0, y) = A(L, y) , \quad \frac{\partial A(x, 0)}{\partial y} = \frac{\partial A(x, L)}{\partial y} . \quad (155) \quad \{\text{E2.262}\}$$

There is still a net flux across the horizontal boundaries at $t = 0$, but the boundary flux is now free to decay away along with the solution. Effectively, we now have an infinitely long row of contiguous flow cell initially threaded by a horizontal magnetic field extending to $\pm\infty$. It is time to reveal that the hitherto unexplained dotted lines on Fig. 14 correspond in fact to solutions computed with such boundary conditions, for the same cellular flow and initial condition as before. The magnetic energy now decays to zero, confirming that the boundaries indeed played a crucial role in the sustenance of the magnetic field in our previous solutions. What is noteworthy is the rate at which it does so. In the absence of the flow and with freely decaying boundary flux, the initial field would diffuse away on a timescale $\tau_\eta \sim L^2/\eta$, which is equal to R_m units of L/u_0 . With the flow turned on, the decay proceeds at an accelerated rate because of the inexorable decrease of the typical length scale associated with the evolving solution, which we argued earlier varied as t^{-1} . What then is the typical timescale for this enhanced dissipation? The decay phase of the field (for $t \gg L/u_0$) is approximately described by

$$\{\text{E2.263}\} \quad \frac{\partial A}{\partial t} = \eta \nabla^2 A . \quad (156)$$

An estimate for the dissipation timescale can be obtained once again via dimensional analysis, by replacing ∇^2 by $1/\ell^2$, as in §0.18 but with the important difference that ℓ is now a function of time:

$$\{\text{E2.264}\} \quad \ell \rightarrow \ell(t) = \left(\frac{L}{t}\right) \left(\frac{L}{u_0}\right) , \quad (157)$$

in view of our previous discussion (cf. Fig. 16 and accompanying text). This leads to

$$\{\text{E2.265}\} \quad \frac{\partial A}{\partial t} \simeq -\frac{\eta u_0^2 t^2}{L^4} A , \quad (158)$$

where the minus sign is introduced “by hand” in view of the fact that $\nabla^2 A < 0$ in the decay phase. Equation (158) integrates to

$$\{\text{E2.266}\} \quad \frac{A(t)}{A_0} = \exp \left[-\frac{\eta u_0^2}{3L^4} t^3 \right] = \exp \left[-\frac{1}{3R_m} \left(\frac{u_0^3 t^3}{L^3} \right) \right] . \quad (159)$$

This last expression indicates that with t measured in units of L/u_0 , the decay time scales as $R_m^{1/3}$. This is indeed a remarkable situation: in the low magnetic diffusivity regime (i.e., high R_m), the flow has in fact *accelerated* the decay of the magnetic field, even though large field intensification can

occur in the early, transient phases of the evolution. This is not at all what a dynamo should be doing!

As it turns out, flux expulsion is even trickier than the foregoing discussion may have led you to believe! Flux expulsion destroys the mean magnetic field component directed *perpendicular* to the flow streamlines. It cannot do a thing to a mean component oriented *parallel* to streamlines. For completely general flow patterns and initial conditions, the dissipative phase with timescale $\propto R_m^{1/3}$ actually characterizes the approach to a state where the advected trace quantity—here the vector potential A —becomes constant *along each streamline*, at a value \bar{A} equal to the initial value of A averaged on each of those streamlines. For the cellular flow and initial conditions used above, this average turns out to be $\bar{A} = 0.5$ *for every streamline*, so that the $R_m^{1/3}$ decay phase corresponds to the true decay of the magnetic field to zero amplitude. If \bar{A} varies from one fieldline to the next, however, the $R_m^{1/3}$ phase is followed by a third decay phase, which proceeds on a timescale $\sim R_m$, since induction no longer operates ($\mathbf{u} \cdot \nabla A = 0$) and the typical length scale for A is once again L . At any rate, even with a more favorable initial condition we have further delayed field dissipation, but we still don't have a dynamo since dissipation will proceed inexorably, at best on the “long” timescale $R_m \times (L/u_0)$.

0.20.5 Flux expulsion in spherical geometry: axisymmetrization

{pch_ssec:axisymm}

You may think that the flux expulsion problem considered in the preceding section has nothing to do with any astronomical objects you are likely to encounter in your astrophysical careers. Wrong!

Consider the evolution of a magnetic field pervading a sphere of electrically conducting fluid, with the solar-like differential rotation profile already encountered previously (§0.19.4 and eqs. (139)–(140)), and with the field having initially the form of a dipole whose axis is inclined by an angle Θ with respect to the rotation axis ($\theta = 0$). Such a magnetic field can be expressed in terms of a vector potential having components:

$$A_r(r, \theta, \phi) = 0 \quad (160) \quad \{\text{eq:2.69a}\}$$

$$A_\theta(r, \theta, \phi) = (R/r)^2 \sin \Theta (\sin \beta \cos \phi - \cos \beta \sin \phi) \quad (161) \quad \{\text{eq:2.69b}\}$$

$$A_\phi(r, \theta, \phi) = (R/r)^2 [\cos \Theta \sin \theta - \sin \Theta \cos \theta (\cos \beta \cos \phi + \sin \beta \sin \phi)] \quad (162) \quad \{\text{eq:2.69c}\}$$

where β is the azimuthal angle locating the projection of the dipole axis on the equatorial plane.

Now, the vector potential for an inclined dipole can be written as the sum of two contributions, the first corresponding to an aligned dipole ($\Theta = 0$), the second to a perpendicular dipole ($\Theta = \pi/2$), their relative magnitude being

equal to $\tan\Theta$. Since the governing equation is linear, the solution for an inclined dipole can be broken into two independent solutions for the aligned and perpendicular dipoles. The former is precisely what we investigated already in §0.19.4, where we concluded there that the shearing of an aligned dipole by an axisymmetric differential rotation would lead to the buildup of a toroidal component, whose magnitude would grow linearly in time at a rate set by the magnitude of the shear.

The solution for a perpendicular dipole is in many way similar to the cellular flow problem of §0.20. You can see how this may be the case by imagining looking from above onto the equatorial plane of the sphere; the fieldlines contained in that plane will have a curvature and will be contained within a circular boundary, yet topologically the situation is similar to the cellular flow studied in the preceding section: the (sheared) flow in the equatorial plane is made of closed, circular streamlines contained within that plane, so that we can expect flux expulsion to occur. The equivalent of the turnover time here is the differential rotation timescale, namely the time for a point located on the equator to perform a full 2π revolution with respect to the poles:

$$\tau_{\text{DR}} = (\Omega_{\text{Equ}} - \Omega_{\text{Pole}})^{-1} = \Omega_{\odot}^{-1}(a_2 + a_4) , \quad (163)$$

where the second equality follows directly from eq. (140). For a freely decaying dipole, the perpendicular component of the initial dipole will then be subjected to flux expulsion, and dissipated away, at a rate far exceeding purely diffusive decay in the high R_m limit, as argued earlier.

But here is the amusing thing; for an observer looking at the magnetic field at the surface of the sphere, the enhanced decay of the perpendicular component of the dipole will translate into a gradual decrease in the inferred tilt axis of the dipole. Figure 18 shows this effect, for the differential rotation profile given by eq. (139) and a magnetic Reynolds number $R_m = 10^3$. Contours of constant B_r are plotted on the surface $r/R = 1$, with the neutral line ($B_r = 0$) plotted as a thicker line. At $t = 0$ the field has the form of a pure dipole tilted by $\pi/3$ with respect to the coordinate axis, and the sphere is oriented so that the observer (you!) is initially looking straight down the magnetic axis of the dipole. Advection by the flow leads to a distortion of the initial field, with the subsequent buildup of small spatial scales in the r - and θ -directions (only the latter can be seen here)¹². After only two turnover times (last frame), the surface field looks highly axisymmetric.

So, in a differentially rotating fluid system with high R_m , flux expulsion leads to the *symmetrization* of any non-axisymmetric magnetic field component initially present —or contemporaneously generated. The efficiency of the symmetrization process should make us a little cautious in assuming that the large-scale magnetic field of the Sun, which one would deem roughly

¹² An animation of this solution, as well as a few others for different R_m and/or tilt angle, can be viewed on the course Web Page.

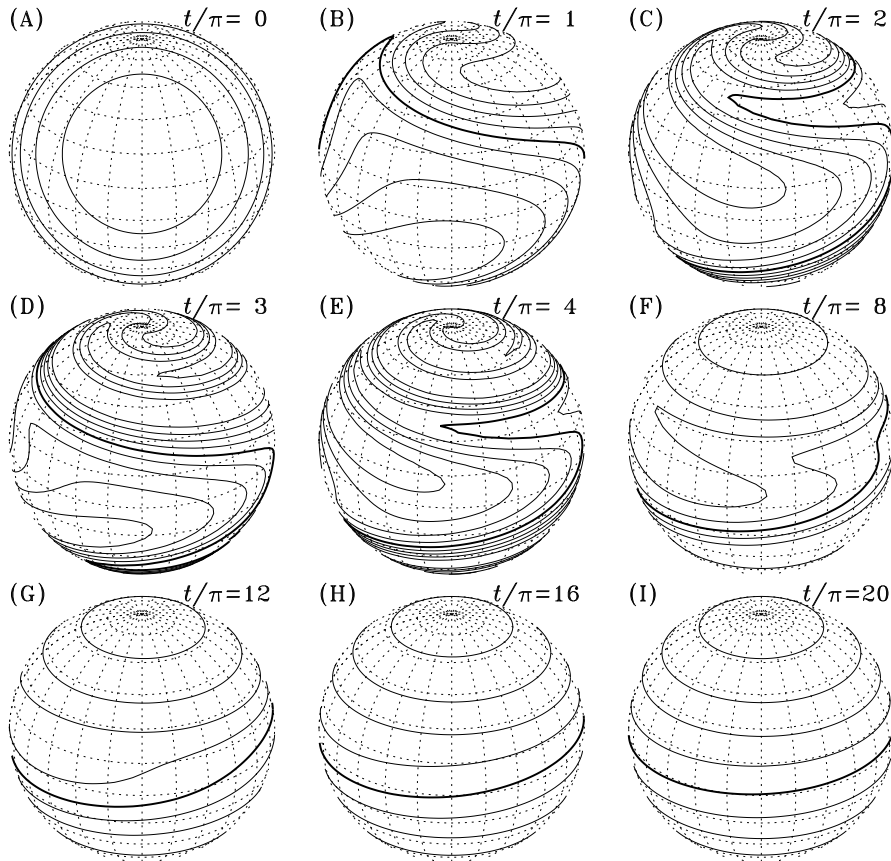


Fig. 18 ^{F2.13} Symmetrization of an inclined dipole in a electrically conducting sphere in a state of solar-like axisymmetric differential rotation. Each panel shows contours of constant B_r at the surface of the sphere, and the solution is matched to a potential in the exterior ($r/R > 1$). The differential rotation is given by eq. (139). Time is given in units of τ_{DR} , in which the turnover period (or differential rotation period) is equal to 2π , and the magnetic Reynolds number is $R_m = 10^3$.

axisymmetric upon consideration of surface things like the sunspot butterfly diagram, is characterized by the same level of axisymmetry in the deep-seated generating layers, where the dynamo is presumed to operate. After all, standing in between is a thick, axisymmetrically differentially rotating convective envelope that must be reckoned with. In fact, observations of coronal density structures in the descending phase of the solar cycle can be interpreted in terms of a large-scale, tilted dipole component, with the tilt angle steadily decreasing over 3–4 years towards solar minimum. Interestingly, the differential rotation timescale for the Sun is ~ 6 months. Are we seeing the axisymmetrization process in operation? Maybe. Axisymmetry is certainly

a very convenient modeling assumption when working on the large scales of the solar magnetic field, but it may be totally wrong.

Axisymmetrization has also been invoked as an explanation for the almost perfectly axisymmetric magnetic field of the planet Saturn, which stands in stark contrast to the other solar system planetary magnetic fields.

{Santidyn} 0.21 Two anti-dynamo theorems

The cellular flow studied in §0.20, although it initially looked encouraging (cf. Fig. 14), proved not to be a dynamo after all. Is this peculiar to the flow defined by eqs. (146)–(147), or is this something more general? Exhaustively testing for dynamo action in all possible kinds of flow geometries is clearly impractical. However, it turns out that one can rule out *a priori* dynamo action in many classes of flows. These demonstrations are known as *anti-dynamo theorems*.

{pch_ssec:zeldovich} 0.21.1 Zeldovich’s theorem

A powerful anti-dynamo theorem due to Ya. B. Zeldovich (1914-87), has a lot to teach us about our cellular flow results. The theorem rules out dynamo action in steady planar flows in cartesian geometry, i.e., flows of the form

$$\{\text{Eadt1}\} \quad \mathbf{u}_2(x, y, z) = u_x(x, y, z)\hat{\mathbf{e}}_x + u_y(x, y, z)\hat{\mathbf{e}}_y \quad (164)$$

in a bounded volume V at the boundaries (∂V) of which the magnetic field vanishes. Note that no other restrictions are placed on the magnetic field, which can depend on all three spatial coordinate as well as time. Nonetheless, in view of eq. (164) it will prove useful to consider separately the z -component of the magnetic field $B_z(x, y, z, t)$ from the (2D) field component in the $[x, y]$ plane (hereafter denoted \mathbf{B}_2). It is readily shown that the z -component of the induction equation then reduces to

$$\{\text{Eadt2}\} \quad \left(\frac{\partial}{\partial t} + \mathbf{u} \cdot \nabla \right) B_z = \eta \nabla^2 B_z \quad (165)$$

for spatially constant magnetic diffusivity. Now, the LHS is just a Lagrangian derivative, yielding the time variation of B_z as one moves along with the fluid. Multiplying this equation by B_z and integrating over V yields, after judicious use of a suitable vector identity and of the divergence theorem:

$$\{\text{Eadt3}\} \quad \frac{1}{2} \int_V \frac{DB_z^2}{Dt} dV = \int_{\partial V} B_z (\nabla B_z) \cdot \mathbf{nd}S - \eta \int_V (\nabla B_z)^2 dV . \quad (166)$$

Now, the first integral on the RHS vanishes since $\mathbf{B} = 0$ on ∂V by assumption. The second integral is positive definite, *therefore B_z always decays on the diffusive timescale* (cf. §0.18).

Consider now the magnetic field \mathbf{B}_2 in $[x, y]$ planes. The most general such 2D field can be written as the sum of a solenoidal and potential component:

$$\mathbf{B}_2(x, y, z, t) = \nabla \times (A\hat{\mathbf{e}}_z) + \nabla\Phi, \quad (167) \quad \{\text{Eadt4}\}$$

where the vector potential A and scalar potential Φ both depend on all three spatial coordinates and time. Evidently, the constraint $\nabla \cdot \mathbf{B} = 0$ implies

$$\nabla_2^2 \Phi = -\frac{\partial B_z}{\partial z}, \quad (168) \quad \{\text{Eadt5}\}$$

where $\nabla_2^2 \equiv \partial^2/\partial x^2 + \partial^2/\partial y^2$ is the 2D Laplacian operator in the $[x, y]$ plane. Clearly, once B_z has resistively dissipated, i.e., for times much larger than the global resistive decay time τ_η , Φ is simply a solution of the 2D Laplace equation $\nabla_2^2 \Phi = 0$.

Here comes the sneaky part. We take the curl of the induction equation. Upon substituting eq. (167), the z -component of the resulting expression yields

$$\nabla \times \nabla \times \left[\frac{\partial(A\hat{\mathbf{e}}_z)}{\partial t} + \mathbf{u}_2 \cdot \nabla(A\hat{\mathbf{e}}_z) - \eta \nabla_2^2(A\hat{\mathbf{e}}_z) - \mathbf{u}_2 \times \nabla\Phi \right] = 0, \quad (169)$$

with $\nabla \cdot (A\hat{\mathbf{e}}_z) = 0$ as a choice of gauge. Note that only one term involving Φ survives, because $\nabla \times \nabla\Phi = 0$ identically. In general, the above expression is only satisfied if the quantity in square brackets itself vanishes, i.e.,

$$\left(\frac{\partial}{\partial t} + \mathbf{u} \cdot \nabla \right) A = \eta \nabla_2^2 A + (\mathbf{u}_2 \times \nabla\Phi) \cdot \hat{\mathbf{e}}_z. \quad (170) \quad \{\text{Eadt6}\}$$

This expression is identical to that obtained above for B_z , except for the presence of the source term $\mathbf{u}_2 \times \nabla\Phi$. However, we just argued that for $t \gg \tau_\eta$, $\nabla_2^2 \Phi = 0$. In addition, \mathbf{B} vanishes on ∂V by assumption, so that the only possible asymptotic interior solutions are of the form $\Phi = \text{const}$, which means that the source term vanishes in the limit $t \gg \tau_\eta$. From this point on eq. (170) is indeed identical to eq. (165), for which we already demonstrated the inevitability of resistive decay. Therefore, dynamo action, i.e., maintenance of a magnetic field against resistive dissipation, is impossible in a planar flow for *any* 3D magnetic field.

0.21.2 Cowling's theorem

`{pch_ssec:cowling}`

Another powerful anti-dynamo theorem, predating in fact Zeldovich's, is due to T.G. Cowling (1906-90). This anti-dynamo theorem is particularly impor-

tant historically, since it rules out dynamo action for 3D but axisymmetric flows and magnetic fields, which happen to be the types of flows and fields one sees in the Sun, at least on the larger spatial scales. Rather than going over one of the many formal proofs of Cowling's theorem found in the literature, let's just follow the underlying logic of our proof of Zeldovich's theorem.

Assuming once again that there are no sources of magnetic field exterior to the domain boundaries, we consider the inductive action of a 3D, steady axisymmetric flow on a 3D axisymmetric magnetic field. Working in spherical polar coordinates (r, θ, ϕ) , we write:

$$\{\text{E2.70a}\} \quad \mathbf{u}(r, \theta) = \frac{1}{\rho} \nabla \times (\Psi(r, \theta) \hat{\mathbf{e}}_\phi) + \varpi \Omega(r, \theta) \hat{\mathbf{e}}_\phi \quad (171)$$

$$\{\text{E2.70bis}\} \quad \mathbf{B}(r, \theta, t) = \nabla \times (A(r, \theta, t) \hat{\mathbf{e}}_\phi) + B(r, \theta, t) \hat{\mathbf{e}}_\phi \quad (172)$$

where $\varpi = r \sin \theta$. Here the vector potential component A and stream function Ψ defines the poloidal components of the flow, and Ω is the angular velocity (units rad s^{-1}). Note that the form of eq. (171) guarantees $\nabla \cdot (\rho \mathbf{u}) = 0$, describing mass conservation in a steady flow. Separation of the (vector) MHD induction equation into two components for the 2D scalar fields A and B , as done in §0.18, now leads to:

$$\{\text{eq:cowa}\} \quad \left(\frac{\partial}{\partial t} + \mathbf{u}_p \cdot \nabla \right) (\varpi A) = \varpi \eta \left(\nabla^2 - \frac{1}{\varpi^2} \right) A, \quad (173)$$

$$\left(\frac{\partial}{\partial t} + \mathbf{u}_p \cdot \nabla \right) \left(\frac{B}{\varpi} \right) = \frac{\eta}{\varpi} \left(\nabla^2 - \frac{1}{\varpi^2} \right) B + \frac{1}{\varpi^2} \frac{d\eta}{dr} \frac{\partial(\varpi B)}{\partial r}$$

$$\{\text{eq:cowb}\} \quad - \left(\frac{B}{\varpi} \right) \nabla \cdot \mathbf{u}_p + \mathbf{B}_p \cdot \nabla \Omega, \quad (174)$$

where \mathbf{B}_p and \mathbf{u}_p are notational shortcuts for the poloidal field and meridional flow. Notice that the vector potential A evolves in a manner entirely independent of the toroidal field B , the latter being conspicuously absent on the RHS of eq. (173). This is not true of the toroidal field B , which is well aware of the poloidal field's presence via the $\nabla \Omega$ shearing term.

The LHS of these expressions is again a Lagrangian derivative for the quantities in parentheses, and the first terms on each RHS are of course diffusion. The next term on the RHS of eq. (174) vanishes for incompressible flows, and remains negligible for very subsonic compressible flows. The last term on the RHS, however, is a source term, in that it can lead to the growth of B as long as A does not decay away. This is the very situation we have considered in §0.19.4, by holding A fixed as per eq. (135). However, there

is no similar source-like term on the RHS of eq. (173), which governs the evolution of A .

This should now start to remind you of Zeldovich's theorem. In fact, eq. (115) is structurally identical to eq. (165), for which we demonstrated the inevitability of resistive decay in the absence of sources exterior to the domain. This means that A will inexorably decay, implying in turn that B will then also decay once A has vanished. Since axisymmetric flows cannot maintain A against Ohmic dissipation, *a 3D axisymmetric flow cannot act as a dynamo for a 3D axisymmetric magnetic field.*¹³ Cowling's theorem is not restricted to spherical geometry, and is readily generalized to any situation where both flow and field showing translational symmetry in one and the same spatial coordinate. Such physical systems are said to have an *ignorable coordinate*.

It is worth pausing and reflecting on what these two antidynamo theorems imply for the cellular flow of §0.20. It was indeed a planar flow ($u_z = 0$), and moreover the magnetic field had an ignorable coordinate ($\partial\mathbf{B}/\partial z \equiv 0$)! We thus fell under the purview of both Zeldovich's and Cowling's theorems, so in retrospect our failure to find dynamo action is now understood.

0.22 The Roberts cell dynamo

{pch_sec:rbc}

Clearly, the way to evade both theorems is to consider flows and fields that are fully three-dimensional, and lack translational symmetry in either the flow or the magnetic field. We now consider one such flow, and examine some of its dynamo properties.

0.22.1 The Roberts cell

The Roberts cell is a spatially periodic, incompressible flow defined over a 2D domain $(x, y) \in [0, 2\pi]$ in terms of a stream function

$$\Psi(x, y) = \cos x + \sin y. \quad (175) \quad \{\text{E3.rc1b}\}$$

so that

$$\mathbf{u}(x, y) = \frac{\partial\Psi(x, y)}{\partial y} \hat{\mathbf{e}}_x - \frac{\partial\Psi(x, y)}{\partial x} \hat{\mathbf{e}}_y + \Psi(x, y) \hat{\mathbf{e}}_z \quad (176) \quad \{\text{E3.rc1a}\}$$

Note that the flow velocity is independent of the z -coordinate, even though the flow has a non-zero z -component. Equations (176)–(175) describes a periodic array of counterrotating flow cells in the $[x, y]$ plane, with a z -component

¹³ A fact often unappreciated is that Cowling's theorem does not rule out the dynamo generation of a *non-axisymmetric* 3D magnetic field by a 3D axisymmetric flow.

that changes sign from one cell to the next; the total flow is then a series of helices, which have the same kinetic helicity $h = \mathbf{u} \cdot \nabla \times \mathbf{u}$ in each cell. The Roberts cell flow represents one example of a *Beltrami flows*, i.e., it satisfies the relation $\nabla \times \mathbf{u} = \alpha \mathbf{u}$, where α is a numerical constant. Such flows are *maximally helical*, in the sense that their vorticity ($\boldsymbol{\omega} \equiv \nabla \times \mathbf{u}$) is everywhere parallel to the flow, which maximizes helicity for a given flow speed. Figure 19 shows one periodic “unit” of the the Roberts cell flow pattern. Take note already of the presence stagnation points where the corners of four contiguous flow cells meet.

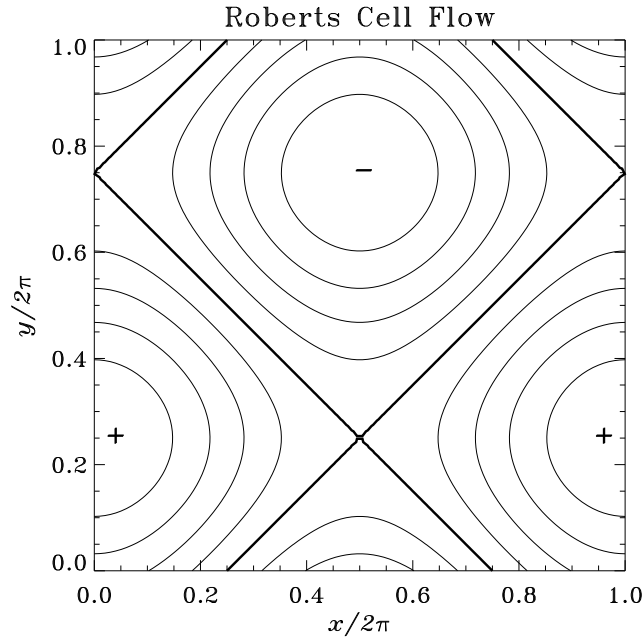


Fig. 19 ^{F3.2} The Roberts cell flow. The flow is periodic in the $[x, y]$ plane, and independent of the z -coordinate (but $u_z \neq 0!$). Flow streamlines are shown projected in the $[x, y]$ plane, and the $+/-$ signs indicate the direction of the z -component of the flow. The thicker contour defines the network of *separatrix surfaces* in the flow, corresponding to cell boundaries and intersecting at stagnation points. The $u_z(x, y)$ isocontours coincide with the projected streamlines.

Let's first pause and consider why one should expect the Roberts cell to evade Cowling's and Zeldovich's theorems. First, note that this is not a planar flow in the sense demanded by Zeldovich's theorem, since we do have three non-vanishing flow components. However, the z -coordinate is ignorable in the sense of Cowling's theorem, since all flow components are independent of z . *If this flow is to evade Cowling's theorem and act as a dynamo, it must act on a magnetic field that is dependent on all three spatial coordinates.*

Consequently, we consider the inductive effects of this flow acting on a fully three dimensional magnetic field $\mathbf{B}(x, y, z, t)$. Since the flow speed is

independent of z , we can expect solutions of the linear induction equation to be separable in z , i.e.:

$$\mathbf{B}(x, y, z, t) = \mathbf{b}(x, y, t)e^{ikz} \quad (177) \quad \{\text{E3.r.c2}\}$$

where k is a (specified) wavevector in the z -direction, and the 2D magnetic amplitude \mathbf{b} is now a complex quantity. We are still dealing with a fully 3D magnetic field, but the problem has been effectively reduced to two spatial dimensions (x, y) , which represents a great computational advantage.

0.22.2 Dynamo solutions

From the dynamo point of view, the idea is again to look for solutions of the induction equations where the magnetic energy does not fall to zero as $t \rightarrow \infty$. In practice this means specifying k , as well as some weak field as an initial condition, and solve the 2D linear initial value problem for $\mathbf{b}(x, y, t)$ resulting from the substitution of eq. (177) into the induction equation:

$$\frac{\partial \mathbf{b}}{\partial t} = (\mathbf{b} \cdot \nabla_{xy})\mathbf{u} - (\mathbf{u} \cdot \nabla_{xy})\mathbf{b} - ik u_z \mathbf{b} + R_m^{-1}(\nabla_{xy}^2 \mathbf{b} - k^2 \mathbf{b}) \quad (178) \quad \{\text{E3.r.c3}\}$$

subjected to periodic boundary conditions on \mathbf{b} , in order to avoid the potentially misleading role of fixed-flux boundary conditions in driving dynamo action, as encountered in §0.20. Here ∇_{xy} and ∇_{xy}^2 are the 2D gradient and Laplacian operators in the $[x, y]$ plane. As before we use as a time unit the turnover time τ_c , which is of order 2π here. All solutions described below were again obtained numerically, starting from a weak, horizontal magnetic field as the initial condition.

The time evolution of the can be divided into three more or less distinct phases, the first two being similar to the case of the 2D cellular flow considered in the preceding chapter: (1) quadratic growth of the magnetic energy for $t \lesssim \tau_c$; (2) flux expulsion for the subsequent few τ_c . However, and unlike the case considered in §0.20, for some values of k the third phase is one of *exponential growth* in the magnetic field (and energy).

Figure 20 shows a typical Roberts cell dynamo solution, for $R_m = 10^2$ and $k = 2$. What is plotted is the real part of the z -component of $\mathbf{b}(x, y, t)$, at time $t \gg \tau_c$. The thick green lines are the separatrices of the flow. One immediately recognizes the workings of flux expulsion, in that very little magnetic flux is present near the center of the flow cells. Instead the field is concentrated in thin sheets parallel to the separatrix surfaces. Given our extensive discussion of flux expulsion in the preceding chapter, it should come as no surprise that the thickness of those sheets scales as $R_m^{-1/2}$. For $t \gg \tau_c$, the field grows exponentially, but the shape of the “planform” remains fixed. In other words, even though we solved the induction equation as an initial

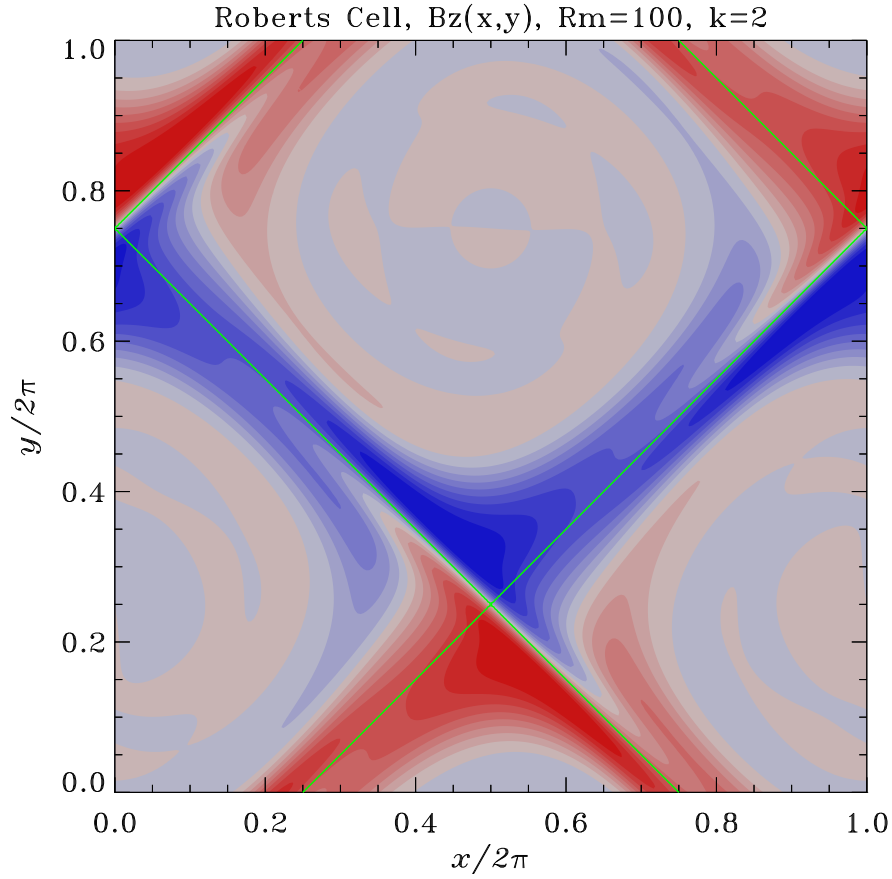


Fig. 20 ^{F3.3} Isocontours for the z -component of the magnetic field in the $[x, y]$ plane, for Roberts cell dynamo solutions with $R_m = 100$ and $k = 2$, in the asymptotic regime $t \gg \tau_c$. The color scale codes the real part of the z -component of $\mathbf{b}(x, y, t)$ (gray-to-blue is negative, gray-to-red positive). The green straight lines indicate the separatrix surfaces of the flow (see Fig. 19). Note the flux expulsion from the cell centers, and the concentration of the magnetic flux in thin sheets pressed against the separatrices. In the $t \gg \tau_c$ regime, the field grows exponentially but the shape of the planform is otherwise steady. Compare this with Fig. 17A and B.

value problem, the solution can be thought of as an eigensolution of the form $\mathbf{B}(x, y, z, t) = \mathbf{b}(x, y)e^{ikz+st}$, with $\text{Re}(s) > 0$ and $\text{Im}(s) = 0$.

In terms of the magnetic energy evolution, the growth rate s of $\mathbf{b}(x, y, t)$ is readily obtained by a linear least-squares fit to the $\log \mathcal{E}_B$ vs t curves in the $t \gg \tau_c$ regime, or more formally defined as

$$\{E3.rc4\} \quad s = \lim_{t \rightarrow \infty} \left[\frac{1}{2t} \log(\mathcal{E}_B) \right]. \quad (179)$$

It turns out that the Roberts cell flows yields dynamo action (i.e., $s > 0$) over wide ranges of wavenumbers k and magnetic Reynolds number R_m . Figure 21 shows the variations in growth rates with k , for various values of R_m . The curves peak at a growth rate value k_{\max} that gradually shifts to higher k as R_m increases. The largest growth rate is $k_{\max} \simeq 0.17$, and occurs at $R_m \simeq 10$. It can be shown (see bibliography) that in the high R_m regimes the following scalings hold:

$$k_{\max} \propto R_m^{1/2}, \quad R_m \gg 1, \quad (180) \quad \{\text{E3.rc5a}\}$$

$$s(k_{\max}) \propto \frac{\log(\log R_m)}{\log R_m}, \quad R_m \gg 1. \quad (181) \quad \{\text{E3.rc5b}\}$$

To understand the origin of these peculiar scaling relations, we need to take

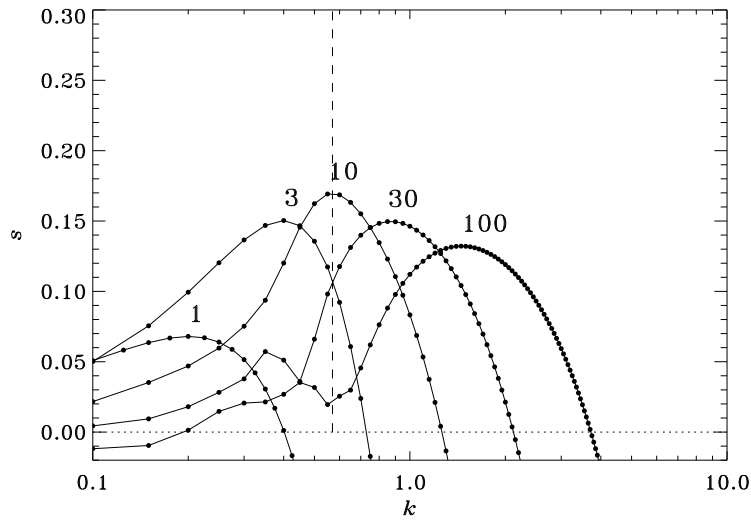


Fig. 21 ^{E3.5} Growth rates of the magnetic energy in the Roberts cell, for sequences of solutions with increasing k and various values of R_m , as labeled near the maxima of the various curves. Growth typically occurs for a restricted range in k , and peaks at a value k_{\max} that increases slowly with increasing R_m . Note however how the corresponding maximum growth rate decreases with increasing R_m . The small “dip” left of the main peaks for the high- R_m solutions is a real feature, although here it is not very well resolved in k .

a closer look at the mechanism through which the magnetic field is amplified by the Roberts cell.

0.22.3 Exponential stretching and stagnation points

{ssec:stagn}

Even cursory examination of Figure 20 suggests that magnetic field amplification in the Roberts cell is somehow associated with the network of separatrices and stagnation points. It will prove convenient in the foregoing analysis and discussion to first introduce new coordinates

$$\{\text{E3.rc6a}\} \quad x' = x - y, \quad y' = x + y + \frac{3\pi}{2}, \quad (182)$$

corresponding to a $3\pi/2$ translation in the y -direction, followed by 45° rotation about the origin in the $[x, y]$ plane. The separatrices are now parallel to the coordinate lines $x' = n\pi$, $y' = n\pi$ ($n = 0, 1, \dots$), and the stream function has become

$$\{\text{E3.rc6b}\} \quad \Psi(x', y') = 2 \sin(x') \sin(y'). \quad (183)$$

Close to the stagnation points, a good approximation to eq. (183) is

$$\{\text{E3.rc6c}\} \quad \Psi(x', y') \simeq 2x'y', \quad x', y' \ll 1 \quad (184)$$

which, if anything else, should now clarify why this is called a hyperbolic stagnation point... Consider now a fluid element flowing in the vicinity of this stagnation point. From a Lagrangian point of view its equations of motion are:

$$\{\text{E3.rc7a}\} \quad \frac{\partial x'}{\partial t} = u_{x'} = 2x', \quad (185)$$

$$\{\text{E3.rc7b}\} \quad \frac{\partial y'}{\partial t} = u_{y'} = -2y', \quad (186)$$

which immediately integrates to

$$\{\text{E3.rc7c}\} \quad x'(t) = x'_0 e^{2t}, \quad y'(t) = y'_0 e^{-2t}, \quad (187)$$

where (x'_0, y'_0) is the location of the fluid element at $t = 0$. Evidently, the fluid element experiences *exponential stretching* in the x' -direction, and corresponding contraction in the y' -direction (since $\nabla \cdot \mathbf{u} = 0!$). Now, recall that in ideal MHD ($R_m = \infty$) a magnetic fieldline obeys an equation identical to that of a line element, and that stretching leads to field amplification as per the mass conservation constraint (§0.19.1). Evidently stagnation point have quite a bit of potential, when it comes to amplifying exponentially a pre-existing magnetic field... provided diffusion and destructive folding can be held at bay. Let's look into how this is achieved in the Roberts Cell.

0.22.4 Mechanism of field amplification in the Roberts cell

We stick to the rotated Roberts cell used above, restrict ourselves to the $R_m \gg 1$ regime, and pick up the field evolution after flux expulsion is completed and the magnetic field is concentrated in thin boundary layers (thickness $\propto R_m^{-1/2}$) pressed against the separatrices (as on Fig. 20).

Consider a x' -directed magnetic fieldline crossing a vertical separatrix, as shown on Figure 22A (gray line labeled “a”). the y' component of the flow is positive on either side of the separatrix, and peaks on the separatrix. Consequently, the fieldline experiences stretching in the y' -direction ($a \rightarrow b \rightarrow c \rightarrow d$ on Fig. 22A). However, the induced y' component of the magnetic field changes sign across the separatrix, so that we seem to be heading towards our dreaded destructive folding. This is where the crucial role of the vertical (z) dimension becomes apparent. Figure 22B is a view of the same configuration in the $[x', z]$ plane, looking down onto the y' axis on part A. At $t = 0$ the fieldlines have no component in the z -direction, but in view of the assumed e^{ikz} spatial dependency the x' component changes sign every half-wavelength k/π . Consider now the inductive action of the z -component of the velocity, which changes sign across the separatrix. After some time interval of order $k/(\pi u_z)$ the configuration of Fig. 22B will have evolved to that shown on part C. Observe what has happened: the fieldlines have been sheared in such a way that y' -components of the magnetic field of like signs have been brought in close proximity. Contrast this to the situation on part B, where magnetic footpoints in closest proximity have oppositely directed y' -components.

The end result of this process is that a y' -directed magnetic field is produced by shearing of the initial x' -directed field, with a phase shift in the z -direction such that destructive folding is avoided. Clearly, this requires *both* a z -component of velocity, *and* a z -dependency in the magnetic field. Either alone won't do the trick.

Now, the same reasoning evidently applies to a y' -directed magnetic fieldline crossing a horizontal separatrix: a x' -directed magnetic field will be induced. That magnetic field will be swept along the horizontal separatrix, get further amplified by exponential stretching as it zooms by the stagnation point, and continue along the vertical separatrix, where it can now serve as a seed field for the production of a y' -directed field. The dynamo “loop” is closed, at any time the rate of field production is proportional to the local field strength, and exponential growth of the field follows. The process works best if the half wavelength k/π is of order of the boundary layer thickness, which in fact is what leads to the scaling law given by eq. (180). The scaling for the growth rate (eq. (181)), in turn, is related to the time spent by a fluid element in the vicinity of the stagnation point.

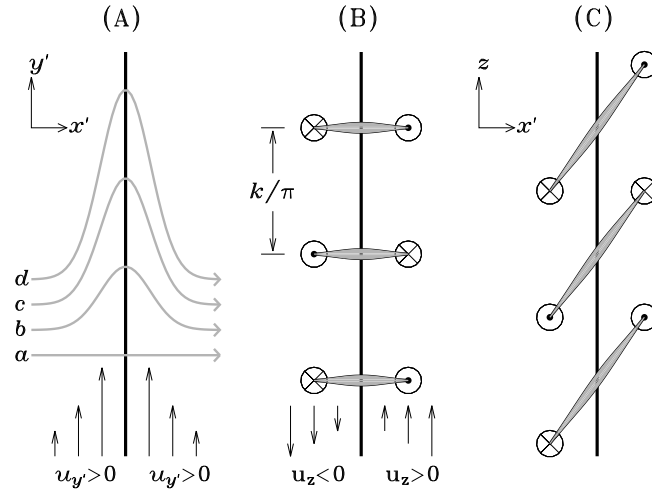


Fig. 22 $\mathbb{F}^{3.6}$ Mechanism of magnetic field amplification in the Roberts cell flow. The diagram is plotted in terms of the rotated $[x', y']$ Roberts cell. The thick vertical line is a separatrix surface, and the gray lines are magnetic fieldlines. Part (A) is a view in the horizontal plane $[x', y']$, and shows the production of a y' -directed magnetic component from an initially x' -directed magnetic field (line labeled “a”). Parts (B) and (C) are views in the $[x', z]$ plane looking down along the y' axis, and illustrate the phase shift in the z -direction of the y' magnetic component caused by the z -component of the velocity. The symbol \odot (\otimes) indicates a magnetic field coming out (into) the plane of the page. Note on part (C) how footpoints of identical polarity are brought in close proximity, thus avoiding the destructive folding that would have otherwise characterized the situation depicted on part B in the $u_z = 0$ 2D case.

0.22.5 Fast versus slow dynamos

One worrisome aspect of the Roberts cell dynamo is the general decrease of the growth rates with increasing R_m (see Fig. 21); worrisome, because the $R_m \rightarrow \infty$ limit is the one relevant to most astrophysically interesting circumstances. A dynamo exhibiting this property is called a *slow dynamo*, in contrast to a *fast dynamo*, which (by definition) retains a finite growth rate as $R_m \rightarrow \infty$, the formal requirement being that

$$\lim_{R_m \rightarrow \infty} s(k_{\max}) > 0. \quad (188)$$

In view of eq. (181), the Roberts cell is thus formally a slow dynamo. However the RHS of eq. (181) is such a slowly decreasing function of R_m that the Roberts cell is arguably the closest thing it could be to a fast dynamo... without formally being one. The distinction hinges on the profound differences between the strict mathematical case of $R_m = \infty$ (ideal MHD), and the more physically relevant limit $R_m \rightarrow \infty$. From the physical point of view, the

distinction is a crucial one. One example will suffice. Recall that in the absence of dissipation magnetic helicity is a conserved quantity in any evolving magnetized fluid:

$$\frac{d\mathcal{H}_B}{dt} = \frac{d}{dt} \int_V \mathbf{A} \cdot \mathbf{B} dV = 0, \quad (189) \quad \{\text{E3.f.d1}\}$$

where $\mathbf{B} = \nabla \times \mathbf{A}$. Dynamo action, in the sense of amplifying a weak initial field, is then clearly impossible except for the subset of initial fields having $\mathcal{H}_B = 0$. This is a very stringent constraint on dynamo action! Go back now to the Roberts cell dynamo in the high- R_m regime. We saw that magnetic structures build up on a horizontal length scale $\propto R_m^{-1/2}$, and that the vertical wavelength of the fastest growing mode also decreases as $R_m^{-1/2}$. *The inexorable shrinking of the length scales ensures that dissipation always continues to operate even in the $R_m \rightarrow \infty$ limit.* This is why the Roberts cell dynamo can evade the constraint of helicity conservation. This is also why it is a slow dynamo. On the other hand, the Vainshtein & Zeldovich Stretch-Twist-Fold dynamo of §0.19, with its growth rate $\sigma = \ln 2$, is a fast dynamo since nothing prevents it from operating in the $R_m \rightarrow \infty$ limit.

But is this really the case? In the flows we have considered up to now, the existence of dynamo action hinges on stretching winning over destructive folding; in the 2D cellular flow of §0.20, destructive folding won over stretching everywhere away from boundaries. In the Roberts cell, destructive folding is avoided only for vertical wavenumbers such that magnetic fields of like signs are brought together, minimizing dissipation. The STF dynamo actually combines stretching and constructive folding, such that folding *reinforces* stretching. The fact that destructive folding is avoided entirely is why the growth rate does not depend on R_m .

Well, upon further consideration it turns out that magnetic diffusivity must play a role in the STF rope dynamo after all. Diffusion comes in at two levels; the first and most obvious one is at the “crossings” formed by the STF sequence. The second and less obvious arises from the fact that as one applies the STF operation n times, the resulting “flux rope” is in fact made up of n closely packed flux ropes, each of cross-section $\propto 2^{-n}$ times smaller than the original circular flux rope, so that the total cross-section looks more like a handful of spaghettis than it does a single monolithic flux rope of strength $\propto 2^n$. If one waits long enough, the magnetic length scale perpendicular to the loop axis shrinks to zero, so that even in the $R_m \rightarrow \infty$ limit dissipation is bound to come into to play.

0.23 The CP flow and fast dynamo action

{pch_sec:CPflow}

It turns out that a simple modification of the Roberts cell flow can turn it into a true fast dynamo. The co-called CP flow (for “Circularly Polarized”) is nothing more than the original Roberts cell flow, with a forced time-dependence. It is once again a spatially periodic, incompressible flow, defined in cartesian coordinate over a 2D domain $(x, y) \in [0, 2\pi]$:

$$\{\text{E3.cp1a}\} \quad u_x(x, y, t) = A \cos(y + \epsilon \sin \omega t) , \quad (190)$$

$$\{\text{E3.cp1b}\} \quad u_y(x, y, t) = C \sin(x + \epsilon \cos \omega t) , \quad (191)$$

$$\{\text{E3.cp1c}\} \quad u_z(x, y, t) = A \sin(y + \epsilon \sin \omega t) + C \cos(x + \epsilon \cos \omega t) . \quad (192)$$

Although the CP flow is not expressed here in terms of a stream function, this is the same as the Roberts Cell flow, except that now the counterrotating flow cells are “precessing” in unison in the $[x, y]$ plane, along circular paths of radius ϵ , undergoing a full revolution in a time interval $2\pi/\omega$. Here and in what follows we set $\omega = 1$, $\epsilon = 1$, $A = C = \sqrt{3/2}$, without any loss of generality.

0.23.1 Dynamo solutions

The CP flow has the same spatial symmetry properties as the Roberts cell, and in particular is invariant in the z -direction. Consequently we again need to seek magnetic solutions with a z -dependency to evade Cowling’s theorem. The magnetic field is again separable in z (eq. (177)), which leads to the 2D form of the induction equation already encountered with the Roberts cell (eq. (178)), subjected to periodic boundary conditions on $\mathbf{b}(x, y, t)$. As before the idea is to pick a value for the vertical wavenumber k , and monitor dynamo action by tracking the growth (or decay) of the magnetic energy via eq. (179).

Computing solutions for varying k soon reveals that dynamo action (i.e., positive growth rates $s(k, R_m)$) occurs in a finite range of vertical wavenumber k , with exponential growth setting in after a time of order of the turnover time. Figure 23 shows snapshot of the vertical magnetic field $b_z(x, y, t)$ in this phase of exponential growth, for a $R_m = 2000$ solutions with $k = 0.57$, which here yields the largest growth rate. The solution is fully time-dependent, and its behavior is best appreciated by viewing it as an animation¹⁴.

¹⁴ which you can do, of course, on the course’s Web Page, and for a few R_m values, moreover...

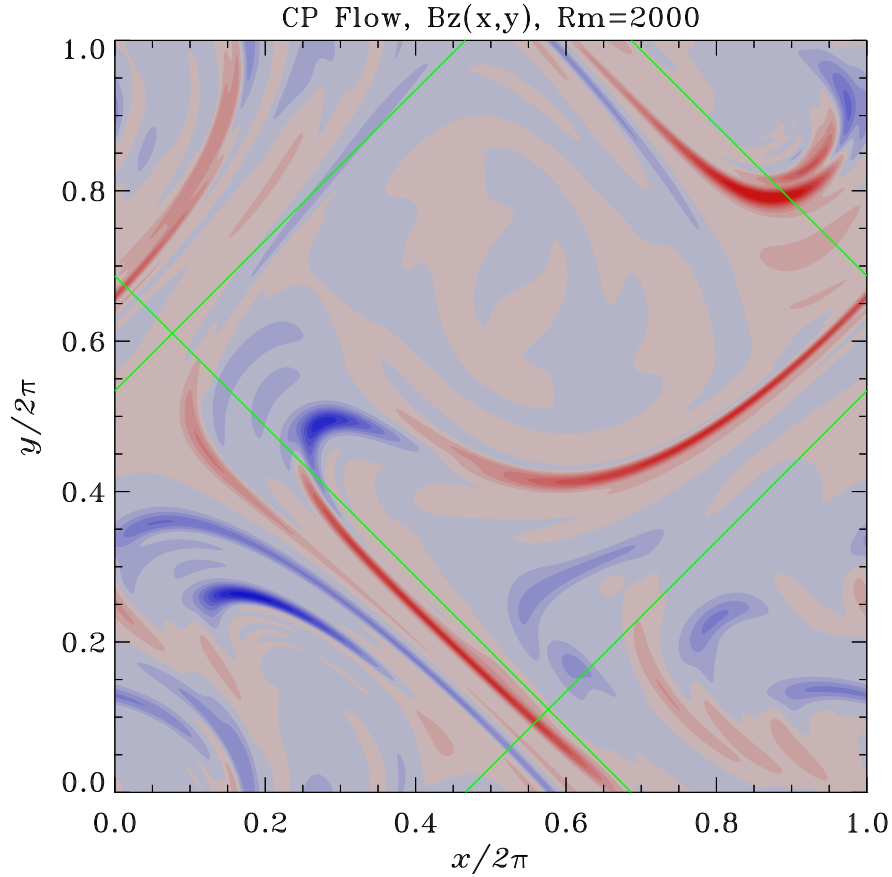


Fig. 23 ^{F3, CPF5} Snapshot of the z -component of the magnetic field in the $[x, y]$ plane, for a CP Flow solution with $R_m = 2000$ and $k = 0.57$, in the asymptotic regime $t \gg \tau_c$. The color scale codes the real part of the z -component of $\mathbf{b}(x, y, t)$ (gray-to-blue is negative, gray-to-red positive). The green straight lines indicate the separatrix surfaces of the underlying pattern of flow cells, and are no longer fixed in space due to the precession of the flow cells. (see eqs. (190)–(192)). This is a strongly time-dependent solution, exhibiting overall exponential growth of the magnetic field.

The solution is characterized by multiple sheets of intense magnetic field, of thickness once again $\propto R_m^{-1/2}$. The magnetic field exhibits *spatial intermittency*, in the sense that if one were to randomly choose a location somewhere in the $[x, y]$ plane, chances are good that only a weakish magnetic field would be found. In high- R_m solutions, strong fields are concentrated in small regions of the domain; in other words, their *filling factor* is small. This can be quantified by computing the *probability density function* (hereafter PDF) of the magnetic field strength, $f(|B_z|)$. This involves measuring B_z at every (x, y) mesh point in the solution domain, and simply counting how many

mesh points have $|B_z|$ between values B and $B + dB$. The result of such a procedure is shown in histogram form on Figure 24. The PDF shows a power-law tail at high field strengths,

$$\{E3.pltail\} \quad f(|B_z|) \propto |B_z|^{-\gamma}, \quad |B_z| \gtrsim 10^{-5}, \quad (193)$$

spanning over four orders of magnitude in field strength, and with $\gamma \simeq 0.75$ here. This indicates that strong field are still far more likely to be detected than if the magnetic field was simply a normally-distributed random variable (for example). The fact that the power law index γ is smaller than unity means that the *largest* local field strength found in the domain will always dominate the computation of the spatially-averaged field strength.

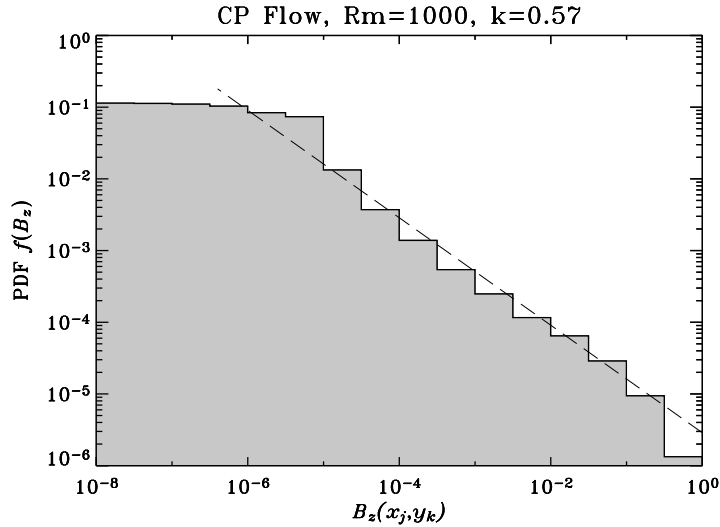


Fig. 24 ^{F3.histo} Probability density function for the (unsigned) strength of the z -component of the magnetic field, for a $R_m = 10^3$, $k = 0.57$ CP flow dynamo. The peak field strength has been normalized to a value of unity. Note the power-law tail at large field strength (straight line in this log-log plot, with slope ~ -0.75).

The CP flow dynamo solutions also exhibit *temporal intermittency*; if one sits at one specific point (x, y) point in the domain and measures B_z at subsequent time steps, a weak B_z is measured most of the time, and only occasionally are large values detected. Once again the PDF shows a power-law tail with slope flatter than -1 indicating that a temporal average of B_z at one location will always be dominated by the largest B_z measured to date.

Unlike in the Roberts cell, the range of k yielding dynamo action does *not* shift significantly to higher k as R_m is increased, and in the high R_m regime the corresponding maximum growth rate k_{\max} does *not* decrease with increasing R_m (see Fig. 21). In the CP flow considered here ($A = C = \sqrt{3}/2$,

$\omega = 1$, $\epsilon = 1$), $k_{\max} \simeq 0.57$, with $s(k_{\max}) \simeq 0.3$ for $R_m \gtrsim 10^2$, as shown on Figure 25 (solid line). Figure 25 indicates that the CP flow operates as a fast dynamo.

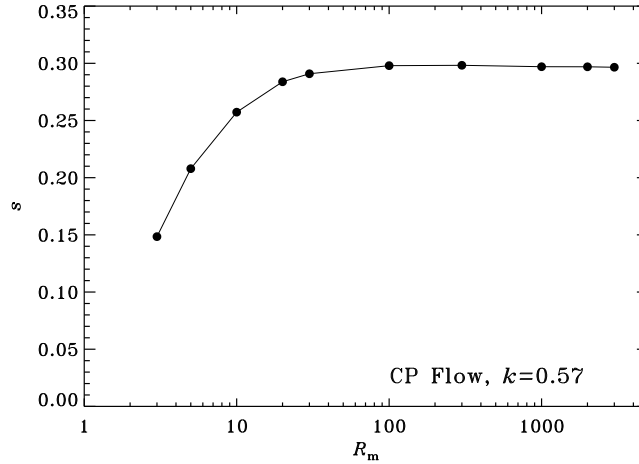


Fig. 25 ^{F3.CPF6} Growth rate of $k = 0.57$ CP flow dynamo solutions, plotted as a function of the magnetic Reynolds number (solid line). The constancy of the growth rate in the high- R_m regime suggests (but does not strictly prove) that this dynamo is fast.

0.23.2 Fast dynamo action and chaotic trajectories

Fast dynamo action in the CP flow turns out to be intimately tied to the presence of *chaotic trajectories* in the flow. In fact, theorems have been proven, demonstrating that

1. A smooth flow cannot be a fast dynamo if $\lambda_L = 0$, so that $\lambda_L > 0$, or, equivalently, the existence of chaotic regions in the flow, is a necessary (although not sufficient) condition for fast dynamo action;
2. In the limit $R_m \rightarrow \infty$, the largest Lyapunov exponent of the flow is an upper bound on the dynamo growth rate.

Proofs of these theorems need not concern us here (but see bibliography), but they once again allow us to *rule out* fast dynamo action in many classes of flows.

The presence or absence of chaotic trajectories in a flow can be quantified in a number of ways, the most straightforward (in principle) being the calculations of the flow's *Lyapunov exponents*. This is another fancy name for a rather simple concept: the rate of exponential divergence of two neighbouring

fluid element located at \mathbf{x}_1 , \mathbf{x}_2 at $t = 0$ somewhere in the flow. The Lyapunov exponent λ_L can be (somewhat loosely) defined via

$$\ell(t) = \ell(0) \exp(\lambda_L t) , \quad (194)$$

where $\ell \equiv \|\mathbf{x}_2 - \mathbf{x}_1\|$ is the length of the *tangent vector* between the two fluid elements. Because there are three independent possible directions in 3D space, one can compute *three* distinct Lyapunov exponents at any given point in the flow, and it can be shown that for an incompressible flow their sum is zero. Now, recalling the simple flux tube stretching example of §0.19, exponential divergence of two points located in the same fieldline within the tube clearly implies exponential increase in the tube's length, and therefore, via eq. (131), exponential increase of the magnetic field strength.

Calculating a *Poincaré section*, as plotted on Figure 26 for our CP flow, is another very useful way to check for chaotic trajectories in a flow. It is constructed by launching tracer particles at $z = 0$ (and $t = 0$), and following their trajectories as they are carried by the flow. At every 2π time interval, the position of each particle is plotted in the $[x, y]$ plane (modulo 2π in x and y , since most particles leave the original 2π -domain within which they were released as a consequence of cell precession). Some particles never venture too far away from their starting position in the $[x, y]$ plane. They end up tracing closed curves which, however distorted they may end up looking, identify regions of space where trajectories are integrable. Other particles, on the other hand, never return to their starting position. If one waited long enough, one such particle would eventually come arbitrarily close to all points in the $[x, y]$ plane outside of the integrable regions. The corresponding particle trajectory is said to be *space filling*, and the associated particle motion chaotic. The region of the $[x, y]$ plane defined by the starting positions of all particles with space filling trajectories is called the *chaotic region* of the flow.

{S3.flux} 0.23.3 Magnetic flux versus magnetic energy

With the CP flow, we definitely have a pretty good dynamo on our hands. But how are those dynamo solutions to be related to the Sun (or other astrophysical bodies)? So far we have concentrated on the magnetic energy as a measure of dynamo action, but in the astrophysical context *magnetic flux* is also important. Consider the following two (related) measures of magnetic flux:

$$\{\text{E3.30}\} \quad \Phi = \langle |\mathbf{B}| \rangle , \quad F = \langle |\mathbf{B}| \rangle , \quad (195)$$

where the angular brackets indicate some sort of suitable spatial average over the whole computational domain. The quantity Φ is nothing but the average magnetic flux, while F is the average *unsigned flux*. Under this notation the magnetic energy can be written as $\mathcal{E}_B = \langle \mathbf{B}^2 \rangle$. Consider now the scaling of

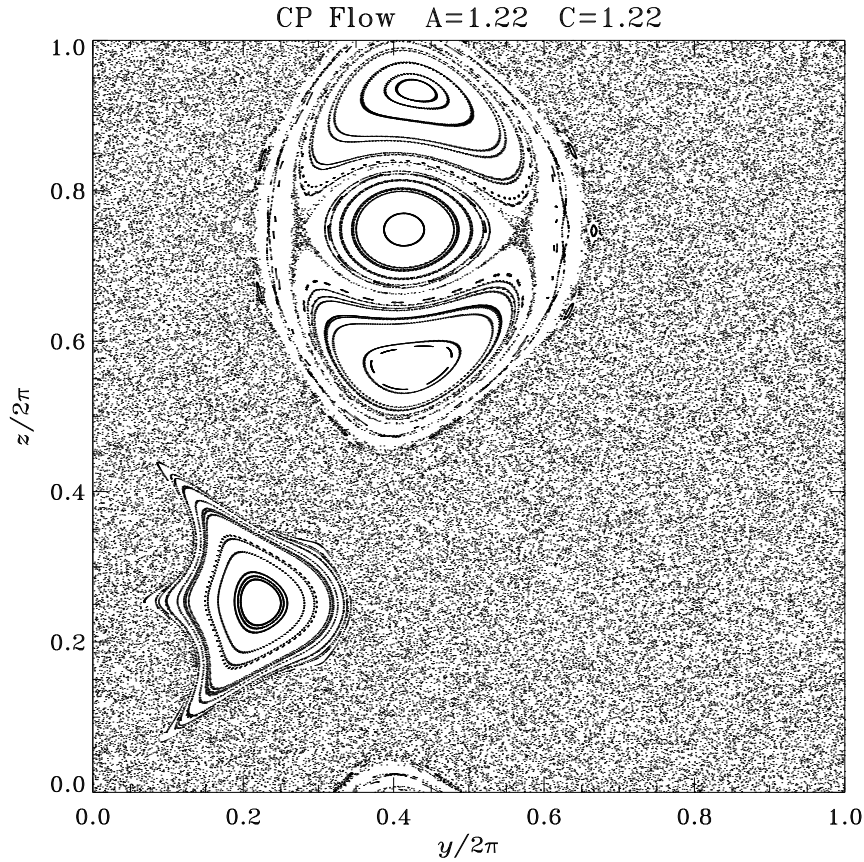


Fig. 26 $\mathbb{P}^{3,CP3}$ Poincaré section for the CP flow, for $\epsilon = 1$, $\omega = 1$, and $A = C = \sqrt{3/2}$. The plot is constructed by repeatedly “launching” particles at $z = 0$, $t = 0$, following their trajectories in time, and plotting their (projected) position (modulo 2π) in the $[x, y]$ plane at interval $\Delta t = 2\pi$. The flow is chaotic within the featureless “salt-and-pepper” regions, and integrable in regions threaded by close curves.

the two following ratios as a function of the magnetic Reynolds number:

$$\{\text{E3.31a}\} \quad \mathcal{R}_1 = \frac{\mathcal{E}_B}{\Phi^2} \propto R_m^n, \quad (196)$$

$$\mathcal{R}_2 = \frac{F^2}{\Phi^2} \propto R_m^\kappa. \quad (197) \quad \{\text{E3.31b}\}$$

A little reflection will reveal that a large value of \mathcal{R}_1 indicates that the magnetic field is concentrated in a small total fractional area of the domain, i.e.,

the *filling factor* is much smaller than unity¹⁵. The ratio \mathcal{R}_2 , on the other hand, is indicative of the dynamo's ability to generate a net signed flux. The exponent κ measures the level of *folding* in the solution; large values of κ indicate that while the dynamo may be vigorously producing magnetic flux on small spatial scales, it does so in a manner such that very little *net* flux is being generated on the spatial scale of the computational domain. Figure 27 shows the variations with R_m of the two ratios defined above. Least squares fits to the curves yields $n = 0.35$ and $\kappa = 0.13$. Positive values for the ex-

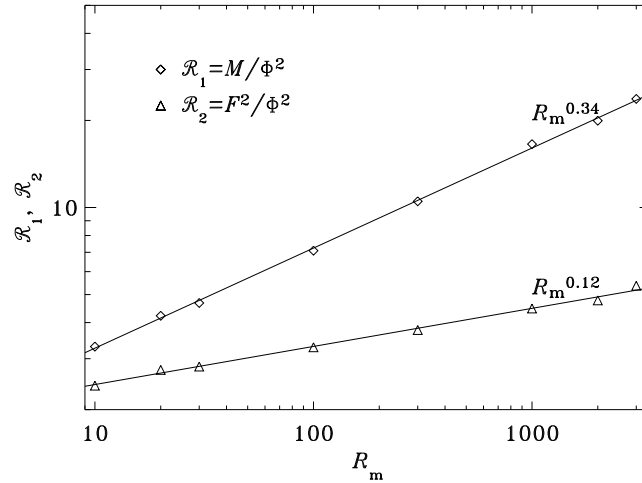


Fig. 27 ^{F3.CPF7} Variations with R_m of the two ratios defined in eqs. (196)–(197). Least squares fits (solid lines) yield power law exponents $n = 0.35$ and $\kappa = 0.13$.

ponents κ and n indicate that the CP flow dynamo is relatively inefficient at producing magnetic flux in the high R_m regime, and even less efficient at producing net signed flux. While other flows yielding fast dynamo action lead to different values for these exponents, in general they seem to always turn out positive, with $\kappa < n$, so that the (relative) inability to produce net signed flux seems to be a generic property of fast dynamos in the high- R_m regime.

0.23.4 Fast dynamo action in the nonlinear regime

We conclude this section by a brief discussion of fast dynamo action in the nonlinear regime. Evidently the exponential growth of the magnetic field will

¹⁵ If you can't figure it out try this: take a magnetic field of strength B_1 crossing a surface area A_1 ; now consider a more intense magnetic field, of strength $B_2 = 4B_1$, concentrated in one quarter of the area A_1 ; calculate \mathcal{E}_B , Φ , and \mathcal{R}_1 ... get it?

be arrested once the Lorentz force becomes large enough to alter the original CP flow. What might the nature of the backreaction on \mathbf{u} look like?

Naively, one might think that the Lorentz force will simply reduce the amplitude of the flow components, leaving the overall geometry of the flow more or less unaffected. That this cannot be the case becomes obvious upon recalling that in the high R_m regimes the eigenfunction is characterized by magnetic structures of typical thickness $\propto R_m^{-1/2}$, while the flow has a typical length scale $\sim 2\pi$ in our dimensionless units. The extreme disparity between these two length scales in the high- R_m regime suggests that the saturation of the dynamo-generated magnetic field will involve alterations of the flow field on small spatial scales, so that a flow very much different from the original CP flow is likely to develop in the nonlinear regime.

That this is indeed what happens was nicely demonstrated some years ago by F. Cattaneo and collaborators (see references in bibliography), who computed simplified nonlinear solutions of dynamo action in a suitably forced CP flow. They could show that

1. the r.m.s. flow velocity in the nonlinearly saturated regime is comparable to that in the original CP flow;
2. magnetic dissipation actually *decreases* in the nonlinear regime;
3. dynamo action is suppressed by the disappearance of chaotic trajectories in the nonlinearly modified flow.

0.24 Dynamo action in turbulent flows

{pch_sec:turbdyn}

The Roberts cell and CP flow are arguably more akin to malfunctioning washing machines than any sensible astrophysical object. Nonetheless many things we have learned throughout this chapter do carry over to more realistic circumstances, and in particular to turbulent, thermally-driven convective fluid motions.

As support for this grand sweeping claim, consider Figure 28 herein. It is a snapshot of a numerical simulation of dynamo action in a stratified, thermally-driven turbulent fluid being heated from below and spatially periodic in the horizontal directions. The fluid is contained in rectangular box of aspect ratio $x : y : z = 10 : 10 : 1$, and we are here looking at the top layer of the simulation box, with the color scale encoding the vertical magnetic field component $B_z(x, y)$. Such thermally-driven turbulent flows in a stratified background have long been known to be characterized by cells of broad upwellings of warm fluid. These cells have a horizontal size set by, among other things, the density scale height within the box; On the other hand, the downwelling of cold fluid needed to satisfy mass conservation end up being concentrated in a network of narrow lanes at the boundaries between adjacent upwelling cells. This asymmetry is due to the vertical pressure and density

gradient in the box: rising fluid expands laterally into the lower density layers above, and descending fluid is compressed laterally in the higher density layers below. Near the top of the simulation box, this leads to the concentration of magnetic structures in the downwelling lanes, as they are continuously being swept horizontally away from the centers of upwelling cells.

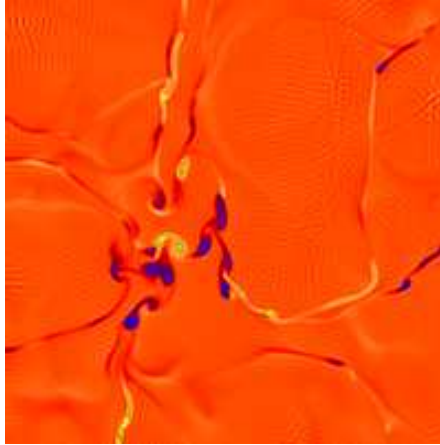


Fig. 28 ^{F3, SMAG1} Closeup on a snapshot of the top “horizontal” $[x, y]$ plane of a MHD numerical simulation of thermally-driven stratified turbulent convection in a box of aspect ratio $x : y : z = 10 : 10 : 1$, at a Viscous Reynolds number of 245 and $R_m = 1225$. The simulation uses a pseudo-spectral spatial discretization scheme, with 1024 collocation points in the x and y directions, and 97 in z . The color scale encodes the vertical (z) component of the magnetic field (orange-to-yellow is positive B_z , orange-to-blue negative). Numerical simulation results kindly provided by F. Cattaneo, University of Chicago.

This convectively-driven turbulent flow acts as a vigorous nonlinear fast dynamo, with a ratio of magnetic to kinetic energy of about 20%. This dynamo, much like that arising in the CP flow,

1. produces magnetic fields that are highly intermittent, both spatially and temporally;
2. produces flux concentrations on scales $\propto R_m^{-1/2}$;
3. produces little or no mean-field, i.e., signed magnetic flux on a spatial scale comparable to the size of the system.

The fundamental physical link between turbulent convection and the CP flow is the presence of chaotic trajectories in both flows, which leads to the expectation that dynamo action should be possible in convection zones of the Sun and stars. Time to move on, then, to the solar magnetic field and its underlying dynamo mechanism(s).

Bibliography:

The first detailed discussions of the diffusive decay of large-scale magnetic fields in astrophysical bodies are due to

Cowling, T.G., *Mon. Not. Roy. Astron. Soc.*, **105**, 166-174 (1945),
 Wrubel, M.H., *Astrophys. J.*, **16**, 291-298 (1952).

On the numerical solution of algebraic eigenvalue problems in general, and on inverse iteration in particular, see

Golub, G.H., & Van Loan, C.F., *Matrix Computations (second edition)*,
 Baltimore: The Johns Hopkins University Press; §7.6.1 and
 7.7.8 (1989).

Press, W.H., Teukolsky, S.A., Vetterling, W.T., & Flannery, B.P., *Numerical Recipes*, Second Ed., (Cambridge: Cambridge University Press), §11.7 (1992).

Chapter 19 of Press *et al.* also contains a good introduction to the art of PDE discretization by finite differences, including the simple explicit schemes used to compute the various solutions discussed in §§0.20, 0.22, and 0.23.

Among the numerous books discussing the calculation of macroscopic transport coefficients such as viscosity and electrical conductivity, starting from a microscopic point of view, I still like the following old classic:

Spitzer, L. Jr., *Physics of Fully Ionized Gases* (second ed.), New York: Wiley Interscience (1962).

On the Stretch-Twist-Fold dynamo, see

Vainshtein, S.I., and Zeldovich, Ya.B., *Soviet Physics Uspekhi*, **15**(2), 159 (1972),

Moffatt, H.K., and Proctor, M.R.E., *J. Fluid Mech.*, **154**, 493-507 (1985).

Magnetic flux expulsion from regions of closed streamlines is discussed in many textbooks dealing with magnetohydrodynamics. Analytical solutions for some specific cellular flows can be found for example in

Moffatt, H.K., *Magnetic field generation in electrically conducting fluids*,
 Cambridge: Cambridge University Press (1978),

Parker, E.N., *Cosmical Magnetic Fields*, Oxford: Clarendon Press, chap. 16 (1979).

but for the first and last word on this topic, you should consult

Weiss, N.O., *Proc. Roy. Soc. London A***293**, 310-328 (1966),

Rhines, P.B., & Young, W.R., *J. Fluid Mech.*, **133**, 133-145 (1983).

The Rhines & Young paper contains clean analytical examples of the two successive dissipative phases with characteristic timescales proportional to $R_m^{1/3}$ and R_m , as discussed in §0.20.4. Flux expulsion is of course not restricted to 2-D flow; for a nice example in 3-D see

Galloway, D.J., & Proctor, M.R.E., *Geophys. Astrophys. Fluid Dyn.*, **24**, 109-136 (1983).

On the possible symmetrization of the Saturnian magnetic field by deep-seated differential rotation, see

Stevenson, D.J., *Geophys. Astrophys. Fluid Dyn.*, **21**, 113-127 (1982),
Kirk, R.L., & Stevenson, D.J., *Astrophys. J.*, **316**, 836-846 (1987).

An insightful discussion of the symmetrization process in more general terms is that of

Rädler, K.-H. 1986, On the effect of differential rotation on axisymmetric and non-axisymmetric magnetic fields in cosmic bodies, in *Proceedings of the Joint Varenna-Abastumani International School and Workshop*, ESA Spec. Pub. SP-251, 569-574.

On anti-dynamo theorems, see

Cowling, T.G., *Mon. Not. Roy. Astron. Soc.*, **94**, 39-48 (1933),
Bullard, E.C., & Gellman, H., *Phil. Trans. R. Soc. London A*, **247**, 213-278 (1954),
Zeldovich, Ya. B., *J. Exp. and Theoretical Physics*, **31**, 154 (1956) [Russian],
Zeldovich, Ya. B., & Ruzmaikin, A.A., *J. Exp. and Theoretical Physics*, **78**, 980 (1980) [Russian],

as well as pages 113-ff and 538-ff, respectively, of the books by Moffatt and Parker listed above. English translations of the last two papers are also reprinted in *Selected Works of Yakov Borisovich Zeldovich*, vol. 1 (ed. J.P. Ostriker, Princeton, 1992).

The mathematical aspects of fast dynamo theory are discussed at length in the book

Childress, S., & Gilbert, A.D., *Stretch, Twist, Fold: The Fast Dynamo*, Berlin: Springer (1995),

although the reader preferring a shorter introduction might want to first work through the three following review articles:

Childress, S., in *Topological Aspects of the Dynamics of Fluids and Plasmas*, eds. H.K. Moffatt *et al.*, Kluwer, 111-147 (1992).
Soward, A.M., in *Lectures on Solar and Planetary Dynamos*, eds. M.R.E. Proctor & A.D. Gilbert, Cambridge University Press, 181 (1994),
Gilbert, A.D., "Dynamo theory", in *Handbook of Mathematical Fluid Dynamics*, vol. 2, eds. S. Frierlander and D. Serre, Elsevier, 355-441 (2003),

On the Roberts cell dynamo, see chapter 5 of the Childress & Gilbert book cited above, as well as

Roberts, G.O., *Phil. Trans. R. Soc. London*, **A271**, 411-454 (1972),

Soward, A.M., *J. Fluid Mech.*, **180**, 267-295 (1987).

The scaling relations given by eqs. (180)–(181) are derived in the second of these papers. On fast dynamo action in the CP flow, begin with:

Galloway, D.J., & Proctor, M.R.E., *Nature*, **356**, 691-693 (1992),

Ponty, Y., Pouquet, A., & Sulem, P.L., *Geophys. Astrophys. Fluid Dyn.*, **79**, 239-257 (1995),

Cattaneo, F., Kim, E.-J., Proctor, M.R.E., & Tao, L., *Phys. Rev. Lett.*, **75**, 1522-1525 (1995).

The discussion in §0.23.3 follows closely that in the Cattaneo *et al.* (1995) paper, and the discussion of nonlinear effects in the CP flow is largely inspired from

Cattaneo, F., Hughes, D.W., & Kim, E.-J., *Phys. Rev. Lett.*, **76**, 2057-2060 (1996),

but on this topic see also

Mininni, P.D., Ponty, Y., Montgomery, D.C., Pinton, J.-F., Politano, H., and Pouquet, A., *Astrophys. J.*, **626**, 853-863 (2005).

The mathematically inclined reader wishing to delve deeper into the theorems for fast dynamo action mentioned in §0.23 will get a solid and character building workout out of

Vishik, M.M., *Geophys. Astrophys. Fluid Dyn.*, **48**, 151-167 (1989),

Klapper, I., & Young, L.S., *Comm. Math. Phys.*, **173**, 623-646 (1995).

On small-scale dynamo action in three-dimensional thermally-driven convective turbulence, see, e.g.,

Cattaneo, F., *Astrophys. J.*, **515**, L39-L42 (1999),

Cattaneo, F., Emonet, T., & Weiss, N.O., *Astrophys. J.*, **588**, 1183-1198 (2003).

Vogler, A., & Schüssler, M., *Astron. Ap.*, **465**, L43-46 (2007).

Chapter 3

Dynamo models of the solar cycle

{pch_chap:dynmod}

*Was einmal gedacht wurde,
kann nicht mehr zurückgenommen werden.*

*Friedrich Dürrenmatt
Die Physiker (1962)*

The time has now come to put together everything (well... almost) we have learned so far to construct dynamo models for solar and stellar magnetic fields. In this and the following chapter we concentrate on the Sun, for which the amount of observational data available constrains dynamo models to a degree much greater than for other stars. Dynamo action in stars other than the Sun will be considered in chapter 5, using solar dynamo models as sky-hooks.

We begin (§0.25) by briefly reviewing the basic properties of the solar magnetic cycle, which are to be (hopefully) reproduced by the (relatively) simple dynamo models to be constructed in the remainder of this chapter. The different classes of solar dynamo models to be discussed differ primarily in the choice they make regarding the physical mechanism responsible for the regeneration of the poloidal magnetic component. They all share the shearing of a poloidal field by differential rotation (§0.19.4) as a source of toroidal field, and all invoke some sort of enhanced, “turbulent” magnetic diffusivity in the solar convective envelope (more on that very shortly!).

We first introduce a statistical-physical theory known as *mean-field electrodynamics*, which will allow us to construct (relatively) simple dynamo models in which the poloidal field is produced through the inductive action of convective turbulence (§0.26). We then look into what currently stands as their main “competitors”, namely solar cycle models based on poloidal field regeneration by the surface decay of active regions, more succinctly known as Babcock-Leighton models (§0.27). We then turn to cycle models relying on various hydrodynamical or MHD instabilities, which can under certain circumstances act as sources of poloidal magnetic fields (§0.28). We close with

an overview of the current state of affairs with regards to investigations of the solar dynamo problem through large-scale MHD simulations of turbulent convection in a thick, stratified rotating shell (§0.29).

{pch_sec:obsycle}

0.25 The solar magnetic field

0.25.1 Sunspots and the photospheric magnetic field

The Sun is the first astronomical object (Earth excluded) in which a magnetic field was detected, through the epoch-making work of George Ellery Hale (1868-1938) and collaborators, in the opening decades of the twentieth century. In 1907-1908, by measuring the Zeeman splitting in magnetically sensitive lines in the spectra of *sunspots* and detecting the polarization of the split spectral components (see Fig. 29), Hale provided the first unambiguous and quantitative demonstration that sunspots are the seat of strong magnetic fields. Not only was this the first detection of a magnetic field outside the Earth, but the inferred magnetic field strength, 0.3T, turned out a few thousand times greater than the Earth’s own magnetic field. It was subsequently realized that the Lorentz force associated with such strong magnetic fields would also impede convective energy transport from below, and therefore lead naturally to the lower temperatures observed within the sunspots, as compared to the surrounding photosphere.

The solar surface magnetic field outside of sunspots, although of much weaker strength, is accessible to direct observation, usually by measuring Zeeman broadening of spectral lines, or the degree of linear or circular polarisation of light emitted from the solar photosphere. The first magnetic maps, or *magnetograms*, of the solar disk were obtained in the late 1950’s by the father-and-son team of Harold D. Babcock (1882-1986) and Horace W. Babcock (1912-2003), and were little more than photographs of stacks of a few dozen horizontal scans of the solar disk displayed on an oscilloscope. Figure 30 (top) is a modern equivalent in pixel form, with the gray scale coding the strength of the normal component of the magnetic field (mid-level gray, $|\mathbf{B}| \lesssim 1 \text{ mT}$; going to white for positive normal field, and to black for negative, peaking around 0.4T in both cases). Comparison with a continuum image (bottom) reveals that the stronger magnetic fields coincide with sunspots, but hefty fields of a few 10^{-2} Tesla can be found within and around groups of sunspots. Away from these “magnetically active regions”, the magnetic field is weaker, concentrated into clumps that collectively make up a spatially fragmented *magnetic network* distributed evenly over the whole surface. Sunspots and active regions, in contrast, are restricted to heliographic latitudes $\lesssim 40^\circ$, and their number waxes and wanes on an eleven year cycle, about which we’ll have a lot more to say in the next chapter.

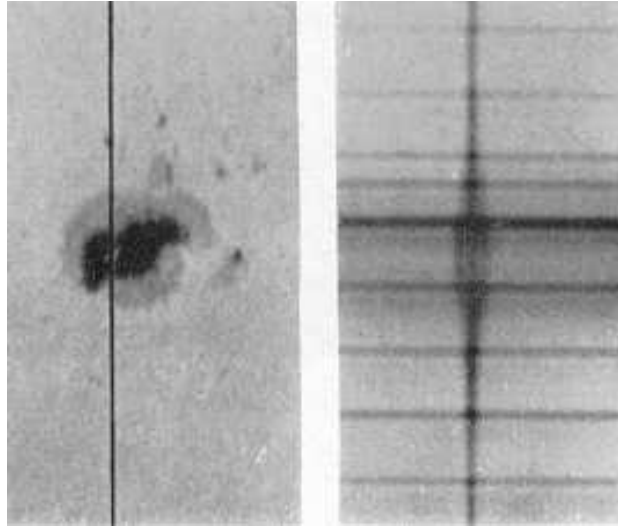


Fig. 29 ^{F1.10} The magnetically-induced Zeeman splitting in the spectrum of a sunspot. The vertical dark line on the left image is the slit having produced the vertical stack of spectra on the right image (with wavelength running horizontally). Reproduced from the 1919 paper by G.E. Hale, F. Ellerman, S.B. Nicholson, and A.H. Joy (in *The Astrophysical Journal*, vol. **49**, pps. 153-178).

Now then, to sum up: far from taking the form of a large-scale, smooth diffuse field as on the Earth, the solar photospheric magnetic field is very fragmented and topologically complex, and shows up concentrated in small magnetized regions separated by field-free plasma. This dichotomy persists down to the smallest spatial scales than can be resolved with current observational techniques. It owes much to the fact that the outer 30% in radius of the Sun is a fluid in a strongly turbulent state. High time-cadence and spatial resolutions of the solar small-scale magnetic field have shown that the associated photospheric magnetic flux is replenished on a very short timescale, commensurate with the convective turnover time immediately below the photosphere. This offers support to a turbulent dynamo-based explanation for the solar small-scale magnetic field, of the type considered in the preceding chapter, away from active regions at least.

From here onwards, we focus on the *large-scale* solar magnetic field, by which we mean the part of the Sun's magnetic field spatially organized on scales commensurate with the solar radius. While it may not be immediately obvious on Fig. 30, sunspots provide one of the better tracers of this large-scale magnetic component.

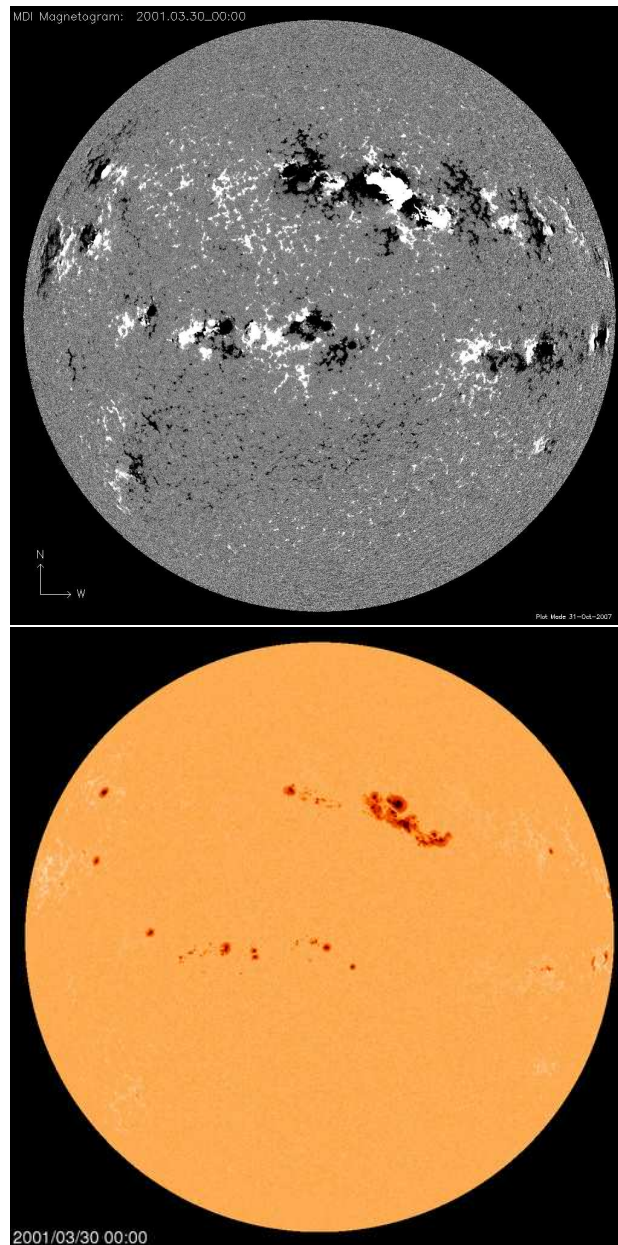


Fig. 30 ^{fig:mdi} Full disk line-of-sight magnetogram (top) and continuum intensity image (bottom) of the solar photosphere, both taken on 30 March 2001 by the MDI instrument onboard the SOHO satellite. The sun's rotation axis is vertical on both images. Public domain images downloaded from the SOHO mission Website: <http://wwwsoho.nascom.nasa.gov>

0.25.2 Hale’s polarity laws

{ssec:hale}

Hale and his collaborators did much more than just measure magnetic fields in sunspots. Through painstaking observations and analyses spanning nearly two decades, they went on to demonstrate the existence of a number of regularities in the magnetic fields of sunspots, now known as *Hale’s polarity laws*. Having noted early on that large sunspots often appear grouped in pairs of opposite magnetic polarities, they could show that:

1. At any given time, the polarities of the leading spots of sunspot pairs are the same in a given solar hemisphere;
2. At any given time, the polarities of the leading spots of sunspot pairs are opposite in the N and S hemispheres;
3. Sunspot polarities reverse in each hemisphere from one 11-yr sunspot cycle to the next;

This polarity ordering is fairly easy to discern on the magnetogram of Fig. 30. The most straightforward interpretation of this common opposite polarity grouping is that we are seeing the surface manifestation of a large-scale toroidal field residing somewhere below the photosphere, having risen upwards and pierced the photosphere in the form of a so-called “ Ω -loop” (see Figure 31), its intersection with the photosphere producing sunspot pairs of opposite polarities. If the zonal unit vector \hat{e}_ϕ is aligned with the direction of solar rotation, then the sign of the deep-seated toroidal component B_ϕ is then given by the the magnetic polarity of the trailing sunspots. This picture of sunspot pairs, taken in conjunction with Hale polarity Laws, therefore indicate that the sun’s internal toroidal field is antisymmetric about the equator.

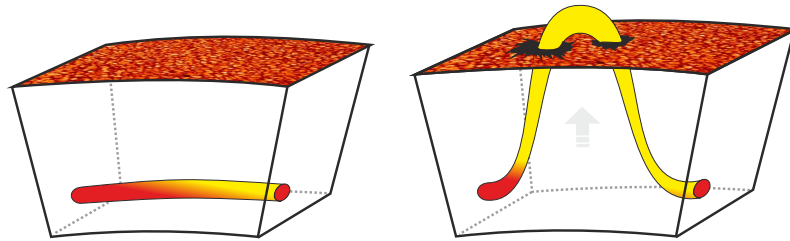


Fig. 31 ^{F1, 11b} Schematic representation of a sunspot pair as the manifestation of an underlying toroidal flux rope having risen through the photosphere as an “ Ω -loop”. At left, the flux tube lies in the azimuthal direction, before destabilisation and buoyant rise through the photosphere (at right). The magnetic fields impedes convective energy transport, so that cooling leads to a collapse of the magnetic field into two sunspots of opposite polarities. Diagram kindly provided by D. Passos.

Another pattern uncovered Hale and collaborators is that the line segment joining two members of a sunspot pair tends to show a systematic tilt angle (γ) with respect to the East-West direction, the sunspot farther ahead (in the direction of solar rotation) being closer to the equator. Although there exists considerable variations in observed tilt angles, statistically the magnitude of the tilt increases with increasing heliocentric latitude. This is known as *Joy's Law*. Least-squares fits to observations yield a parametric representation of the form:

$$\sin \gamma = 0.5 \cos(\theta) \quad (198)$$

where θ is the usual polar angle. This pattern plays an important role in some of the solar cycle models to be considered later in this chapter. This is because the existence of a finite, systematic tilt implies a net dipole moment, which can contribute to the solar poloidal field.

The hemispheric antisymmetry evidenced by Hale's polarity Laws can be readily produced by the shearing of a large-scale poloidal field by a differential rotation symmetric about the equatorial plane, exactly as we modeled already in §0.19.4. The very existence of Hale's polarity Laws thus suggests the presence of a large-scale poloidal component to the solar magnetic field; its detection was beyond the capability of Hale's instruments, but later observations clearly established its existence, and its close connection to the internal magnetic field through the solar magnetic cycle.

0.25.3 The magnetic cycle

Figure 32 is a synoptic (time-latitude) diagram of the longitudinally-averaged photospheric radial magnetic field component, covering three sunspot cycles. Such a diagram is constructed by averaging magnetograms (like the one on Fig. 30) in longitude over each successive solar rotation, and stacking side-by-side the resulting latitudinal distribution of ϕ -averaged magnetic field to form a temporal sequence. The most immediately striking global patterns apparent on Figure 32 are certainly the cyclic variations on a ~ 22 yr period, accompanying polarity reversals, and the (anti)symmetry about the solar equator.

The magnetic signal present within the latitudinal band extending 30 degrees or so on either side of the equator is the magnetographic imprint of sunspots. Their strong magnetic fields (~ 0.1 T) almost average out on such synoptic diagram, because, as already noted, they tend to appear in close pairs of opposite magnetic polarities with comparable (unsigned) magnetic flux. At the beginning of a sunspot cycle (e.g., 1976, 1986, 1996 on Fig. 32), sunspots are observed at relatively high ($\sim 40^\circ$) heliocentric latitudes, but emerge at lower and lower latitudes as the cycle proceeds, until at the end of the cycle they are seen mostly near the equator, at which time spots announc-

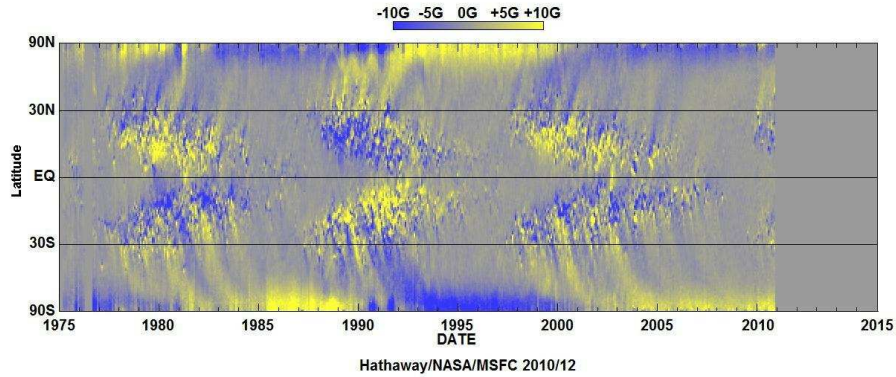


Fig. 32 ^{fig:synoptic} A synoptic magnetogram covering the last three sunspot cycles. The radial component of the sun's magnetic field is azimuthally averaged over a solar rotation, and the resulting latitudinal strips stacked one against the other in the form of a time-latitude diagram. Recall that $1\text{T} \equiv 10^4\text{Gauss}$. Data and graphics courtesy of David Hathaway, NASA/MSFC. [<http://solarscience.msfc.nasa.gov/images/magbfly.jpg>]

ing the onset of the next cycle begin to appear again at $\sim 40^\circ$ latitude. This results in the so-called “butterfly diagram” of sunspot distribution, about which we’ll have more to say in the following chapter. Cycle maximum (as measured by sunspot number) occurs about midway along each butterfly, when sunspot coverage is maximal at about 15 degrees latitude, here 1980, 1991 and 2002.

At high heliocentric latitude ($\gtrsim 50^\circ$) the synoptic magnetograms are dominated by a well-defined dipole component, with strength $\sim 10^{-3}\text{T}$, showing a clean pattern of polarity changes occurring at or near sunspot maximum. For example, during the 1976—1986 cycle the toroidal field was negative in the N-hemisphere, and the Northern polar field reversed from positive to negative magnetic polarity; taken at face value, Figure 32 then indicates that the high latitude poloidal field lags the toroidal field by a phase interval $\Delta\varphi \simeq \pi/2$.

At mid-latitudes the most prominent feature is a fairly regular poleward drift of magnetic fields originating in sunspot latitudes, presumably released there by the decay of sunspot and active regions. It is quite possible that this poleward transport of magnetic flux from active region belts contributes to the polarity reversal of the polar fields.

A $\sim 10^{-3}\text{T}$ polar field pervading a polar cap of $\sim 30^\circ$ angular width, as on Fig. 32, adds up to a poloidal magnetic flux of $\sim 10^{14}\text{Wb}$. The total unsigned flux emerging in active regions, taken to be representative of the solar internal toroidal magnetic component, adds up to a few 10^{17}Wb over a full sunspot cycle. This is usually taken to indicate that the solar internal magnetic field is dominated by its toroidal magnetic component.

We will return to sunspots and their spatiotemporal variations in numbers in §0.30, when we consider the origin of fluctuations in the solar cycle. For the time being we will just concentrate on what they tell us about the strength of the sun's *internal* magnetic field.

{SSfluxropes} 0.25.4 Sunspots as tracers of the sun's internal magnetic field

In translating the cartoon of Figure 31 into a quantitative physical model, we have a number of issues that need to be clarified. The first is to identify the region(s) of the solar interior from which the flux ropes originate. The magnetic pressure ($\propto \mathbf{B}^2$) within a strongly magnetized flux tube leads to a density deficit in order to reach pressure equilibrium with the surrounding plasma. The resulting upward buoyancy force can bring the tube to the surface, which is good and needed, but it turns out that for tubes located within the bulk of the convection zone the rise time is far too short to allow field amplification to a level commensurate with observed sunspot field strengths. This has led to the conclusion that the solar magnetic field is stored—maybe even produced—not in the convective envelope proper, but rather immediately below it, within the tachocline.

Considerable efforts have gone into making models of the storage, destabilization and buoyant rise of thin magnetic flux tubes through the solar convective envelope (see bibliography at the end of this chapter). In most cases, flux tubes are treated as structureless, flux-carrying material lines—the so-called thin flux tube approximation—and so these kinds of calculations cannot properly take into account the interaction of the tube with the surrounding turbulent fluid motions. With this caveat in mind, thin flux tube modeling has produced the following two important results:

1. The flux ropes rise essentially radially if they have a field strength in excess of $B \gtrsim 6\text{--}10\text{ T}$; otherwise the Coriolis force deflects the rising flux tubes to high latitudes.
2. The flux ropes emerge without any tilt for $B \gtrsim 100\text{ T}$, and with tilts compatible with Joy's Law for fields strengths in the range $6\text{--}16\text{ T}$.

The basic physical mechanism underlying these two remarkable results is the same: if the rise time of the flux ropes is of the order of the solar rotation period, the Coriolis force has an important influence. It is the Coriolis force that, upon acting on the internal flow developing along the length of the flux rope during its rise, gives rise to the twist that, upon emergence, manifests itself as Joy's Law. If the field is strong enough for the rise time to be much shorter than the rotation period, then the rising flux rope does not “feel” the rotation, rises radially, and emerges without a tilt. If on the other hand the magnetic field is too weak, the Coriolis force deflects the rising flux rope on

a trajectory running parallel to the rotation axis, resulting in emergence at high heliographic latitudes.

Now, this is great stuff: the observed emergence of sunspots at low heliocentric latitudes puts a *lower* limit on the strength of the participating flux ropes; Joy's Law, on the other hand, translate into an *upper* limit on the field strength. One concludes that the sunspot-forming toroidal flux ropes must have magnetic field strengths in the rather narrow range

$$6 \lesssim B \lesssim 16 \text{ T} . \quad (199)$$

While some level of field amplification is likely during the (ill-understood) process of flux tube formation from the spatially diffuse large-scale magnetic field produced by the dynamo, these modelling results are usually taken to indicate that the large-scale toroidal magnetic field at or below the base of the convective envelope, where stability analyses indicate sunspots-forming toroidal flux rope are formed stored, must have a strength in the range of a few tenths to a few Teslas. By most estimates, the associated magnetic energy density is at least comparable, and perhaps quite a bit larger, than the kinetic energy density of the turbulent fluid motions driving dynamo action.

0.25.5 A solar dynamo shopping list

{pch_ssec:shop}

To close this very brief overview, let's now collect a short list of fundamental observational features that a physical model of the solar large-scale magnetic field should reproduce (anything related to amplitude fluctuation being deferred to chapter 4):

1. A large-scale magnetic field, axisymmetric to a good approximation and antisymmetric about the solar equator;
2. A cyclic variation of this large-scale magnetic field, characterized by polarity reversals with a ~ 20 yr oscillation period;
3. An internal toroidal field of strength $\sim 0.1\text{--}1$ T, concentrated at low solar latitudes ($\lesssim 45^\circ$, say), and migrating equatorward in the course of the cycle with minimal spatiotemporal overlap between successive cycles;
4. A large-scale surface poloidal field of a few 10^{-3} T, migrating poleward in the course of the cycle, and reversing polarity at sunspot maximum.

These properties do not square well with fast dynamo action in turbulent flow; in particular, the Sun's large-scale magnetic field component is characterized by a substantial signed (hemispheric) magnetic flux, for which something else than fast dynamo action is needed. It turns out that the turbulent nature of the flow in the solar convective envelope can still do the trick, but to examine this we will need to adopt a statistical approach to turbulence and to the associated flow-field interactions.

0.26 Mean-field dynamo models

{pch_sec:mfsoldyn}

The ‘toy’ dynamo flows considered in §§0.22 and §0.23 exemplified the fact that high- R_m turbulent flows can be quite effective at producing a lot of small-scale magnetic fields, where “small-scales” is roughly $R_m^{-1/2}$ times the length scale of the flow. At the solar surface, the latter is around $\sim 10^6$ m and the former $\sim 10^8$ (for granulation), which yields very small scales indeed, ~ 100 m! So, at some level, the small-scale magnetic fields on the sun and stars are already taken care of. It turns out that under certain conditions, solar/stellar convective turbulence can also produce magnetic fields with a mean component building up on large spatial scales. These *mean-field dynamo models* remain arguably the most “popular” descriptive models for dynamo action in the Sun and stars, but also in planetary metallic cores, stellar accretion disks, and even galactic disks. Accordingly, we will look into the formulation of these models at some depth¹⁶.

{pch_ssec:MFE}

0.26.1 Mean-field electrodynamics

The fundamental idea on which mean field theory rests is the *two scales approach*, which consists of a decomposition of the field variables into mean and fluctuating parts. This process naturally implies that an averaging procedure can meaningfully be defined. The derivation of mean field theory can proceed equally from the choice of space averages, time averages or ensemble averages. In the context of axisymmetric dynamo models, longitudinal averages impose themselves rather naturally. For the time being let’s just define our averaging operator as:

$$\{E4.1\} \quad \langle A \rangle = \frac{1}{\lambda^3} \int_V A \, d\mathbf{x} . \quad (200)$$

We also assume that the velocity and magnetic field can be decomposed into a mean and fluctuating part so that

$$\{E4.2\} \quad \mathbf{U} = \langle \mathbf{U} \rangle + \mathbf{u} , \quad \text{and} \quad \mathbf{B} = \langle \mathbf{B} \rangle + \mathbf{b} . \quad (201)$$

The decomposition (201) makes sense provided $\langle \mathbf{u} \rangle = \langle \mathbf{b} \rangle = 0$. The physical interpretation of (201) is as follows. The velocity and magnetic fields are characterized by a slowly varying component, $\langle \mathbf{U} \rangle$ and $\langle \mathbf{B} \rangle$, which vary on the characteristic large scale L , plus rapidly fluctuating parts, \mathbf{u} and \mathbf{b} , which vary on the much smaller scale ℓ . The volume averages are computed over some intermediate scale λ such that

¹⁶ §§0.26.1 through 0.26.6 are to a large extent adapted from class notes written by Thomas J. Bogdan for the graduate class APAS7500 we co-taught in 1997 at the University of Colorado at Boulder.

$$\{E4.3\} \quad \ell \ll \lambda \ll L . \quad (202)$$

Whenever (202) is satisfied we say that we have a “good” scale separation.

The objective of mean field theory is to produce a closed set of equations for the mean quantities. Substituting (201) into the induction equation (58), and averaging, we obtain equations for the mean and fluctuating quantities, namely

$$\frac{\partial \langle \mathbf{B} \rangle}{\partial t} = \nabla \times (\langle \mathbf{U} \rangle \times \langle \mathbf{B} \rangle + \mathcal{E} + \eta \nabla \times \langle \mathbf{B} \rangle), \quad (203) \quad \{E4.4\}$$

and

$$\frac{\partial \mathbf{b}}{\partial t} = \nabla \times (\langle \mathbf{U} \rangle \times \mathbf{b} + \mathbf{u} \times \langle \mathbf{B} \rangle + \mathbf{G} + \eta \nabla \times \mathbf{b}), \quad (204) \quad \{E4.5\}$$

where

$$\mathcal{E} = \langle \mathbf{u} \times \mathbf{b} \rangle, \quad \text{and} \quad \mathbf{G} = \mathbf{u} \times \mathbf{b} - \langle \mathbf{u} \times \mathbf{b} \rangle . \quad (205) \quad \{E4.6\}$$

The important thing is that (203) now contains a source term, \mathcal{E} , associated with the average of products of fluctuations, which in general does *not* vanish upon averaging even though \mathbf{u} and \mathbf{b} individually do. The term \mathcal{E} , which is called the mean electromotive force, or emf for short, plays a central role in this theory.

Now, the whole point of the mean-field procedure is to avoid having to deal explicitly with the small scales, so we do not want to be integrating eq. (204) explicitly. But then we have a closure problem: eq. (203) is a 3-component vector equation, for the six components of $\langle \mathbf{B} \rangle$ and \mathbf{b} (leaving the flow out of the picture for the moment). Therefore it is clear that to solve (203), \mathcal{E} must be expressed as some function of $\langle \mathbf{U} \rangle$ and $\langle \mathbf{B} \rangle$.

In order to obtain the the desired expression, we note that (204) is a *linear* equation for \mathbf{b} with the term $\nabla \times (\mathbf{u} \times \langle \mathbf{B} \rangle)$ acting as a source. There must therefore exist a *linear* relationship between \mathbf{B} and \mathbf{b} , and hence, one between \mathbf{B} and $\langle \mathbf{u} \times \mathbf{b} \rangle$. The latter relationship can be expressed formally by the following series

$$\mathcal{E}_i = \alpha_{ij} \langle B \rangle_j + \beta_{ijk} \partial_k \langle B \rangle_j + \gamma_{ijkl} \partial_j \partial_k \langle B \rangle_l + \dots, \quad (206) \quad \{E4.7\}$$

where the tensorial coefficients, α , β , γ , and so forth must depend on $\langle \mathbf{U} \rangle$, what we might loosely term the *statistics* of the turbulent velocity fluctuations, \mathbf{u} , and perhaps on the diffusivity η —but *not* on $\langle \mathbf{B} \rangle$. In this sense, equations (203) and (206), constitute a closed set of equations for the evolution of $\langle \mathbf{B} \rangle$. The convergence of the series representation provided by equation (206) can be anticipated in those cases where the good separation of scales applies. For in these cases each successive derivative in equation (206) is smaller than the previous one by approximately a factor of $\ell/L \ll 1$. With

any luck, we may expect equation (206) to be dominated by the first few terms.

{pch_ssec:alpha} 0.26.2 The α -effect

We have already remarked that \mathcal{E} in (203) acts as a source term for the mean field. It is instructive to examine the contributions to \mathcal{E} deriving from the individual terms in the expansion (206). The first contribution is associated with the second-rank tensor, α_{ij} , thus

$$\{\text{E4.8}\} \quad \mathcal{E}_i^{(1)} = \alpha_{ij} \langle B \rangle_j . \quad (207)$$

The first thing to note is that α_{ij} must be a pseudo-tensor since it establishes a linear relationship between a polar vector—the mean emf, and an axial vector—the mean magnetic field. We can divide α_{ij} into its symmetric and antisymmetric parts, thus¹⁷

$$\{\text{E4.9}\} \quad \alpha_{ij} = \alpha_{ij}^s - \epsilon_{ijk} a_k , \quad (208)$$

where $2a_k = -\epsilon_{ijk} \alpha_{ij}$. From (207) we have

$$\{\text{E4.10}\} \quad \mathcal{E}_i^{(1)} = \alpha_{ij}^s \langle B \rangle_j + (\mathbf{a} \times \langle \mathbf{B} \rangle)_i . \quad (209)$$

The effect of the antisymmetric part is to provide an additional advective velocity (not in general solenoidal), so that the effective mean velocity becomes $\langle \mathbf{U} \rangle + \mathbf{a}$. It can result in *turbulent pumping* of magnetic fields. The nature of the symmetric part is most easily illustrated in the case when \mathbf{u} is an *isotropic* random field.¹⁸ Then \mathbf{a} is zero, α_{ij} must be an isotropic tensor of the form $\alpha_{ij} = \alpha \delta_{ij}$, and (209) reduces to

¹⁷ Here, ϵ_{ijk} is the Levi-Civita tensor density, also known as the unit alternating tensor, and has the values $\epsilon_{ijk} = 0$ when i, j, k are not all different, $\epsilon_{ijk} = +1$ or -1 when i, j, k are all different and in cyclic, or acyclic, order respectively. A particularly useful formula is (Einstein summation over repeated indices in force):

$$\epsilon_{ijk} \epsilon_{klm} = \delta_{il} \delta_{jm} - \delta_{im} \delta_{jl} ,$$

where δ_{ij} is the Kronecker-delta, and has the value $\delta_{ij} = 0$ if i, j are different, and $\delta_{ij} = 1$ when $i = j$.

¹⁸ Throughout the rest of this chapter, we will have cause to repeatedly refer to the statistical properties of the turbulent velocity field. In order to avoid confusion we state the following definitions: a (random) field is *stationary* if its probability density function (pdf) is time independent, it is *homogeneous* if its pdf is independent of position, it is *isotropic* if its pdf is independent of orientation (or equivalently, invariant under rotations), and it is *reflectionally symmetric* if its pdf is invariant under parity reversal. We should note that isotropy and reflectional symmetry are taken here to be distinct properties, although this protocol is not universally accepted.

$$\{\text{E4.11}\} \quad \boldsymbol{\mathcal{E}}^{(1)} = \alpha \langle \mathbf{B} \rangle . \quad (210)$$

Using Ohm's law, this component of the emf is found to generate a contribution to the mean current of the form

$$\mathbf{j}^{(1)} = \alpha \sigma_e \langle \mathbf{B} \rangle , \quad (211) \quad \{\text{E4.12}\}$$

where σ_e is the electrical conductivity. For nonzero α , equation (211) implies the appearance of a mean current everywhere *parallel* to the mean magnetic field—the so-called α -effect. This is in sharp contrast to the more conventional case where the induced current $\sigma_e(\mathbf{U} \times \mathbf{B})$ is *perpendicular* to the magnetic field. We are used to thinking as electrical currents being the source of magnetic fields (think of the Biot-Savart Law, of the pre-Maxwellian form of Ampère's Law); but a mechanically forced magnetic field can become a source of electrical current. That's really what induction is all about.

In the context of axisymmetric large-scale astrophysical magnetic fields, the importance of the α -effect is immediately apparent. We recall from our deliberations in §0.19.4 that a toroidal field could be generated from a poloidal one by differential rotation (velocity shear). The α -effect makes it possible to drive a mean toroidal current parallel to the mean toroidal field, which, in turn will regenerate a poloidal field thereby closing the dynamo cycle.

To appreciate the physical nature of the α -effect we pause to examine the original 1955 physical picture put forth by E.N. Parker. We define a cyclonic event to be the rising of a fluid element associated with a definite twist, say anticlockwise when seen from below (see Figure 33). In spherical geometry, we then consider the effect of many such events, distributed randomly in longitude and time, on an initially purely toroidal field line. Each cyclonic event creates an elemental loop of field with an associated current distribution that will have a component parallel to the initial field if the angle of rotation is less than π and antiparallel if it is greater. By assuming that the individual events are short lived we can rule out rotations of more than 2π . It is clear that the combined effect of many such events is to give rise to a net current with a component along $\langle \mathbf{B} \rangle$.

An important property of α is its pseudoscalar nature, i.e., α changes sign under parity transformations. This implies that α can be nonzero only if the statistics of \mathbf{u} lacks reflectional symmetry. In other words the velocity field must have a definite chirality. For example, on Fig. 33 there is a definite relationship between vertical displacements and sense of twist, set by the Coriolis force. This is similar to the Roberts cell flow, where the sign of the z -component changed sign in step with the sense of rotation in contiguous flow cells. In general the lack of reflectional symmetry of the fluid velocity manifests itself through a nonzero value of the mean fluid helicity $\langle \mathbf{u} \cdot (\nabla \times \mathbf{u}) \rangle$, itself a pseudo scalar. The Roberts Cell and CP flows introduced in the preceding chapter are two examples of flows lacking reflectional symmetry.

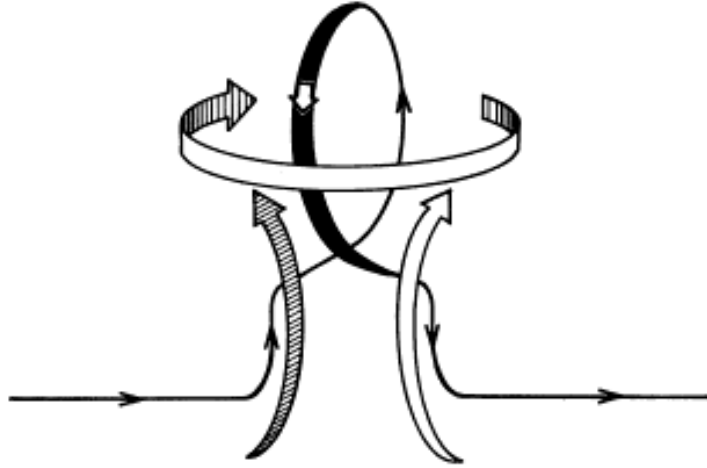


Fig. 33 $\mathbb{A}^{4.1a}$ sketch of magnetic line of force entrained by a cyclonic, rising fluid element in the frozen-in limit. Note that the resulting cyclonic loop can be viewed as resulting from an element of electric current flowing parallel to the original, uniform magnetic field. [from: Parker 1970, *The Astrophysical Journal*, vol. **162**, Figure 1].

As we shall presently see there is an important relation between fluid helicity and the α -effect.

It is important to establish those cases in which α and β can rigorously be computed from knowledge of \mathbf{u} . Not counting methods based on the direct numerical solutions of the induction equation, there are two distinct ways to proceed. In both cases the success of the approach depends on some simplification of equation (204). In one case the term \mathbf{G} is neglected leading to the so-called first order smoothing approximation (FOSA). In the other, the term $\eta\nabla^2\mathbf{b}$ is neglected, leading to the Lagrangian approximation. The two approaches are complementary in the sense that the former is applicable (for most physically relevant circumstances) when the diffusivity is large and the latter when it is small. Even these two most severe simplifying assumptions do not exactly lead to simple mathematics, and to add insult to injury the parameter regimes for which they are expected to hold do not square well with what we think we know about solar interior conditions. The closest we can get to the Sun and stars, in a tractable manner, is the so-called Second-Order Correlation Approximation (SOCA), which neglects cross-correlations between the different velocity components but retains the possibility that the intensity of turbulence itself can vary with position. Under this assumption of near-isotropy, we then have

$$\langle u_j u_k \rangle = \frac{1}{3} \langle \mathbf{u}^2 \rangle \delta_{jk} . \quad (212)$$

This leads to a simple diagonal form for the α tensor:

{eq:SOCAalpha}

$$\alpha = -\frac{1}{3} \tau_c \langle \mathbf{u} \cdot (\nabla \times \mathbf{u}) \rangle, \quad [\text{m s}^{-1}] \quad (213)$$

where τ_c is the correlation time for the turbulent flow. Equation (213) tells us that the α -effect is a direct function of the helicity of turbulent component of the flow; think back of Parker’s picture of twisted magnetic fieldlines (Fig. 33) and convince yourself that this is indeed how it should be for the “cartoon” to work.

If one assumes that the (mild) inhomogeneity arises from the stratification, the (mild) anisotropy from the Coriolis force, and the lifetime of turbulent eddies is commensurate with their turnover time, then eq. (213) can be brought to the form:

$$\alpha = -\frac{1}{3} \tau_c^2 u^2 \boldsymbol{\Omega} \cdot \nabla \ln(\rho u), \quad (214) \quad \{\text{eq:SOCAalpha2}\}$$

where $u = \sqrt{\langle \mathbf{u}^2 \rangle}$ is the local r.m.s. turbulent velocity, and $\boldsymbol{\Omega}$ is the angular velocity vector. With the turbulent velocity increasing outwards through the convective envelope faster than the density decreases, eq. (214) would “predict” an α -effect varying as $\cos \theta$ and positive (negative) in the solar Northern (Southern) hemisphere. Such expression can be validated through MHD numerical simulations of turbulent flows including an externally-imposed weak magnetic field, and from the simulation statistics compute α by appropriate averaging¹⁹. There has been many such simulations, which, almost surprisingly, have corroborated the expressions obtained from SOCA. The key parameter is the so-called Coriolis number, defined as the ratio of rotation period to convective turnover time:

$$\text{Co} = 2\Omega\tau_c, \quad (215) \quad \{\text{eq:SOCAalpha3}\}$$

equivalent to the inverse of the Rossby number of common usage in atmospheric sciences. Estimates for this quantity in the Sun, with τ_c estimated from mixing length theory, yield $\text{Co} \ll 1$ in the outer convection zone, up to $\text{Co} \sim 1$ –10 at the base of the convective envelope. For $\text{Co} \lesssim 1$, the $\alpha_{\phi\phi}$ component of the α -tensor, which is the term responsible for poloidal field regeneration in axisymmetric mean-field models, does turn out positive in the bulk of the convection zone, with a $\sim \cos \theta$ latitudinal dependency. At larger rotation rate, the peak in $\alpha_{\phi\phi}$ is displaced from the pole to lower latitudes, reaching $\sim 30^\circ$ at $\text{Co} \sim 10$. These simulations also produce a sign change in all components of the α -tensor at the very base of the convective envelope, with the region of negative $\alpha_{\phi\phi}$ growing in size as Co increases from 1 to 10.

The above expressions for the α -coefficients are predicated on the small-scale field \mathbf{b} being much weaker than the mean-field $\langle \mathbf{B} \rangle$, a situation expected to hold only in the $\text{R}_m \ll 1$ regime, or if the coherence time of the turbulent

¹⁹ See the papers by Ossendrijver *et al.* (2001) and Käpylä *et al.* (2006) cited in the bibliography, on which the foregoing discussion is based.

flow is much smaller than its turnover time. The first condition is the regime entirely opposite to that expected in solar/stellar interiors, while the second is at best marginally satisfied. High- R_m MHD turbulence simulations indicate that in this regime one has in fact $\mathbf{b} \gtrsim \langle \mathbf{B} \rangle$, and that eq. (213) should be replaced by:

$$\{\text{eq:Pouqalpha}\} \quad \alpha = -\frac{1}{3} \tau_c \left(\langle \mathbf{u} \cdot (\nabla \times \mathbf{u}) \rangle - \frac{1}{\rho} \langle \mathbf{j} \cdot \mathbf{b} \rangle \right). \quad (216)$$

with $\mathbf{j} = (\nabla \times \mathbf{b})/\mu_0$. Notice that the second term on the RHS, corresponding to the current helicity associated with the small-scale magnetic field, has a sign opposite to that kinetic helicity. This says once again, in essence, that the Lorentz force opposes the twisting of the large-scale magnetic field by the turbulent flow. This impact of current helicity on the α -effect represents a potentially powerful quenching mechanism for the α -effect, a topic we shall revisit further below.

\{pch_ssec:pumpturb\} 0.26.3 Turbulent pumping

The non-diagonal part of the α tensor provides a contribution to the turbulent emf taking the form of a non-solenoidal advective velocity (second term on RHS of eq. (209)). This can also be measured in numerical simulations, which indicate that the predominant effect is a downward pumping driven by the stratification, with magnetic fields being expelled from the high-diffusivity regions to the low diffusivity regions. In the presence of rotation turbulent pumping also takes place in the latitudinal direction, with a velocity reaching values of the order of a few meters per second at high rotation rates ($\text{Co} = 10$).

Although turbulent pumping is seldom explicitly included in the simple mean-field dynamo models to be discussed presently, its impact on dynamo action in the sun and solar-type stars is likely important; this is because it can offset flux loss through magnetic buoyancy, and favors accumulation of magnetic fields in the tachocline, where the large shear and low magnetic diffusivity are conducive to the production of strong toroidal flux rope-like structures, believed to give rise to sunspots following their destabilization, buoyant rise through the convection zone and surface emergence.

\{pch_ssec:etaturb\} 0.26.4 The turbulent diffusivity

We now turn to the next term in the expansion (206), namely

$$\{\text{E4.13}\} \quad \mathcal{E}_i^{(2)} = \beta_{ijk} \partial_k \langle B \rangle_j. \quad (217)$$

The physical interpretation of the third-rank pseudotensor, β_{ijk} , is again most easily gained when \mathbf{u} is isotropic, in which case $\beta_{ijk} = \beta\epsilon_{ijk}$, where β is a scalar, and so we have

$$\nabla \times \mathcal{E}^{(2)} = \nabla \times (-\beta \nabla \times \langle \mathbf{B} \rangle) = \beta \nabla^2 \langle \mathbf{B} \rangle. \quad (218) \quad \{\text{E4.14}\}$$

We recognize the scalar β as an additional contribution to the effective diffusivity of $\langle \mathbf{B} \rangle$, which thus becomes $\eta_e \equiv \eta + \beta$. In cases where $\beta \gg \eta$ one refers to $\eta_e \approx \beta$ as the *turbulent diffusivity*. For homogeneous and isotropic turbulence, it can be formally related to the intensity of turbulence as

$$\beta = \frac{1}{3} \tau_c \langle \mathbf{u}^2 \rangle, \quad [\text{m}^2 \text{s}^{-1}], \quad (219) \quad \{\text{eq:SOCAbeta}\}$$

where τ_c is once again the correlation time of the turbulent flow²⁰. Equation (219) states that the turbulent diffusivity is more efficient when the turbulence is more vigorous, which makes intuitive sense since, in order to destroy the magnetic field by folding, the flow must do work against the Lorentz force.

Simple mixing length models of solar convection suggest $u \sim 10 \text{ m s}^{-1}$ and $\tau_c \sim 1$ month at the base of the convection zone ($r/R_\odot \sim 0.7$), which then leads to $\beta \sim 10^8 \text{ m}^2 \text{ s}^{-1}$. This is very much larger than the ordinary magnetic diffusivity $\eta_c \sim 1 \text{ m}^2 \text{ s}^{-1}$, so that we indeed expect $\beta \gg \eta$. This is why, back in the previous chapter (§§0.18.5 and 0.19.4, whenever trying to model the “real” Sun we made use of a magnetic diffusivity profile characterized by a sharp increase when going from the radiative core to the overlying convective envelope (viz. eq. (128) and dash-dotted line on Fig. 7). Note also that the magnetic diffusion time (62) obtained using the above numerical estimate β for the solar convection zone (with $\ell \sim 0.3R_\odot$) is now ~ 10 yr, which is commensurate to the solar cycle period, and suggests that (turbulent) dissipation can be expected to play an important role in solar cycle models.

0.26.5 The mean-field dynamo equations

In summary, our heuristic treatment of mean-field electrodynamics has led us to an evolution equation for the large-scale magnetic field, $\langle \mathbf{B} \rangle$, which takes account of coherences between fluctuation-fluctuation interactions of the small-scale turbulent magnetic and velocity fields. For homogeneous, stationary, and isotropic velocity turbulence, this equation assumes the particularly elegant and physically intuitive form

²⁰ This expression that still holds under SOCA, in which case β becomes a function of position in the flow.

$$\boxed{\frac{\partial \langle \mathbf{B} \rangle}{\partial t} = \nabla \times (\langle \mathbf{U} \rangle \times \langle \mathbf{B} \rangle + \alpha \langle \mathbf{B} \rangle - \beta \nabla \times \langle \mathbf{B} \rangle),} \quad (220) \quad \{\text{E4.15}\}$$

which, according to SOCA, should remain valid in the case of mildly-inhomogeneous, mildly anisotropic turbulence as well, with α and β then given by eqs. (213) and (219). The fluctuation-fluctuation interactions enter this equation through the electromotive force described by the α -effect, and the turbulent diffusion of the mean magnetic field accounted for by β . In principle, these coefficients can be calculated from the lowest-order statistics of the turbulent flow, namely the spatial distribution of turbulent intensity, as measured by $\langle \mathbf{u}^2 \rangle$.

The fact remains that more often than not, and certainly in all mean-field dynamo models to be considered in what follows, the mean-field coefficients α and β will be chosen *a priori*, although we will take care to embody in these choices what we have learned from our brief excursion into mean-field theory. Consequently, the resulting dynamo models will have a descriptive, rather than predictive value. We will be picking numerically “reasonable” turbulent dynamo coefficient that “do the right thing” for the Sun, and see how the resulting models behave as we change other aspects of the model, or, later on, apply them to stars other than the Sun. Yet, as the following example will show, we can still learn a lot from mean-field electrodynamics, even though we have foregone strict physical and mathematical determinism.

{pch_ssec:dynwave}

0.26.6 Dynamo waves

The sunspot butterfly diagram suggests that the deep-seated toroidal magnetic flux system giving rise to sunspots migrates equatorward in the course of the cycle. It turns out that this remarkable pattern can arise naturally in the context of cycle models based on mean-field electrodynamics.

Consider a local cartesian coordinate system oriented so that the direction y corresponds to an ignorable coordinate ($\partial/\partial y = 0$), which we associate with the azimuthal direction in an axisymmetric spherical system, and with x and z mapping onto the latitudinal and radial directions, respectively. Consider now the action of a spatially constant α -effect acting in conjunction with a vertically-sheared flow:

$$\{\text{E4.73}\} \quad \langle \mathbf{U} \rangle = \Omega z \hat{\mathbf{e}}_y, \quad (221)$$

where Ω is a constant [units: s^{-1}]. We shall further assume that the mean-field coefficients α [units: m s^{-1}] and $\eta_e = \beta + \eta$ [units: $\text{m}^2 \text{s}^{-1}$] are constant. The cartesian equivalent of eq. (114) is now

$$\{\text{E4.73b}\} \quad \langle \mathbf{B} \rangle(x, z, t) = \nabla \times (A(x, z, t) \hat{\mathbf{e}}_y) + B(x, z, t) \hat{\mathbf{e}}_y. \quad (222)$$

Substitution of eqs. (221) and (222) into our mean-field induction equation (220) leads to

$$\{\text{E4.74a}\} \quad \frac{\partial A}{\partial t} - \eta_e \left(\frac{\partial^2 A}{\partial x^2} + \frac{\partial^2 A}{\partial z^2} \right) = \alpha B, \quad (223)$$

$$\frac{\partial B}{\partial t} - \eta_e \left(\frac{\partial^2 B}{\partial x^2} + \frac{\partial^2 B}{\partial z^2} \right) = -\Omega \frac{\partial A}{\partial x} + \alpha \left(\frac{\partial^2 A}{\partial x^2} + \frac{\partial^2 A}{\partial z^2} \right). \quad (224) \quad \{\text{E4.74b}\}$$

The two terms on the RHS of this equation parameterize the α -effect and the Ω -effect. Recall that the Ω -effect describes generation of a toroidal magnetic field by the shearing of a poloidal field (as in §0.19.4). The (mean-field) α -effect accounts for the regeneration of *both* poloidal and toroidal magnetic fields due to the chirality, or handedness, of the turbulent flow field. These two terms offer the possibility of dynamo action overcoming the magnetic diffusion term which resides on the LHS of this equation.

Equations (223)–(224) are again PDEs with constant coefficients. We can therefore seek elementary plane-wave solutions of the form

$$\begin{bmatrix} A(x, z, t) \\ B(x, z, t) \end{bmatrix} = \begin{bmatrix} a \\ b \end{bmatrix} \exp[\lambda t + ik(z \cos \vartheta + x \sin \vartheta)]. \quad (225) \quad \{\text{E4.75}\}$$

We may assume that $k \geq 0$ and $0 \leq \vartheta \leq 2\pi$ are prescribed (real) parameters, where the latter sets the orientation of the wavevector in the $[x, z]$ plane. If equation (225) is substituted into eqs. (223)–(224), the requirement that there be nontrivial eigenvectors leads to the dispersion relation:

$$(\lambda + \eta_e k^2)^2 = \alpha k (\alpha k + i\Omega \sin \vartheta). \quad (226) \quad \{\text{E4.76}\}$$

This is a quadratic (complex) polynomial in λ , with the two solutions:

$$\begin{aligned} \lambda_{\pm} = & -\eta_e k^2 \pm \sqrt{\frac{|\alpha|k}{2}} \left\{ \left(\sqrt{\Omega^2 \sin^2 \vartheta + \alpha^2 k^2} + |\alpha|k \right)^{\frac{1}{2}} \right. \\ & \left. + i \operatorname{sign}(\Omega \alpha \sin \vartheta) \left(\sqrt{\Omega^2 \sin^2 \vartheta + \alpha^2 k^2} - |\alpha|k \right)^{\frac{1}{2}} \right\}. \quad (227) \quad \{\text{E4.77}\} \end{aligned}$$

The λ_- solution can only produce a disturbance which decays with the passage of time, so our hope rests on the λ_+ root, with dynamo action occurring when $\operatorname{Re}(\lambda_+) > 0$. Examination of equation (227) indicates that an exponentially growing *dynamo wave* is obtained when $0 < k < k_*$, where the critical wavenumber k_* is one of the (six) roots of the equation,

$$k_*^6 - \frac{\alpha^2}{\eta_e^2} k_*^4 - \frac{\alpha^2 \Omega^2}{4\eta_e^4} \sin^2 \vartheta = 0 . \quad (228) \quad \{\text{E4.78}\}$$

If $k_* \rightarrow 0$ then the “window” for dynamo action disappears. This occurs when $\alpha \rightarrow 0$, in agreement with Cowling’s theorem. From a physical perspective it makes a good deal of sense that the dynamo window inhabits the small-wavenumber, large-wavelength, end of the range of possible parameters. Clearly dynamo waves with rapid spatial fluctuations are susceptible to severe damping due to the enhanced diffusivity $\eta_e \approx \beta$. On the other hand, if the spatial variations of $\langle \mathbf{A} \rangle$ are too large, then there is very little $\langle \mathbf{B} \rangle$ for the α -effect to work on, and so the dynamo process again stalls as $k \rightarrow 0$.

Equation (228) can be solved exactly as a cubic equation for $\zeta \equiv k_*^2$, but for our purposes it is sufficient to simply estimate k_* by inspection of eq. (228) in the limiting cases of “strong” shear, usually most relevant to dynamo action in the sun and stars:

$$\{\text{E4.79}\} \quad k_* \approx \left[\frac{|\alpha \Omega \sin \vartheta|}{2\eta_e^2} \right]^{\frac{1}{3}}, \quad |\alpha| \ll \sqrt{\eta_e |\Omega \sin \vartheta|} . \quad (229)$$

We use the word “wave” to describe these exponentially growing solutions of the mean field equations, because it is clear from equation (227) that $\text{Im}(\lambda_+) \neq 0$. Note also that the direction of propagation clearly depends upon the sign of the product of α and Ω , and that the largest growth rate will occur for $\vartheta = \pi/2$, i.e., wave propagating in the “latitudinal” x -direction, which is a most excellent first step towards reproducing the sunspot butterfly diagram!

0.26.7 The axisymmetric mean-field dynamo equations

{pch_ssec:mfeqs}

We now proceed to reformulate the mean-field induction equation (220) into a form suitable for axisymmetric large-scale magnetic fields pervading a sphere of electrically conducting fluid. We proceed as we did way back in §0.21, which is to express the poloidal field as the curl of a toroidal vector potential, and restrict the large-scale flow to the axisymmetric forms given by eq. (171), with the magnetic diffusivity restricted to vary at most only with r . It will also prove convenient to express the resulting equations in nondimensional form. Toward this end we opt to scale all lengths in terms of R , and time in terms of the diffusion time $\tau = R^2/\eta_e$ based on the (turbulent) diffusivity in the convective envelope, which we assume to be provided by the (scalar) β -term of mean-field electrodynamics. Henceforth dropping the averaging brackets for notational simplicity, the poloidal/toroidal separation procedure applied to the mean-field dynamo equation (220) now leads to

$$\{\text{E5.14a}\} \quad \frac{\partial A}{\partial t} = \eta \left(\nabla^2 - \frac{1}{\varpi^2} \right) A - \frac{R_m}{\varpi} \mathbf{u}_p \cdot \nabla (\varpi A) + C_\alpha \alpha B , \quad (230)$$

$$\begin{aligned} \frac{\partial B}{\partial t} = & \eta \left(\nabla^2 - \frac{1}{\varpi^2} \right) B + \frac{1}{\varpi} \left(\frac{d\eta}{dr} \right) \frac{\partial(\varpi B)}{\partial r} - R_m \varpi \nabla \cdot \left(\frac{B}{\varpi} \mathbf{u}_p \right) \\ & + C_\Omega \varpi (\nabla \times A) \cdot (\nabla \Omega) + C_\alpha \hat{\mathbf{e}}_\phi \cdot \nabla \times [\alpha \nabla \times (A \hat{\mathbf{e}}_\phi)] , \end{aligned} \quad (231) \quad \{\text{E5.14b}\}$$

where the following three nondimensional numbers have materialized:

$$C_\alpha = \frac{\alpha_e R}{\eta_e} , \quad (232) \quad \{\text{E5.17a}\}$$

$$C_\Omega = \frac{\Omega_e R^2}{\eta_e} , \quad (233) \quad \{\text{E5.17b}\}$$

$$R_m = \frac{u_e R}{\eta_e} , \quad (234) \quad \{\text{E5.17c}\}$$

with α_e (dimension m s^{-1}), u_e (dimension $\text{m}^2 \text{s}^{-1}$) and Ω_e (dimension s^{-1}) as reference values for the α -effect, meridional flow and envelope rotation, respectively. Remember that the functionals α , η , \mathbf{u}_p and Ω are hereafter dimensionless. The quantities C_α and C_Ω are *dynamo numbers*, measuring the importance of inductive versus diffusive effects on the RHS of eqs. (230)–(231). The third dimensionless number, R_m , is a magnetic Reynolds number, which here measures the relative importance of advection (by meridional circulation) versus diffusion in the transport of A and B in meridional planes. For simplicity of notation, we continue to use η for the total magnetic diffusivity, retaining the possibility of variation with depth and with the understanding that within the convective envelope this now includes the (dominant) contribution from the β -term of mean-field theory.

Equations (230)–(231) will hereafter be referred to as the *dynamo equations* (rather than the technically preferable but cumbersome “axisymmetric mean-field dynamo equations”). Structurally, they only differs from eqs. (115)–(116) by the presence of not one but two new source terms on the RHS, both associated with the α -effect. The appearance of this term in eq. (230) is crucial, since this is what allows us to evade Cowling’s theorem. Acting in conjunction with the new α -effect term in eq. (231), it makes dynamo action possible in the absence of a large-scale shear, i.e., with $\nabla \Omega = 0$ in eq. (231). Such dynamos are known as α^2 dynamos, and regenerate their magnetic field entirely via the inductive action of small-scale turbulence. Traditionally, dynamo action in planetary cores has been assumed to belong to this variety (at least from the point of view of mean-field theory). Another possibility is that the shearing terms entirely dominates over the α -effect term, in which case the latter is altogether dropped out of eq. (231). This leads to the $\alpha\Omega$ dynamo model, which is believed to be most appropriate to the Sun and solar-type stars. Finally, retaining both source terms in eq. (231) defines, you

guessed it I hope, the $\alpha^2\Omega$ dynamo model. This has received comparatively little attention in the context of solar/stellar dynamos, since (simple) a priori estimates of the dynamo numbers C_α and C_Ω usually yield $C_\alpha/C_\Omega \ll 1$; caution is however warranted if dynamo action takes place in a thin shell, in which case the α -term can still dominate toroidal field production.

In general, solutions are sought in a meridional plane of a sphere of radius R , and as with the diffusive problem of §0.18, are matched to a potential field in the exterior ($r/R > 1$), and regularity requires that $A(r, 0) = A(r, \pi) = 0$ and $B(r, 0) = B(r, \pi) = 0$ be imposed on the symmetry axis. In practice it is often useful to solve explicitly for mode having odd and even symmetry with respect to the equatorial plane. To do so, one simply solves the dynamo equations in a meridional *quadrant*, and imposes the following boundary conditions along the equatorial plane. For a dipole-like antisymmetric mode,

$$\{\text{E5.22b}\} \quad \frac{\partial A(r, \pi/2)}{\partial \theta} = 0, \quad B(r, \pi/2) = 0, \quad [\text{Antisymmetric}], \quad (235)$$

while for symmetric (quadrupole-like) modes one sets instead

$$\{\text{E5.22c}\} \quad A(r, \pi/2) = 0, \quad \frac{\partial B(r, \pi/2)}{\partial \theta} = 0, \quad [\text{Symmetric}]. \quad (236)$$

We are now ready, if not to rock, at least to roll...

{pch_ssec:lmf}

0.26.8 Linear $\alpha\Omega$ dynamo solutions

In constructing mean-field dynamos for the sun, it has been a common procedure to neglect meridional circulation, on the grounds that it is a very weak flow (but more on this further below), and to adopt the $\alpha\Omega$ model formulation, on the grounds that with $R \simeq 7 \times 10^8$ m, $\Omega_0 \sim 10^{-6}$ rad s $^{-1}$, and $\alpha_0 \sim 1$ m s $^{-1}$, one finds $C_\alpha/C_\Omega \sim 10^3$, independently of the assumed (and poorly constrained) value for the turbulent diffusivity. We also restrict the models to the kinematic regime, i.e., all flow fields posed *priori* and deemed steady ($\partial/\partial t = 0$). Equations (230)—(231) then reduce to the so-called $\alpha\Omega$ dynamo equations:

$$\{\text{E5.18a}\} \quad \frac{\partial A}{\partial t} = \left(\nabla^2 - \frac{1}{\varpi^2} \right) A + C_\alpha \alpha B, \quad (237)$$

$$\{\text{E5.18b}\} \quad \frac{\partial B}{\partial t} = \left(\nabla^2 - \frac{1}{\varpi^2} \right) B + C_\Omega \varpi (\nabla \times A \hat{\mathbf{e}}_\phi) \cdot (\nabla \Omega) + \frac{1}{\varpi} \frac{d\eta}{dr} \frac{\partial(\varpi B)}{\partial r}. \quad (238)$$

In the spirit of producing a model that is solar-like we use a fixed value $C_\Omega = 2.5 \times 10^4$, obtained assuming $\Omega_e \equiv \Omega_{Eq} \sim 10^{-6}$ rad s $^{-1}$ and $\eta_e = 5 \times 10^7$ m 2 s $^{-1}$, which leads to a diffusion time $\tau = R^2/\eta_e \simeq 300$ yr.

In the parameter regime characterizing the strongly turbulent solar convection zone, the strength and spatial variation of the α -effect cannot be computed in any reliable manner from first principles, so this will remain the major unknown of the model. In accordance with the $\alpha\Omega$ approximation of the dynamo equations, we restrict ourselves to cases where $|C_\alpha| \ll C_\Omega$. For the dimensionless functional $\alpha(r, \theta)$ we use an expression of the form

$$\alpha(r, \theta) = f(r)g(\theta) , \quad (239) \quad \{\text{E5.26a}\}$$

where

$$f(r) = \frac{1}{4} \left[1 + \operatorname{erf} \left(\frac{r - r_c}{w} \right) \right] \left[1 - \operatorname{erf} \left(\frac{r - 0.8}{w} \right) \right] . \quad (240) \quad \{\text{E5.26b}\}$$

This combination of error functions concentrates the α -effect in the bottom half of the envelope, and lets it vanish smoothly below, just as the net magnetic diffusivity does (i.e., we again set $r_c/R = 0.7$ and $w/R = 0.05$). Various lines of argument point to an α -effect peaking at the bottom of the convective envelope, since there the convective turnover time is commensurate with the solar rotation period, a most favorable setup for the type of toroidal field twisting at the root of the α -effect. Likewise, the hemispheric dependence of the Coriolis force suggests that the α -effect should be positive in the Northern hemisphere, and change sign across the equator ($\theta = \pi/2$). The “minimal” latitudinal dependency is thus

$$g(\theta) = \cos \theta . \quad (241) \quad \{\text{E5.26c}\}$$

The C_α dimensionless number, measuring the strength of the α -effect, is treated as a free parameter of the model. You may be shocked by the fact that we are, in a very very cavalier manner, effectively treating the α -effect as a (almost) free-function; this sorry situation is unfortunately the rule rather than the exception in mean-field dynamo modelling²¹.

With α , β and the large-scale flow given, the $\alpha\Omega$ dynamo equations (237)—(238) become linear in the mean-field \mathbf{B} . With none of the PDE coefficients depending explicitly on time, one can seek eigensolutions of the form

$$\begin{bmatrix} A(r, \theta, t) \\ B(r, \theta, t) \end{bmatrix} = \begin{bmatrix} a(r, \theta) \\ b(r, \theta) \end{bmatrix} e^{\lambda t} , \quad (242) \quad \{\text{E5.19b}\}$$

where the amplitudes a and b are in general complex quantities. Substituting eqs. (242) into the $\alpha\Omega$ dynamo equations yields a classical linear eigenvalue problem. It will prove convenient to write the eigenvalue explicitly as

$$\lambda = \sigma + i\omega , \quad (243) \quad \{\text{E5.20}\}$$

²¹ References to some of the more noteworthy exceptions are provided in the bibliography at the end of this chapter.

so that σ is the growth rate and ω the cyclic frequency, both expressed in terms of the inverse diffusion time $\tau^{-1} = \eta/R^2$. In a model for the (oscillatory) solar dynamo, we are looking for solutions where $\sigma > 0$ and $\omega \neq 0$.

Armed (and dangerous) with the above model, we plow ahead and solve numerically the $\alpha\Omega$ dynamo equations as a 2D eigenvalue problem. We first produce a sequence of solutions for increasing values of $|C_\alpha|$, holding C_Ω fixed at its “solar” value 2.5×10^4 , Figure 34 shows the variation of the growth rate σ and frequency ω as a function of C_α . Four sequences are shown, corresponding to modes that are either antisymmetric or symmetric with respect to the equatorial plane (“A” and “S” respectively), computed with either positive or negative C_α . For $|C_\alpha|$ smaller than some threshold value, the induction terms make too small a contribution to the RHS of eq. (237), leaving the dissipation terms dominant, so that solutions all have $\sigma < 0$, as per Cowling’s theorem. As $|C_\alpha|$ increases, the growth rate eventually reaches $\sigma = 0$. At this point we also have $\omega \neq 0$, so that the corresponding solution oscillates with neither growth of decay of its amplitude. Further increases of $|C_\alpha|$ lead to $\sigma > 0$. We are now finally in the dynamo regime, where a weak initial field is amplified exponentially in time.

Computing similar sequences for the same model but different values of C_Ω soon reveals that the onset of dynamo activity ($\sigma > 0$) is controlled by the *product* of C_α and C_Ω :

$$\{\text{E5.13}\} \quad D \equiv C_\alpha \times C_\Omega = \frac{\alpha_0 \Omega_0 R^3}{\eta_e^2} . \quad (244)$$

The value of D for which $\sigma = 0$ is called the *critical dynamo number* (denoted D_{crit}). This, at least, is similar to what we found for the analytical solution of §0.26.6. Modes having $\sigma < 0$ are called *subcritical*, and those having $\sigma > 0$ *supercritical*. Note on Fig. 34 how little the growth rate and dynamo frequency depend on the assumed solution parity.

Here the first mode to become supercritical is the negative C_α mode, for which $D_{\text{crit}} = -0.9 \times 10^5$, followed shortly by the positive C_α mode ($D_{\text{crit}} = -1.1 \times 10^5$). The dynamo frequency for these critical modes is $\omega \simeq 300$, which corresponds to a full cycle period of ~ 6 yr. This is within a factor of four of the observed full solar cycle period. Once again we should not be too impressed by this, since we have quite a bit of margin of manoeuver in specifying numerical values for η_e and C_α , and there is no reason to believe that the Sun should be exactly exactly at the critical threshold for dynamo action.

Figure 35 shows a half a cycle of the dynamo solution, in the form of snapshots of the toroidal (color scale) and poloidal (fieldlines) eigenfunctions in a meridional plane, with the rotation/symmetry axis oriented vertically²². The four frames are separated by a phase interval $\varphi = \pi/3$, so that panel (D)

²² Animation of this solution, and another for its cousin with negative C_α , can be viewed on the course Web Page.

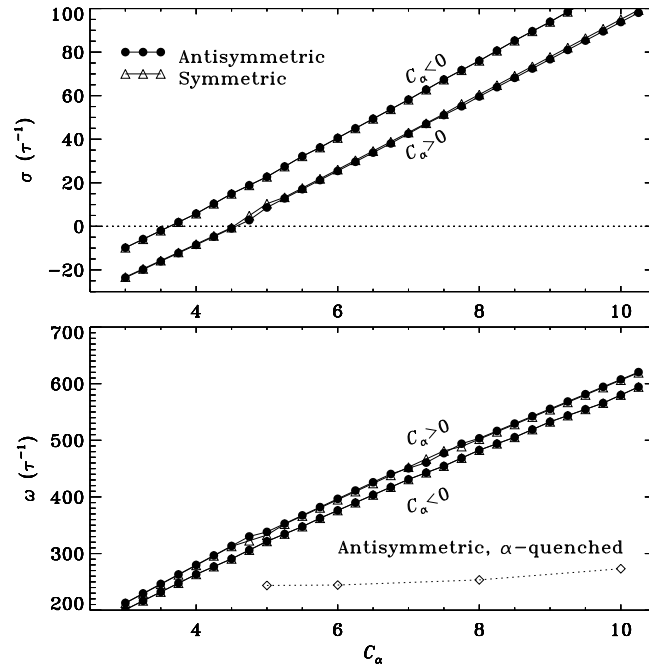


Fig. 34 ^{fig:grates} Variations of the dynamo growth rate (A) and frequency (B) as a function of increasing $|C_\alpha|$ in the minimal $\alpha\Omega$ model with solar-like internal differential rotation. Sequences are shown for either positive or negative dynamo number (as labeled), and symmetric (triangles) or antisymmetric (dots) parity. Modes having $\sigma < 0$ are decaying, and modes with $\sigma > 0$ exponentially growing. Here modes with A or S parity have very nearly identical eigenvalues. In this model the first mode to reach criticality is the negative C_α mode, for which $D_{\text{crit}} = -0.9 \times 10^5$. The positive C_α mode reaches criticality at $D_{\text{crit}} = 1.1 \times 10^5$. The diamonds on panel (B) correspond to the dynamo frequency measured in a nonlinear version of the same minimal $\alpha\Omega$ model, including algebraic α -quenching, to be discussed in §0.26.10.

is identical to (A) except for reversed magnetic polarities in both magnetic components. Such linear eigensolutions leave the *absolute* magnitude of the magnetic field undetermined, but the relative magnitude of the poloidal to toroidal components is found to scale as $\sim |C_\alpha/C_\Omega|$

The toroidal field peaks in the vicinity of the core-envelope interface, which is not surprising since, in view of eqs. (139)—(140), the radial shear is maximal there and the magnetic diffusivity and α -effect are undergoing their fastest variation with depth. But why is the amplitude of the dynamo mode vanishing so rapidly below the core-envelope interface? After all, the poloidal and toroidal diffusive eigenmodes investigated in §0.18 were truly global, and the adopted contrast in magnetic diffusivity between core and envelope should favor stronger fields in the lower diffusivity core. The crucial differ-

BaseCZ $\alpha \sim \cos\theta$ $C_\alpha = +5$ $C_\Omega = 25000$ $Rm = 0$

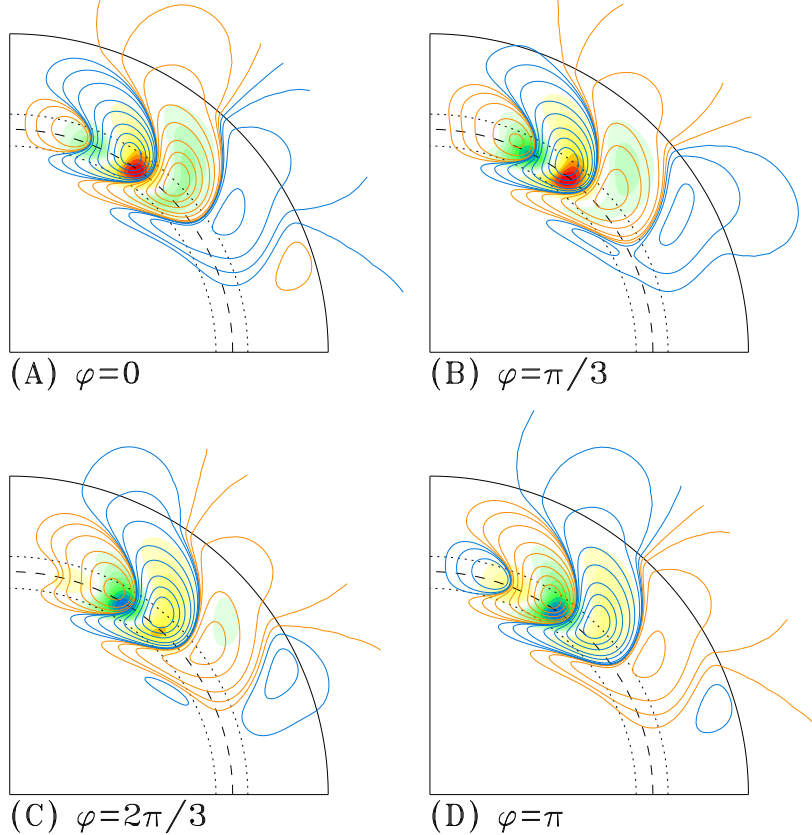


Fig. 35 fig:aolin Four snapshots in meridional planes of our minimal linear $\alpha\Omega$ dynamo solution with defining parameters $C_\Omega = 25000$, $\eta_e/\eta_c = 10$, $\eta_e = 5 \times 10^7 \text{ m}^2 \text{ s}^{-1}$. With $C_\alpha = +5$, this is a mildly supercritical solution (cf. Fig. 34). The toroidal field is plotted as filled contours (green to blue for negative B , yellow to red for positive B , normalized to the peak strength and with increments $\Delta B = 0.2$), on which poloidal fieldlines are superimposed (blue for clockwise-oriented fieldlines, orange for counterclockwise orientation). The dashed line is the core-envelope interface at $r_c/R = 0.7$. The four snapshots shown here cover a half magnetic cycle, i.e., panel (D) is identical to (A) except for reversed magnetic polarities.

ence lies with the oscillatory nature of the solution: because the magnetic field produced in the vicinity of the core-envelope interface is oscillating with alternating polarities, its penetration depth in the core is limited by the electromagnetic skin depth $\ell = \sqrt{2\eta_c/\omega}$ (§0.20), with η_c the core diffusivity. Having assumed $\eta_e = 5 \times 10^7 \text{ m}^2 \text{ s}^{-1}$, we have $\eta_c = \eta_e \Delta\eta = 5 \times 10^6 \text{ m}^2 \text{ s}^{-1}$. A dimensionless dynamo frequency $\omega \simeq 300$ corresponds to $3 \times 10^{-8} \text{ s}^{-1}$, so that $\ell/R \simeq 0.026$, quite small indeed.

Careful examination of 35A→D also reveals that the toroidal/poloidal flux systems present in the shear layer first show up at high-latitudes, and then *migrate equatorward* to finally disappear at mid-latitudes in the course of the half-cycle. If you haven't already guessed it, what we are seeing on Figure 35 is the spherical equivalent of the dynamo waves investigated in §0.26.6 for the cartesian case with uniform α -effect and shear. In more general terms, the dynamo wave travel in a direction \mathbf{s} given by

$$\mathbf{s} = \alpha \nabla \Omega \times \hat{\mathbf{e}}_\phi, \quad (245) \quad \{\text{eq:wave}\}$$

i.e., along isocontours of angular velocity. This result is known as the *Parker-Yoshimura sign rule*. Here with a negative $\partial\Omega/\partial r$ in the high-latitude region of the tachocline, a positive α -effect results in an equatorward propagation of the dynamo wave.

0.26.9 Nonlinearities and α -quenching

{pch_ssec:alpqu}

Obviously, the exponential growth characterizing supercritical ($\sigma > 0$) linear solutions must stop once the Lorentz force associated with the growing magnetic field becomes dynamically significant for the inductive flow. This magnetic backreaction can show up here in two distinct ways:

1. Reduction of the differential rotation,
2. Reduction of turbulent velocities, and therefore of the α -effect (and perhaps also of the turbulent magnetic diffusivity).

Because the solar surface and internal differential rotation shows very little dependence on the phase of the solar cycle, it has been customary to assume that magnetic backreaction occurs at the level of the α -effect. In the mean-field spirit of *not* solving dynamical equations for the small-scales, it has been standard practice to *assume* a dependence of α on B that “does the right thing”, namely reducing the α -effect once the magnetic field becomes “strong enough”, the latter usually taken to mean when the growing dynamo-generated mean magnetic field reaches a magnitude such that its energy per unit volume is comparable to the kinetic energy of the underlying turbulent fluid motions:

$$\frac{B_{\text{eq}}^2}{2\mu_0} = \frac{\rho u_t^2}{2} \rightarrow B_{\text{eq}} = u_t \sqrt{\mu_0 \rho}. \quad (246)$$

This expression defines the *equipartition field strength*, denoted B_{eq} , which varies from ~ 1 T at the base of the solar convective envelope, to ~ 0.1 T in the surface layers. It has become common practice to introduce an *ad hoc* algebraic nonlinear quenching of α (and sometimes η_e as well) directly on the mean-toroidal field B by writing:

$$\alpha \rightarrow \alpha(B) = \frac{\alpha_0}{1 + (B/B_{\text{eq}})^2} . \quad (247) \quad \{\text{eq:alpquench}\}$$

Needless to say, this remains an *extreme* oversimplification of the complex interaction between flow and field that is known to characterize MHD turbulence, but its wide usage in solar dynamo modeling makes it a nonlinearity of choice for the illustrative purpose of this section.

0.26.10 Kinematic $\alpha\Omega$ models with α -quenching

{pch_ssec:nlmf}

With algebraic α -effect included in the poloidal source term, the mean-field $\alpha\Omega$ equations are now nonlinear, and are best solved as an initial-boundary-value problem. The initial condition is an arbitrary seed field of very low amplitude, in the sense that $B \ll B_{\text{eq}}$ everywhere in the domain. Boundary conditions remain the same as for the linear analysis of the preceding section.

Consider again the minimal $\alpha\Omega$ model of §0.26.8, where the α -effect assumes its simplest possible latitudinal dependency, $\propto \cos\theta$. We use again $C_\Omega = 2.5 \times 10^4$ and positive $C_\alpha \geq 5$, so that the corresponding linear solution are in the supercritical regime (see Figure 34). With a very weak \mathbf{B} as initial condition, early on the model is essentially linear and exponential growth is expected. This is indeed what is observed, as can be seen on Fig. 36, showing time series of the total magnetic energy in the simulation domain for increasing values of C_α , all above criticality. Eventually however, B starts to become comparable to B_{eq} in the region where the α -effect operates, leading to a break in exponential growth, and eventual saturation at some constant value of magnetic energy. Evidently, α -quenching is doing what it was designed to do! Note how the saturation energy level increases with increasing C_α , an intuitively satisfying behavior since solutions with larger C_α have a more powerful poloidal source term. The cycle frequency for these solutions is plotted as diamonds on Fig. 34B and, unlike in the linear solutions, now shows very little increase with increasing C_α . Moreover, the dynamo frequency of these α -quenched solutions are found to be slightly *smaller* than the frequency of the linear critical mode (here by some 10–15%), a behavior that is typical of mean-field dynamo models. Yet the overall form of the dynamo solutions closely resembles that of the linear eigenfunctions plotted on Fig. 35. Indeed, the full cycle period is here $P/\tau \simeq 0.027$, which translates into 9 yr for our adopted $\eta_e = 5 \times 10^7 \text{ m}^2 \text{ s}^{-1}$, i.e., a little over a factor of two shorter than the real thing. Not bad!

As a solar cycle model, these dynamo solutions do suffer from one obvious problem: magnetic activity is concentrated at too high latitudes (see Fig. 35). This is a direct consequence of the assumed $\cos\theta$ dependency for the α -effect. One obvious way to push the dynamo mode towards the equator is to concentrate the α -effect at low latitude. This is not as *ad hoc* as one may think, given that the numerical simulation results discussed in §0.26.2

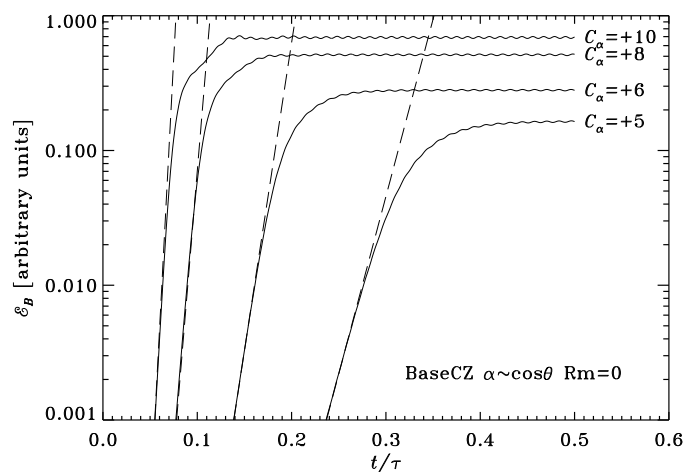


Fig. 36 fig:2tsao Time series of magnetic energy for a set of $\alpha\Omega$ dynamo solutions using our minimal $\alpha\Omega$ model including algebraic α -quenching, and different values for C_α , as labeled. Magnetic energy is expressed in arbitrary units. The dashed line indicates the exponential growth phase characterizing the linear regime.

do indicate that in the high rotation regime ($Co \gtrsim 4$), the peak in the α -effect is indeed displaced to low latitudes. We therefore proceed using now a latitudinal dependency in $\propto \sin^2 \theta \cos \theta$ for the α -effect.

Figure 37 shows a selection of three $\alpha\Omega$ dynamo solutions, in the form of time-latitude diagrams of the toroidal field extracted at the core-envelope interface, here $r_c/R_\odot = 0.7$. If sunspot-producing toroidal flux ropes form in regions of peak toroidal field strength, and if those ropes rise radially to the surface, then such diagrams are directly comparable to the sunspot butterfly diagram. These three models all have $C_\Omega = 25000$, $|C_\alpha| = 10$, $\Delta\eta = 0.1$, and $\eta_e = 5 \times 10^7 \text{ m}^2 \text{ s}^{-1}$. To facilitate comparison between solutions, antisymmetric parity is *imposed* via the boundary condition at the equator. On such diagrams, the latitudinal propagation of dynamo waves shows up as a “tilt” of the flux contours away from the vertical direction.

The first solution, on Figure 37A, is once again our basic solution of Fig. 35, with an α -effect varying in $\cos \theta$. The other two use an α -effect varying in $\sin^2 \theta \cos \theta$, and so manage to produce dynamo action that materializes in two more or less distinct branches, one associated with the negative radial shear in the high latitude part of the tachocline, the other with the positive shear in the low-latitude tachocline. These two branches propagate in opposite directions, in agreement with the Parker-Yoshimura sign rule, since the α -effect here does not change sign within an hemisphere, but the radial gradient of Ω does.

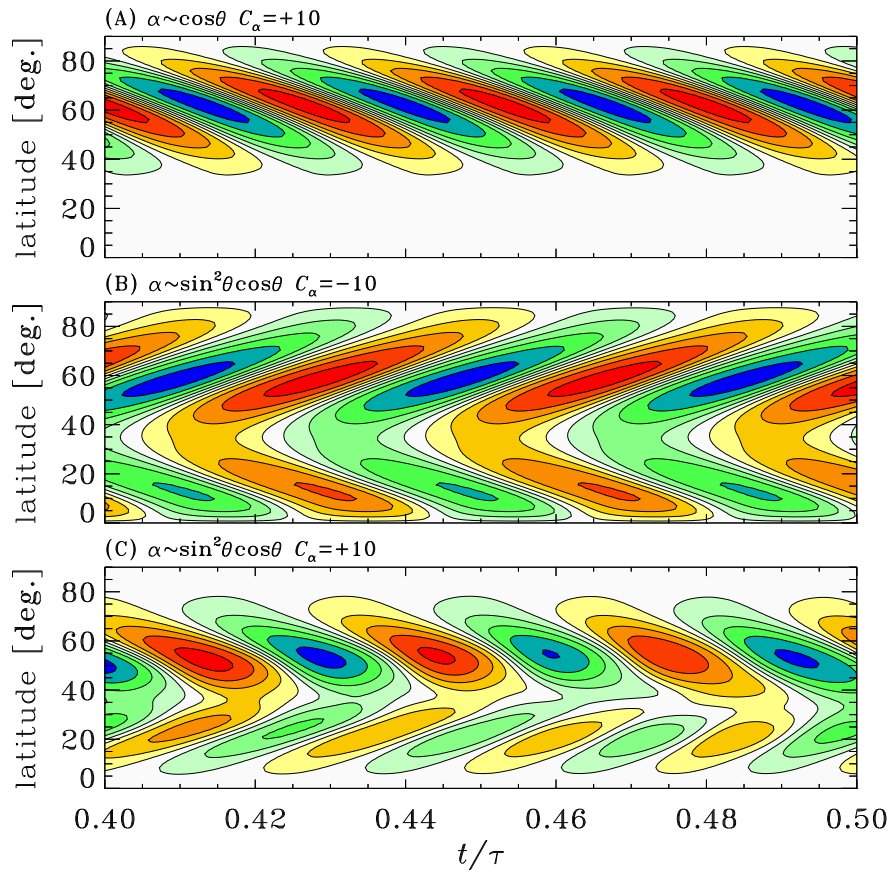


Fig. 37 fig:aosolns Northern hemisphere time-latitude (“butterfly”) diagrams for a selection of nonlinear $\alpha\Omega$ dynamo solutions including α -quenching, constructed at the depth $r/R_{\odot} = 0.7$ corresponding to the core-envelope interface. Isocontours of toroidal field are normalized to their peak amplitudes, and plotted for increments $\Delta B/\max(B) = 0.2$, with yellow-to-red (green-to-blue) contours corresponding to $B > 0$ (< 0). The assumed latitudinal dependence of the α -effect is on given each panel. Other model ingredients as on Fig. 35. Note the co-existence of two distinct cycles in the solution shown on panel C, with periods differing by about 25%.

It is noteworthy that co-existing dynamo branches, as on Fig. 37B and C, can have distinct dynamo periods, which in nonlinearly saturated solutions leads to long-term amplitude modulation. Such modulations are typically not expected in dynamo models where the only nonlinearity present is a simple algebraic quenching formula such as eq. (247). Note that this does not occur for the $C_{\alpha} < 0$ solution, where both branches propagate away from each other, but share a common latitude of origin and so are phased-locked at the onset (cf. Fig. 37B). We are seeing here a first example of potentially

distinct dynamo modes interfering with one another, a direct consequence of the complex profile of solar internal differential rotation.

The solution of Fig. 37B is characterized by a low-latitude equatorially propagating branch, and a full cycle period of 16 yr, which is getting pretty close to the “target” 22yr. But again the strong high-latitude, poleward-propagating branch has no counterpart in the sunspot butterfly diagram. This is often summarily dealt with by flatly zeroing out the α -effect at latitudes higher than $\sim 40^\circ$, but this is clearly not a very satisfying approach. Let’s try something else instead.

0.26.11 Enters meridional circulation: flux transport dynamos

{pch_ssec:aocm}

Meridional circulation is unavoidable in turbulent, compressible rotating convective shells. It basically results from an imbalance between Reynolds stresses and buoyancy forces. The $\sim 15 \text{ m s}^{-1}$ poleward flow observed at the surface has been detected helioseismically, down to $r/R_\odot \simeq 0.85$ without significant departure from the poleward direction, except locally and very close to the surface, in the vicinity of active region belts. Mass conservation evidently requires an equatorward flow deeper down.

Meridional circulation can boldly transport the dynamo-generated magnetic field (terms $\propto \mathbf{u}_p \cdot \nabla$ in eqs. (173)–(174)), and therefore, for a (presumably) solar-like equatorward return flow that is vigorous enough, can overpower the Parker-Yoshimura rule and produce equatorward propagation no matter what the sign of the α -effect is. At low circulation speeds, the primary effect is a Doppler shift of the dynamo wave, leading to a small change in the cycle period. The behavioral turnover from dynamo wave-like solutions to circulation-dominated magnetic field transport sets in when the circulation speed in the dynamo region becomes comparable to the propagation speed of the dynamo wave. In the circulation-dominated regime, the cycle period loses sensitivity to the assumed turbulent diffusivity value, and becomes determined primarily by the circulation’s turnover time. Solar cycle models achieving equatorward propagation of the deep-seated toroidal field in this manner are often called *flux transport dynamos*.

These properties of dynamo solutions with meridional flows can be cleanly demonstrated in simple $\alpha\Omega$ models using a purely radial shear at the core-envelope interface (see references in bibliography), but with a solar-like differential rotation profile the situation turns out to be far more complex. Consider for example the three $\alpha\Omega$ dynamo solutions of Fig. 37, now recomputed including a meridional flow taking the form of a single cell per meridional quadrant, directed poleward in the outer convective envelope and with the equatorward return flow closing at the core-envelope interface, as illustrated on Figure 38A²³. As R_m is increased, for the solution of Fig. 37A, the dynamo

²³ An animation of this solution can be viewed on the course Web Page.

is decaying in $10^2 \lesssim R_m \lesssim 600$, and then kicks in again at $R_m \simeq 800$ with a double-branched structure in its butterfly diagram. The negative- C_α solution (Fig. 37B), on the other hand transits to a steady mode around $R_m \sim 10^2$ that persists at least up to $R_m = 5000$; The solution of Fig. 37C, develops a dominant equatorial branch at $R_m \sim 200$, but a dominant high-latitude branch takes over from $R_m \sim 10^3$ onward.

Figure 38B through I shows half a cycle of our $\alpha \propto \cos \theta$ reference solution, now for parameter values $C_\alpha = 0.5$, $C_\Omega = 5 \times 10^5$, $\Delta\eta = 0.1$, and $R_m = 2500$, which for an envelope diffusivity reduced to $\eta_e = 5 \times 10^6 \text{ m}^2 \text{ s}^{-1}$ corresponds to a solar-like surface poleward flow and differential rotation. The transport of the magnetic field by meridional circulation is clearly apparent, and concentrates the toroidal field to low latitudes, which is great from the point of view of the sunspot butterfly diagram. Note also how poloidal fieldlines suffer very strong stretching in the latitudinal direction within the tachocline (panels C through F), a direct consequence of shearing—in addition to plain transport—by the equatorward flow. One interesting consequence is that induction of the toroidal field is now effected primarily by the *latitudinal* shear within the tachocline, with the radial shear, although larger in magnitude, playing a lesser role since $B_r/B_\theta \ll 1$. The meridional flow also has a profound impact on the magnetic field evolution at $r = R$, as it concentrates the poloidal field in the polar regions. This leads to a large amplification factor through magnetic flux conservation, so that dynamo solutions such as shown on Fig. 38 are typically characterized by very large polar field strengths, here 0.07 T, for an equipartition field strength $B_{\text{eq}} = 0.5 \text{ T}$ in eq. (247). This is only a factor of 4 or so smaller than the toroidal field in the tachocline, even though we have here $C_\alpha/C_\Omega = 10^{-6}$. This concentrated poloidal field, when advected downwards to the polar regions of the tachocline, is responsible for the strong polar branch often seen in the butterfly diagram of dynamo solutions including a rapid meridional flow.

It is noteworthy that to produce a butterfly-like time-latitude diagram of the toroidal field at the core-envelope interface, the required value of R_m in conjunction with the observed surface meridional flow speed and reasonable profile for the internal return flow, ends up requiring a rather low envelope magnetic diffusivity, $\lesssim 10^7 \text{ m}^2 \text{ s}^{-1}$, which stands at the very low end of the range suggested by mean-field estimates such as provided by eq. (219). Still, kinematic $\alpha\Omega$ mean-field models including meridional circulation and simple algebraic α -quenching can produce equatorially-concentrated and equatorially propagating dynamo modes with a period resembling that of the solar cycle for realistic, solar-like differential rotation and circulation profiles. Nice and fine, but it turns out we have another potential problem on our hands.

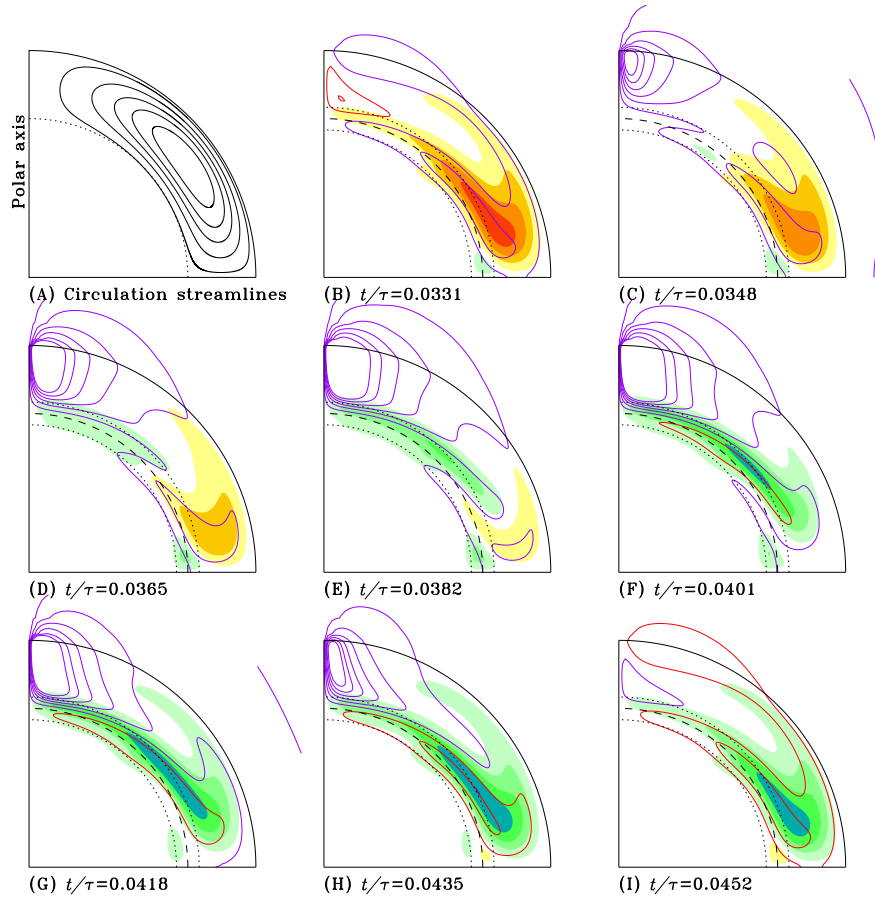


Fig. 38 Snapshots covering half a cycle of an $\alpha\Omega$ dynamo solution including meridional circulation, starting at the time of polarity reversal in the polar surface field. Meridional circulation streamlines are plotted on panel A, the flow being poleward at the surface and equatorward at the core-envelope interface. Color coding of the toroidal field and poloidal fieldlines as on Fig. 35. This α -quenched solution uses the same differential rotation, diffusivity, and α -effect profiles as on Fig. 35, with parameter values $C_\alpha = 0.5$, $C_\Omega = 5 \times 10^5$, $\Delta\eta = 0.1$, $R_m = 2500$. Note the strong amplification of the surface polar fields, the latitudinal stretching of poloidal fieldlines by the meridional flow at the core-envelope interface, and the weak, secondary dynamo mode in the equatorial region of the tachocline. Fig: aocmsoln

{pch_ssec:intrf}

0.26.12 Interface dynamos

The α -quenching expression (eq. (247)) used in the two preceding sections amounts to saying that dynamo action saturates once the mean, dynamo-generated field reaches an energy density comparable to that of the driving turbulent fluid motions, i.e., $B_{\text{eq}} \sim \sqrt{\mu_0 \rho} u$, where u is the turbulent velocity amplitude. This appears eminently sensible, since from that point on a toroidal fieldline would have sufficient tension to resist deformation by cyclonic turbulence, and so could no longer feed the α -effect. At the base of the solar convective envelope, one finds $B_{\text{eq}} \sim 0.1$ T, for $u \sim 10$ m s⁻¹, according to standard mixing length theory of convection. However, various calculations and numerical simulations have indicated that long before the mean toroidal field B reaches this strength, the helical turbulence reaches equipartition with the *small-scale*, turbulent component of the magnetic field. Such calculations also indicate that the ratio between the small-scale and mean magnetic components should itself scale as $R_m^{1/2}$, where $R_m = v\ell/\eta$ is a magnetic Reynolds number based on the turbulent speed but *microscopic* magnetic diffusivity. This then leads to the alternate quenching expression

{eq:stralpqench}

$$\alpha \rightarrow \alpha(B) = \frac{\alpha_0}{1 + R_m (B/B_{\text{eq}})^2} . \quad (248)$$

known in the literature as *strong α -quenching* or *catastrophic quenching*. Since $R_m \sim 10^8$ in the solar convection zone, this leads to quenching of the α -effect for very low amplitudes of the mean magnetic field, of order 10^{-5} T. Even though significant field amplification is likely in the formation of a toroidal flux rope from the dynamo-generated magnetic field, we are now a very long way from the 6–16 T demanded by simulations of buoyantly rising flux ropes and sunspot formation.

A way out of this difficulty was proposed by E.N. Parker in the form of *interface dynamos*. The idea is beautifully simple: if the toroidal field quenches the α -effect, amplify and store the toroidal field away from where the α -effect is operating! Parker showed that in a situation where a radial shear and α -effect are segregated on either side of a discontinuity in magnetic diffusivity taken to coincide with the core-envelope interface, the constant coefficient $\alpha\Omega$ dynamo equations considered already in §0.26.6 support solutions in the form of travelling surface waves localized on the discontinuity. The key aspect of Parker's (linear, cartesian, analytical) solution is that for supercritical dynamo waves, the ratio of peak toroidal field strength on either side of the discontinuity surface is found to scale with the diffusivity ratio as

{eq:bphirat}

$$\frac{\max(B_e)}{\max(B_c)} \sim \left(\frac{\eta_e}{\eta_c} \right)^{-1/2} . \quad (249)$$

If the core diffusivity η_c assumes the microscopic value, and that the envelope diffusivity (η_e) is of turbulent origin so that $\eta_e \sim \ell u$, then the toroidal field

strength ratio then scales as $\sim (u\ell/\eta_c)^{1/2} \equiv R_m^{1/2}$. This is precisely the factor needed to bypass strong α -quenching, at least as embodied in eq. (248).

As an illustrative example, Figure 39A shows a series radial cuts of the toroidal magnetic component at 15° latitude, spanning half a cycle in a numerical interface solution with $C_\Omega = 2.5 \times 10^5$, $C_\alpha = +10$, and a core-to-envelope diffusivity contrast $\Delta\eta = 10^{-2}$. The differential rotation and magnetic diffusivity profiles are the same as before, but here the α -effect is now (even more artificially) concentrated towards the equator, by imposing a latitudinal dependency $\alpha \sim \sin(4\theta)$ for $\pi/4 \leq \theta \leq 3\pi/4$, and zero otherwise.

This model does achieve the kind of toroidal field amplification one would like to see in interface dynamos. Notice how the toroidal field peaks below the core-envelope interface (vertical dotted line), well below the α -effect region and near the peak in radial shear. Figure 39B shows how the ratio of peak toroidal field below and above r_c varies with the imposed diffusivity contrast $\Delta\eta$. The dashed line is the dependency expected from eq. (249). For relatively low diffusivity contrast, $-1.5 \leq \log(\Delta\eta) \lesssim 0$, both the toroidal field ratio and dynamo period increase as $\sim (\Delta\eta)^{-1/2}$. Below $\log(\Delta\eta) \sim -1.5$, the $\max(B)$ -ratio increases more slowly, and the cycle period falls, as can be seen on Fig. 39C. This is basically an electromagnetic skin-depth effect; unlike in the original picture proposed by Parker, here the poloidal field must diffuse down a finite distance into the tachocline before shearing into a toroidal component can commence. With this distance set by our adopted profile of $\Omega(r, \theta)$, as $\Delta\eta$ becomes very small there comes a point where the dynamo period is such that the poloidal field cannot diffuse as deep as the peak in radial shear in the course of a half cycle. The dynamo then runs on a weaker shear, thus yielding a smaller field strength ratio and weaker overall cycle.

0.27 Babcock-Leighton models

{pch_sec:BLsoldyn}

Solar cycle models based on what is now called the Babcock-Leighton mechanism were first developed in the early 1960's, yet they were temporarily eclipsed by the rise of mean-field electrodynamics a few years later. Their revival was motivated in part by the fact that synoptic magnetographic monitoring over solar cycles 21 and 22 has offered strong evidence that the surface polar field reversals are triggered by the decay of active regions (see Fig. 32). The crucial question is whether this is a mere side-effect of dynamo action taking place independently somewhere in the solar interior, or a dominant contribution to the dynamo process itself.

Figure 40 illustrates the basic idea of the Babcock-Leighton mechanism. Consider the bipolar magnetic regions (BMR) sketched on the right. Recall that each of these is the photospheric manifestation of a toroidal flux rope emerging as an Ω -loop (see Fig. 31). The *leading* (*trailing*) component of each BMR is that located ahead (behind) with respect to the direction of the Sun's

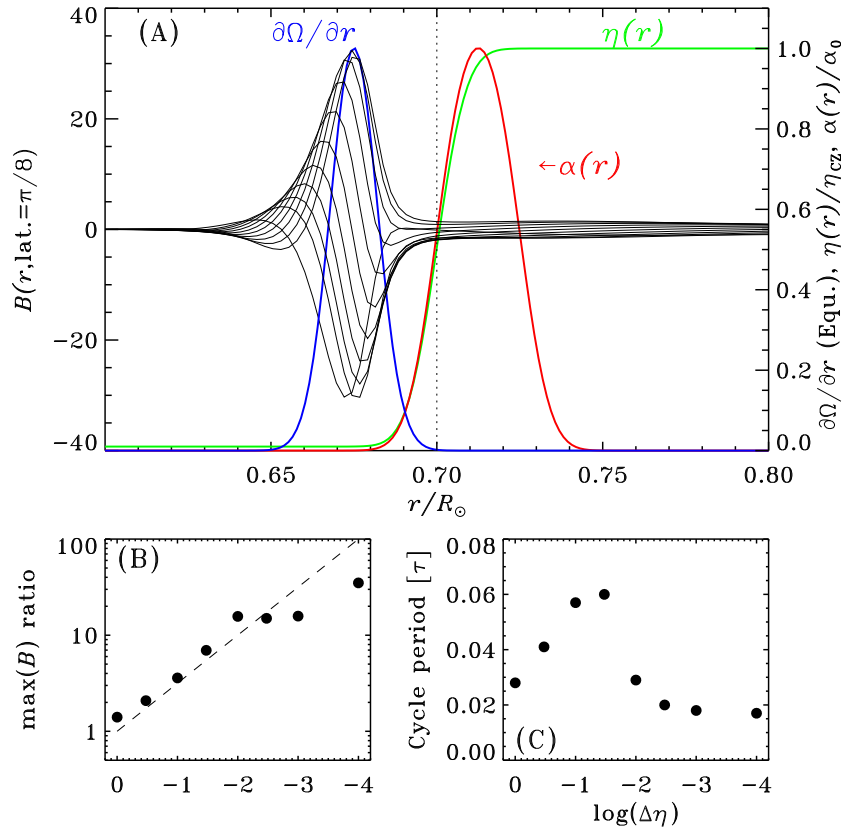


Fig. 39 A representative interface dynamo model in spherical geometry. This solution has $C_\Omega = 2.5 \times 10^5$, $C_\alpha = +10$, and a core-to-envelope diffusivity contrast of 10^{-2} . Panel (A) shows a series of radial cuts of the toroidal field at latitude 15° . The (normalized) radial profiles of magnetic diffusivity, α -effect, and radial shear are also shown, again at latitude 15° . The core-envelope interface is again at $r/R_\odot = 0.7$ (dotted line), where the magnetic diffusivity varies near-discontinuously. Panels (B) and (C) show the variations of the core-to-envelope peak toroidal field strength and dynamo period with the diffusivity contrast, for a sequence of otherwise identical dynamo solutions.

{fig:intrf}

rotation. Joy's Law states that, on average, the leading component is located at lower latitude than the trailing component, so that a line joining each component of the pair makes an angle with respect to the E-W line. Hale's polarity law also informs us that the leading/trailing magnetic polarity pattern is opposite in each hemisphere, a reflection of the equatorial antisymmetry of the underlying toroidal flux system. Horace W. Babcock (1912–2003) demonstrated empirically from his early magnetographic observation of the

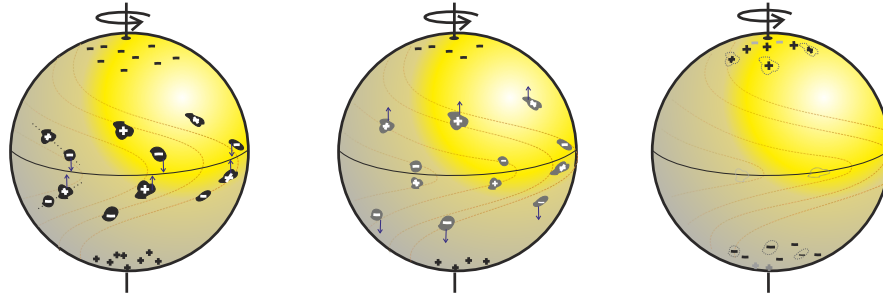


Fig. 40 ^{F6.1} Cartoon of the Babcock-Leighton mechanism. At left, a number of bipolar magnetic regions (BMR) have emerged, with opposite leading/following polarity patterns in each hemisphere, as per Hale's polarity Law. After some time (middle), the BMRs have started decaying, with the leading components experiencing diffusive cancellation across the equator, while the trailing components have moved to higher latitudes. At later time, (right), the net effect is the buildup of an hemispheric flux of opposite polarity in the N and S hemisphere, i.e., a net dipole moment (see text) Diagram kindly provided by D. Passos.

sun's surface solar magnetic field that as the BMRs decay (presumably under the influence of turbulent convection), the trailing components drift to higher latitudes, leaving the leading components at lower latitudes, as sketched on Fig. 40 (middle). Babcock also argued that the trailing polarity poloidal flux released to high latitude by the cumulative effects of the emergence and subsequent decay of many BMRs was responsible for the reversal of the sun's large-scale dipolar field (right).

More germane from the dynamo point of view, the Babcock-Leighton mechanism taps into the (formerly) toroidal flux in the BMRs to produce a poloidal magnetic component. To the degree that a positive dipole moment is being produced from a toroidal field that is positive in the N-hemisphere, this is a bit like a positive α -effect in mean-field theory. In both cases the Coriolis force is the agent imparting a twist on a magnetic field; with the α -effect this process occurs on the small spatial scales and operates on individual magnetic fieldlines. In contrast, the Babcock-Leighton mechanism operates on the large scales, the twist being imparted via the the Coriolis force acting on the flow generated along the axis of a buoyantly rising magnetic flux tube.

0.27.1 Sunspot decay and the Babcock-Leighton mechanism

{pch_ssec:BLmech}

Evidently this mechanism can operate as sketched on Figure 40 provided the magnetic flux in the leading and trailing components of each (decaying) BMR are separated in latitude faster than they can diffusively cancel with one another. Moreover, the leading components must end up at low enough latitudes for diffusive cancellation to take place across the equator. This is

not trivial to achieve, and we now take a more quantitative look at the Babcock-Leighton mechanism, first with a simple 2D numerical model.

The starting point of the model is the grand sweeping assumption that, once the sunspots making up the bipolar active region lose their cohesiveness, their subsequent evolution can be approximated by the passive advection and resistive decay of the radial magnetic field component. This drastic simplification does away with any dynamical effect associated with magnetic tension and pressure within the spots, as well as any anchoring with the underlying toroidal flux system. The model is further simplified by treating the evolution of B_r as a two-dimensional transport problem on a spherical surface corresponding to the solar photosphere. Consequently, no subduction of the radial field can take place.

Even under these simplifying assumptions, the evolution is still governed by the MHD induction equation, specifically its r -component. The imposed flow is made of an axisymmetric surface “meridional circulation”, basically a poleward-converging flow in the latitudinal direction on the sphere, and differential rotation in the azimuthal direction:

$$\{\text{E6.10a}\} \quad \mathbf{u}(\theta) = u_\theta(\theta)\hat{\mathbf{e}}_\theta + \Omega_S(\theta)R\sin\theta\hat{\mathbf{e}}_\phi, \quad (250)$$

where Ω_S is the solar-like surface differential rotation profile used in the preceding chapter (see eq. (140)). Note that in general $\nabla \cdot \mathbf{u} \neq 0$ here, a direct consequence of working on a spherical surface without possibility of subduction. The r -component of the induction equation is cast in non-dimensional form, by expressing length in units of the solar radius R , and time in units of $\tau_c = R/u_0$, i.e., the advection time associated with the meridional flow. Introducing a new latitudinal variable $\mu = \cos\theta$, neglecting all radial derivatives, and evaluating the resulting expression at $r/R = 1$ results in²⁴:

$$\{\text{E6.11}\} \quad \begin{aligned} \frac{\partial B_r}{\partial t} = \frac{\partial}{\partial \mu} \left[(1 - \mu^2)^{1/2} u_\theta B_r \right] - \frac{\Omega_S}{R_u} \frac{\partial B_r}{\partial \phi} \\ + \frac{1}{R_m} \left[\frac{\partial}{\partial \mu} \left((1 - \mu^2) \frac{\partial B_r}{\partial \mu} \right) + \frac{1}{(1 - \mu^2)} \frac{\partial^2 B_r}{\partial \phi^2} \right]. \end{aligned} \quad (251)$$

The solutions are defined in terms of the two nondimensional numbers:

²⁴ This expression is best obtained through the original form of the MHD induction equation (58), rather than the (usually) equivalent form where the $-\nabla \times (\eta \nabla \times \mathbf{B})$ is expressed as $\eta \nabla^2 \mathbf{B}$; the distinction hinges on the metric terms resulting from the application of the Laplacian operator on a vector quantity. Some of have to be forced to zero if radial diffusion is explicitly ignored. More specifically, in the case considered here this would mean zeroing the $-2B_r/r^2$ term in the diffusive term on the RHS of the r -component of the induction equation in spherical coordinates, as given in Appendix 1

$$\{E6.12\} \quad R_m = \frac{u_0 R}{\eta}, \quad R_u = \frac{u_0}{\Omega_0 R}. \quad (252)$$

with u_0 a characteristic speed for the meridional flow, and η the net magnetic diffusivity, assumed constant over the spherical surface defining the solution domain. Using $\Omega_0 = 3 \times 10^{-6} \text{ rad s}^{-1}$, $u_0 = 15 \text{ m s}^{-1}$, and $\eta = 6 \times 10^8 \text{ m}^2 \text{ s}^{-1}$ yields $\tau_c \simeq 1.5 \text{ yr}$, $R_m \simeq 20$ and $R_u \simeq 10^{-2}$. The former is really a measure of the (turbulent) magnetic diffusivity, and is the only free parameter of the model, as u_0 is well constrained by surface Doppler measurements. The corresponding magnetic diffusion time is $\tau_\eta = R^2/\eta \simeq 26 \text{ yr}$, so that $\tau_c/\tau_\eta \ll 1$.

Figure 41 shows a representative solution²⁵, computed assuming a simple analytic form for the meridional flow, namely $u_\theta(\theta) = 2u_0 \sin \theta \cos \theta$. The initial condition (panel A, $t = 0$) describes a series of eight BMRs, four per hemisphere, equally spaced 90° apart at latitudes $\pm 15^\circ$. Each BMR consists of two Gaussian profiles of opposite sign and adding up to zero net flux, with angular separation $d = 10^\circ$ and with a line joining the center of the two Gaussians tilted with respect to the E-W direction²⁶ by an angle γ , itself related to the latitude θ_0 of the BMR's midpoint according to the Joy's Law-like relation:

$$\sin \gamma = 0.5 \cos \theta_0. \quad (253) \quad \{E6.13\}$$

The symmetry of the flow and initial condition on $B_r(\theta, \phi)$ means that the problem can be solved in a single hemisphere with $B_r = 0$ enforced in the equatorial plane, in a 90° wide longitudinal wedge with periodic boundary conditions in ϕ .

The combined effect of circulation, diffusion and differential rotation is to concentrate the magnetic polarity of the trailing “spot” to high latitude, while the polarity of the leading spot dominates at lower latitudes, but experiences diffusive cancellation with the opposite polarity leading flux from its “cousin” in the other solar hemisphere. At mid-latitudes, the effect of differential rotation is to stretch longitudinally the unipolar regions originally associated with each member of the BMR, causing the development of thin banded structures of opposite magnetic polarities, thus enhancing dissipation.

The combined effects of these advection-diffusion processes is to separate in latitude the two polarities of the BMR. This is readily seen upon calculating the longitudinally averaged latitudinal profiles of B_r , as shown on Fig. 42 for the same six successive epochs corresponding to the snapshots on Fig. 41. The poleward displacement of the trailing polarity “bump” is the equivalent to Babcock's original cartoon (cf. Fig. 40). The time required to achieve this

²⁵ An animation of which can of course be viewed on the course Web Page.

²⁶ Remember that this is meant to represent the result of a toroidal flux rope erupting through the surface, so that in this case the underlying toroidal field is positive in the Northern hemisphere, which is the polarity of the trailing “spot”, as measured with respect to the direction of rotation, from left to right here.

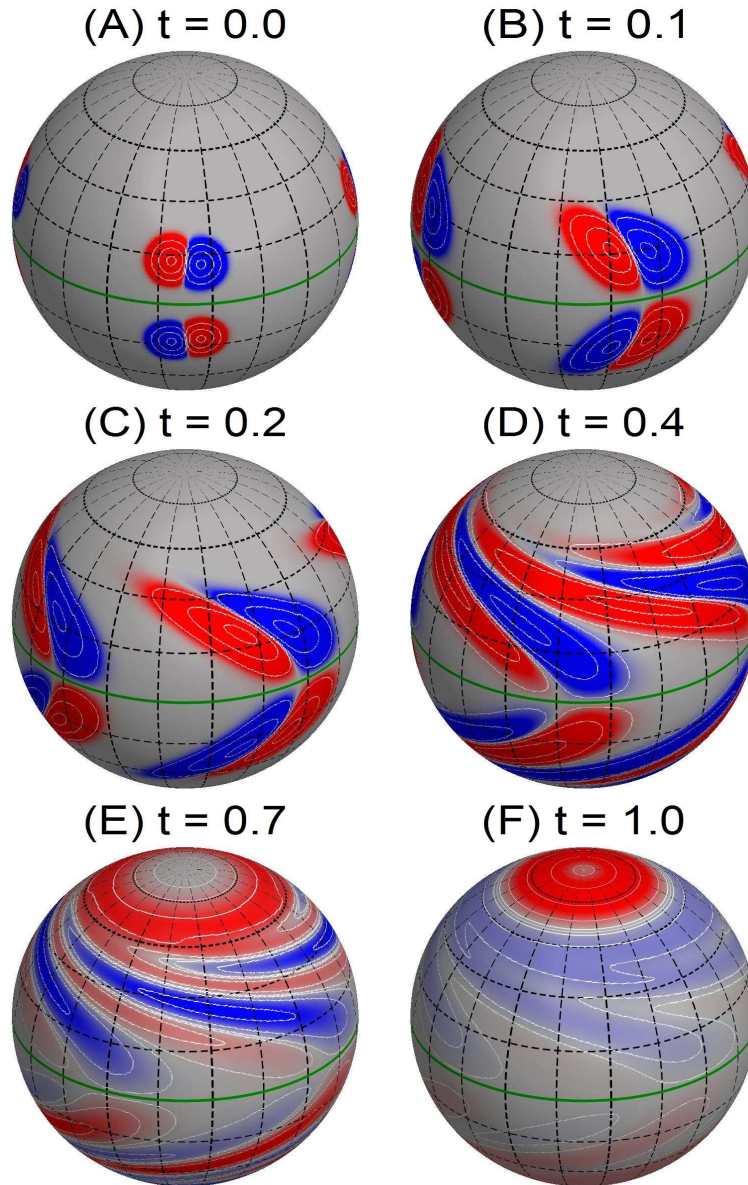


Fig. 41 ^{E6.2} Evolution of the surface radial magnetic field for two sets of four BMRs equally spaced in longitude, and initially located at latitudes $\pm 15^\circ$, with opposite polarity ordering in each hemisphere, as per Hale's polarity Laws. The surface field evolves in response to diffusion and advective transport by differential rotation and a poleward meridional flow, as described by the 2D advection-diffusion equation (251). Parameter values are $R_u = 10^{-2}$ and $R_m = 50$, with time given in units of the meridional flow's characteristic time $\tau_c = R/u_0$.

here is $t/\tau_c \sim 1$, and scales as $(R_m/R_u)^{1/3}$. The significant amplification of the trailing polarity bump from $t/\tau_c \gtrsim 0.5$ onward is a direct consequence of magnetic flux conservation in the poleward-converging meridional flow. Notice also the strong latitudinal gradient in B_r at the equator (dotted line) early in the evolution; the associated trans-equatorial diffusive polarity cancellation affects preferentially the leading spots of each pairs, since the trailing spots are located slightly farther away from the equator.

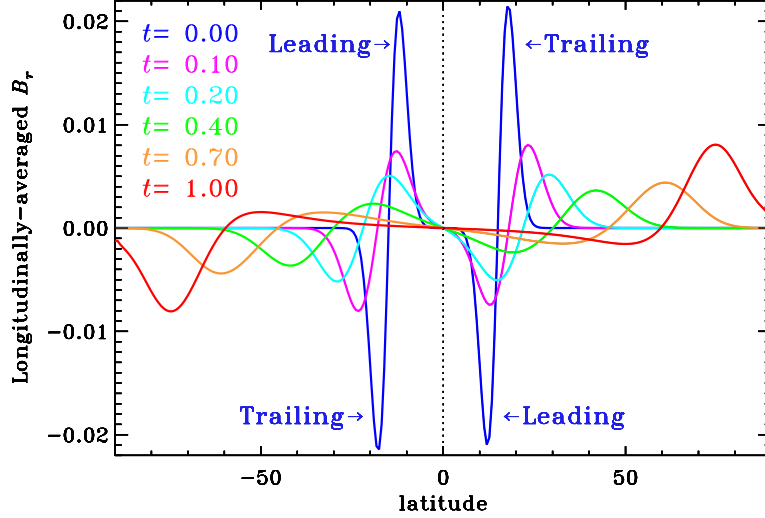


Fig. 42 ^{F6.2B} Latitudinal profile of the longitudinally averaged vertical magnetic field, at the six epochs plotted on Fig. 41. The strong signal at $t = 0$ results entirely from the slight misalignment of the emerging BMRs with respect to the E-W direction. By one turnover time, two polar caps of oppositely-signed magnetic field have built up, amounting to a net dipole moment (see text).

Consider again the mean signed and unsigned magnetic flux:

$$\Phi = \langle |B_r| \rangle, \quad F = \langle |B_r| \rangle, \quad (254) \quad \{\text{E6.13a}\}$$

where the averaging operator is now defined on the spherical surface, for the Northern and Southern hemisphere separately:

$$\langle B_r \rangle = \int_0^{2\pi} \int_{-\pi/2(0)}^{0(\pi/2)} B_r(\theta, \phi) \sin \theta d\theta d\phi. \quad (255) \quad \{\text{E6.13b}\}$$

Figure 43 shows the time-evolution of the signed (Φ , solid line) and unsigned (F , dashed) fluxes in the Northern hemisphere, for the solution of Fig. 41. The unsigned flux decreases rapidly at first, then settles into a slower decay

phase. Meanwhile a small but significant hemispheric signed flux is building up. This is a direct consequence of (negative) flux cancellation across the equator, mediated by diffusion, and is the Babcock-Leighton mechanism in action. Note the dual, conflicting role of diffusion here; it is needed for cross-hemispheric flux cancellation, yet must be small enough to allow the survival of a significant trailing polarity flux on timescales of order τ_c .

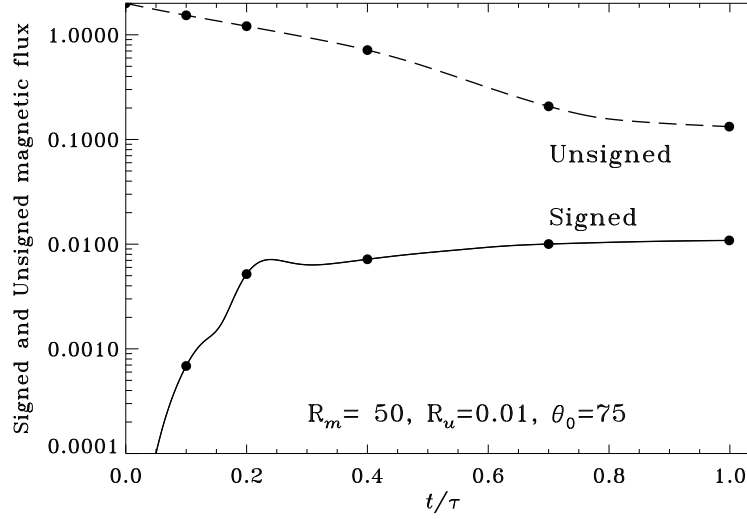


Fig. 43 ^{E6.3} Evolution of the Northern hemisphere signed (solid line) and unsigned (dashed line) magnetic flux for the solution of Fig. 41. The solid dots mark the times at which the snapshots and longitudinal averages are plotted on Figs. 41 and 42.

The efficiency (Ξ) of the Babcock-Leighton mechanism, i.e., converting toroidal to poloidal field, can be defined as the ratio of the signed flux at $t = \tau_c$ to the BMR's initial unsigned flux:

$$\{E6.13c\} \quad \Xi = 2 \frac{\Phi(t = \tau_c)}{F(t = 0)}. \quad (256)$$

Note that Ξ is independent of the assumed initial field strength of the BMRs since eq. (251) is linear in B_r . Looking back at Fig. 43, one would eyeball the efficiency at about 1% in converting the BMR flux to polar cap signed flux. This conversion efficiency turns out to be a rather complex function of BMR parameters; it is expected to grow as with increasing tilt γ , and therefore should increase with latitudes as per Joy's Law, yet proximity to the equator favors transequatorial diffusive flux cancellation of the leading component; moreover, having $du_\theta/d\theta < 0$ favors the separation of the two BMR components, thus minimizing diffusive flux cancellation between the leading and

trailing components. These competing effects lead to a toroidal-to-poloidal conversion efficiency peaking for BMRs emerging at fairly low latitudes, the exact value depending on the latitudinal variation of the adopted surface meridional flow profile. At any rate, we noted already (§0.25) that the sun's polar cap flux peaks at solar minimum, at a value amounting to $\sim 0.1\%$ of the cycle-integrated active region (unsigned) flux; the efficiency required of the Babcock-Leighton mechanism is indeed quite modest.

0.27.2 Axisymmetrization revisited

Take another look at Fig. 41; at $t = 0$ (panel A) the surface magnetic field distribution is highly non-axisymmetric. By $t/\tau_c = 0.7$ (panel E), however, the field distribution shows a far less pronounced ϕ -dependency, especially at high latitudes where in fact B_r is nearly axisymmetric, and by $t/\tau_c = 1$ (panel F) there is little non-axisymmetric field left over the surface. This should remind you of something we encountered earlier: axisymmetrization of a non-axisymmetric magnetic field by an axisymmetric differential rotation (§0.20.5), the spherical analog of flux expulsion. In fact a closer look at the behavior of the unsigned flux on Fig. 43A (dashed line) already shows a hint of the two-timescale behavior we have come to expect of axisymmetrization: the rapid destruction of the non-axisymmetric flux component and slower ($\sim \tau_\eta$) diffusive decay of the remaining axisymmetric flux distribution.

Since the spherical harmonics represent a complete and nicely orthonormal functional basis on the sphere, it follows that the initial condition for the simulation of Fig. 41 can be written as

$$B_r^0(\theta, \phi) = \sum_{l=0}^{\infty} \sum_{m=-l}^{+l} b_{lm} Y_{lm}(\theta, \phi), \quad (257) \quad \{\text{E6.15}\}$$

where the Y_{lm} 's are the spherical harmonics:

$$Y_{lm}(\theta, \phi) = \sqrt{\frac{2l+1}{4\pi} \frac{(l-m)!}{(l+m)!}} P_l^m(\cos\theta) e^{im\phi}, \quad (258) \quad \{\text{E6.15b}\}$$

and with the coefficients b_{lm} given by

$$b_{lm} = \int_0^{2\pi} \int_0^\pi B_r^0(r, \theta) Y_{lm}^*(\theta, \phi), \quad (259) \quad \{\text{E6.16}\}$$

where the “*” indicates complex conjugation. Now, axisymmetrization will wipe out all $m \neq 0$ modes, leaving only the $m = 0$ modes to decay away on the slower diffusive timescale²⁷. Therefore, at the end of the axisymmetrization

²⁷ With $\mathbf{u} = 0$, the decay rate of those remaining modes are given by the eigenvalues of the 2D pure resistive decay problem, much like in §0.18.

process, the radial field distribution now has the form:

$$B_r(\theta) = \sum_{l=0}^{\infty} \sqrt{\frac{2l+1}{4\pi}} b_{l0} P_l^0(\cos \theta), \quad t/\tau_c \gg R_u. \quad (260) \quad \{\text{E6.17}\}$$

which now describes an axisymmetric poloidal magnetic field. Voilà!

0.27.3 Dynamo models based on the Babcock-Leighton mechanism

So now we understand how the Babcock-Leighton mechanism can convert a toroidal magnetic field into a poloidal component, and therefore act as a poloidal source term in eq. (173). Now we need to construct a solar cycle model based on this idea. One big difference with the $\alpha\Omega$ models considered in §0.26 is that the two source regions are now spatially segregated: production of the toroidal field takes place in the tachocline, as before, but now production of the poloidal field takes place in the surface layers.

The mode of operation of a generic solar cycle model based on the Babcock-Leighton mechanism is illustrated in cartoon form on Figure 44. Let P_n represent the amplitude of the high-latitude, surface (“A”) poloidal magnetic field in the late phases of cycle n , i.e., after the polar field has reversed. The poloidal field P_n is advected downward by meridional circulation (A→B), where it then starts to be sheared by the differential rotation while being also advected equatorward (B→C). This leads to the growth of a new low-latitude (C) toroidal flux system, T_{n+1} , which becomes buoyantly unstable (C→D) and starts producing sunspots (D), which subsequently decay and release the poloidal flux P_{n+1} associated with the new cycle $n+1$. Poleward advection and accumulation of this new flux at high latitudes (D→A) then obliterates the old poloidal flux P_n , and the above sequence of steps begins anew. Meridional circulation clearly plays a key role in this “conveyor belt” model of the solar cycle, by providing the needed link between the two spatially segregated source regions. Under this configuration, Babcock-Leighton solar cycle models operate as flux-transport dynamos.

0.27.4 The Babcock-Leighton poloidal source term

The definition of the Babcock-Leighton source term S to be inserted in eq. (173) is evidently the crux of the model. The dynamo solutions presented in what follows use the following:

$$\{\text{eq:BLsource}\} \quad S(r, \theta, B(t)) = s_0 f(r) g(\theta) h(B) B(r_c, \theta, t). \quad (261)$$

with

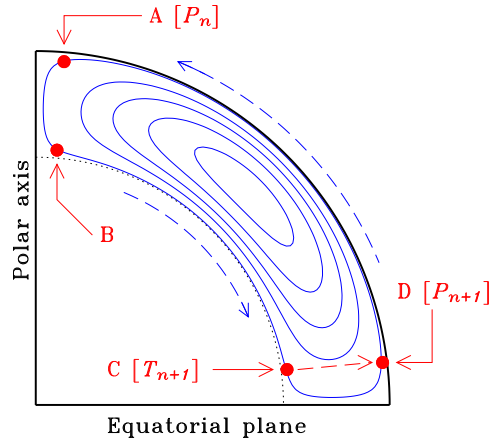


Fig. 44 Operation of a solar cycle model based on the Babcock-Leighton mechanism. The diagram is drawn in a meridional quadrant of the sun, with streamlines of meridional circulation plotted in blue. Poloidal field having accumulated in the surface polar regions (“A”) at cycle n must first be advected down to the core-envelope interface (dotted line) before production of the toroidal field for cycle $n + 1$ can take place (B→C). Buoyant rise of flux rope to the surface (C→D) is a process taking place on a much shorter timescale. Fig: BLc belt

$$\{\text{eq:BLs1}\} \quad f(r) = \frac{1}{2} \left[1 + \operatorname{erf} \left(\frac{r - r_2}{d_2} \right) \right] \left[1 - \operatorname{erf} \left(\frac{r - r_3}{d_3} \right) \right], \quad (262)$$

$$g(\theta) = \sin \theta \cos \theta, \quad (263) \quad \{\text{eq:BLs2}\}$$

$$h(B) = \left[1 + \operatorname{erf} \left(\frac{B(r_c, \theta, t) - B_1}{w_1} \right) \right] \left[1 - \operatorname{erf} \left(\frac{B(r_c, \theta, t) - B_2}{w_2} \right) \right], \quad (264) \quad \{\text{eq:BLs3}\}$$

where s_0 is a numerical coefficient setting the strength of the source term (corresponding dynamo number being $C_S = s_0 R / \eta_0$), and with the various remaining numerical coefficients taking the values $r_2/R = 0.95$, $r_3/R = 1$, $d_2/R = d_3/R = 10^{-2}$, $B_1 = 6$, $B_2 = 10$, $w_1 = 2$, and $w_2 = 8$, the latter four all in Tesla. Note that the dependency on B is *non-local*, i.e., it involves the toroidal field evaluated at the core-envelope interface r_c , (but at the same polar angle θ). This nonlocality in B represents the fact that the strength of the source term is proportional to the field strength in the bipolar active region, itself presumably reflecting the strength of the diffuse toroidal field near the core-envelope interface, where the magnetic flux ropes eventually giving rise to the bipolar active region originate. The combination of error functions in eq. (264) restricts the operating range of the model to a finite interval in toroidal field strength, and is motivated by simulations of the stability and

buoyantly rise thin flux tubes, as discussed in §0.25.4. Other equally reasonable prescriptions and modelling approaches are of course possible (see bibliography at the end of this chapter).

At any rate, inserting this source term into eq. (173) is what we need to bypass Cowling’s theorem and produce a viable dynamo model. The non-locality of S notwithstanding, at this point the model equations are definitely mean-field like. Yet no averaging on small scales is involved. What is implicit in eq. (261) is some sort of averaging process at least in longitude and time, over many BMR emergences.

{pch_ssec:BLsoln} 0.27.5 A sample solution

Figure 45 shows a series of meridional quadrant snapshot of one such Babcock-Leighton dynamo solution, in the now usual format²⁸. The Figure covers a half-cycle, corresponding to one sunspot cycle, starting approximately at the time one would identify as sunspot minimum, with sunspot maximum (based on magnetic energy as a SSN proxy) occurring between panels E and F, and reversal of the polar field shortly thereafter, between panels F and G. As with the advection-dominated $\alpha\Omega$ solution of the preceding section, this solution is characterized by an equatorward propagation of the toroidal field in the tachocline driven by the meridional flow. The turnover time of the meridional flow is here again the primary determinant of the cycle period. With $\eta_e = 3 \times 10^{11} \text{ cm}^2 \text{ s}^{-1}$, this solution has a nicely solar-like half-period of 12.4 yr. All in all, this is once again a reasonable representation of the cyclic spatiotemporal evolution of the solar large-scale magnetic field.

The strong toroidal fields building up within the polar regions of the tachocline in the course of the cycle (see panel C through G on Fig. 45) are entirely unrelated to the adopted latitudinal dependency of the Babcock-Leighton source term. It results instead from the strong polar field advected downwards by the meridional flow, inducing a toroidal component through the inductive action of both the latitudinal shear within the convective envelope, and the negative radial shear in the polar regions of the tachocline. Here this toroidal component mostly decays away under the influence of Ohmic dissipation, and contributes very little to the production of the next cycle’s poloidal component, which builds up at lower latitude (panel E) and is then carried poleward by the meridional flow (panels E→H).

Although it exhibits the desired equatorward propagation, the toroidal field butterfly diagram on Fig. 45A peaks at much higher latitude ($\sim 45^\circ$) than the sunspot butterfly diagram ($\sim 15^\circ\text{--}20^\circ$). This occurs because this is a solution with high magnetic diffusivity contrast, where meridional circulation closes at the core-envelope interface, so that the *latitudinal* component

²⁸ ...and guess where you have to go to view an animation of this solution...?

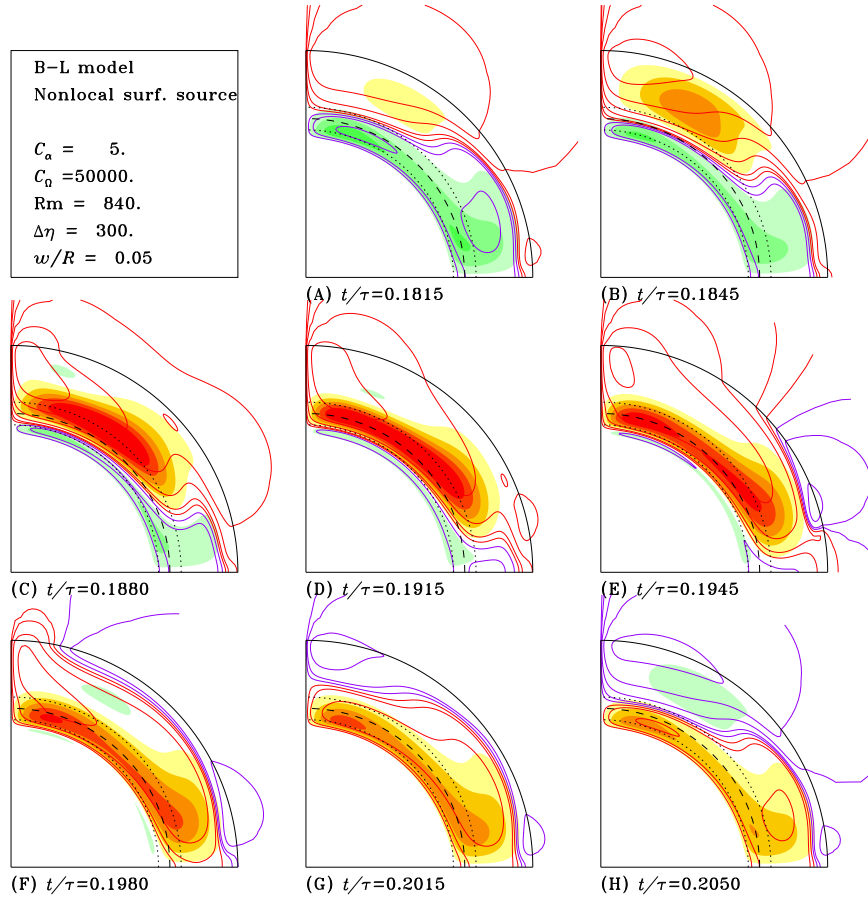


Fig. 45 Snapshot covering half a cycle of a Babcock-Leighton dynamo solution, as described in the text. Color coding of the toroidal field and poloidal fieldlines as on Fig. 35. This solution uses the same differential rotation, magnetic diffusivity and meridional circulation profile as for the advection-dominated $\alpha\Omega$ solution of §0.26.11, but now with the non-local surface source term defined through eq. (261), with parameter values $C_\alpha = 5$, $C_\Omega = 5 \times 10^4$, $\Delta\eta = 0.003$, $Rm = 840$. Note again the strong amplification of the surface polar fields, the latitudinal stretching of poloidal fieldlines by the meridional flow at the core-envelope interface. FIG. 45 CONT.

of differential rotation dominates the production of the toroidal field. This difficulty can be alleviated by letting the meridional circulation penetrate below the core-envelope interface, but this often leads to the production of a strong polar branch, again a consequence of both the strong radial shear present in the high-latitude portion of the tachocline, and of the concentration of the poloidal field taking place in the high latitude-surface layer prior to this field being advected down into the tachocline by meridional circulation (viz. Figs. 44 and 45). Another interesting option to avoid excessive

polar field amplification is to rely on turbulent pumping to carry the surface field downward into the convection zone faster than it can accumulate at the poles.

A noteworthy property of this class of model is the dependency of the cycle period on model parameters; over a wide portion of parameter space, the meridional flow speed is found to be the primary determinant of the cycle period (P). This behavior arises because, in these models, the two source regions are spatially segregated, and the time required for circulation to carry the poloidal field generated at the surface down to the tachocline is what effectively sets the cycle period. The corresponding time delay introduced in the dynamo process has rich dynamical consequences, to be discussed in §0.33 below. On the other hand, P is found to depend very weakly on the assumed values of the source term amplitude s_0 , and turbulent diffusivity η_e ; this is very much unlike the behavior typically found in mean-field models, where P scales nearly as η_e^{-1} in α -quenched $\alpha\Omega$ mean-field models.

{pch_sec:instsoldyn}

0.28 Models based on HD and MHD instabilities

In the presence of stratification and rotation, a number of hydrodynamical (HD) and magnetohydrodynamical (MHD) instabilities associated with the presence of a strong toroidal field in the stably stratified, radiative portion of the tachocline can lead to the growth of disturbances with a net kinetic helicity, which under suitable circumstances can produce a toroidal electromotive force, and therefore act as a source of poloidal field. Different types of solar cycle models have been constructed in this manner, two promising ones being briefly reviewed in this section. In both cases the resulting dynamo models end up being described by something closely resembling our now well-known axisymmetric mean-field dynamo equations, the novel poloidal field regeneration mechanisms being once again subsumed in an α -effect-like source term appearing of the RHS of eq. (173).

{pch_ssec:jmhd}

0.28.1 Models based on shear instabilities

Hydrodynamical stability analyses of the latitudinal shear profile in the solar tachocline indicate that the latter may be unstable to non-axisymmetric perturbations, with the instabilities planforms characterized by a net kinetic helicity, which, loosely inspired by eq. (213), allows the construction of an azimuthally-averaged α -effect-like source term that is directly proportional to the large-scale toroidal component, just as in mean-field electrodynamics.

Figure 46 shows representative time-latitude diagrams of the toroidal magnetic field at the core-envelope interface, and surface radial field, for a flux

transport dynamo solution based on this poloidal source mechanism. This is a solar-like solution with a mid-latitude surface meridional (poleward) flow speed of 17 m s^{-1} , envelope diffusivity $\eta_T = 5 \times 10^7 \text{ m}^2 \text{ s}^{-1}$, a core-to-envelope magnetic diffusivity contrast $\Delta\eta = 10^{-3}$, and a simple α -quenching-like amplitude nonlinearity²⁹. Note the equatorward migration of the deep toroidal field, set here by the meridional flow in the deep envelope, and the poleward migration and intensification of the surface poloidal field, again a direct consequence of advection by meridional circulation, as in the mean-field dynamo models discussed in §0.26.11) in the advection-dominated, high R_m regime. The three-lobe structure of each spatiotemporal cycle in the butterfly diagram reflects the latitudinal structure in kinetic helicity profiles associated with the instability planforms.

The primary weakness of these models, in their present form, is their reliance on a linear stability analysis that altogether ignores the destabilizing effect of magnetic fields, especially since stability analyses have shown that the MHD version of the instability is easier to excite for toroidal field strengths of the magnitude believe to characterize the solar tachocline. Moreover, the planforms in the MHD version of the instability are highly dependent on the assumed underlying toroidal field profile, so that the kinetic helicity can be expected to (1) have a time-dependent latitudinal distribution, and (2) be intricately dependent on the mean toroidal field in a manner that is unlikely to be reproduced by a simple amplitude-limiting quenching formula.

0.28.2 Models based on flux tube instabilities

{pch_ssec:instb}

As briefly discussed in §0.25, modelling of the rise of thin toroidal flux tubes throughout the solar convection zone has met with great success, in particular in reproducing the latitudes of emergence and tilt angles of bipolar sunspot pairs. It is also possible to use the thin-flux tube approximation to study the stability of toroidal flux ropes stored immediately below the base of the convection zone, to investigate the conditions under which they can actually be destabilized and give rise to sunspots. Once the tube destabilizes, calculations show that under the influence of rotation, the correlation between the flow and field perturbations is such as to yield a mean azimuthal electromotive force, equivalent to a *positive* α -effect in the N-hemisphere.

Figure 47 shows a stability diagram for this flux tube instability, in the form of growth rate contours in a 2D parameter space comprised of flux tube strength and latitudinal position at the core-envelope interface. The key is now to identify regions where weak instability arises (growth rates $\gtrsim 1 \text{ yr}$). In the case shown on Fig. 47, these regions are restricted to flux tube strengths in the approximate range 6–15 T.

²⁹ See the Dikpati & Gilman (2001) paper cited in the bibliography for more details.

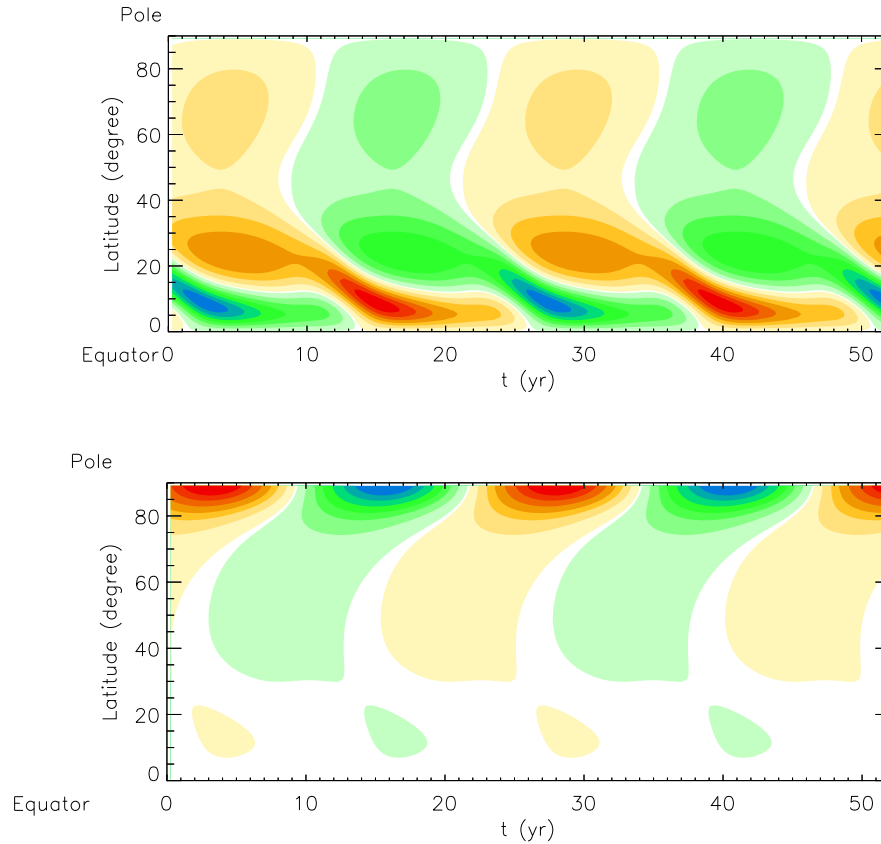


Fig. 46 Time-latitude “butterfly” diagrams of the toroidal field at the core-envelope interface (top), and surface radial field (bottom) for a representative dynamo solution with the tachocline α -effect of Dikpati & Gilman. This solution has a solar-like half-period of eleven years. Note how the deep toroidal field peaks at very low latitudes, in good agreement with the sunspot butterfly diagram. For this solution the equatorial deep toroidal field and polar surface radial field lag each other by $\sim \pi$, but other parameter settings can bring this lag closer to the observed $\pi/2$. Diagrams kindly provided by M. Dikpati.

{fig:shwsolns}

Although it has not yet been comprehensively studied, this dynamo mechanism has a number of very attractive properties. It operates without difficulty in the strong field regime (in fact in *requires* strong fields to operate). It also naturally yields dynamo action concentrated at low latitudes. Difficulties include the need of a relatively finely tuned magnetic diffusivity to achieve a solar-like dynamo period, and a relatively finely-tuned level of subadiabaticity in the overshoot layer for the instability to kick on and off at the appropriate toroidal field strengths.

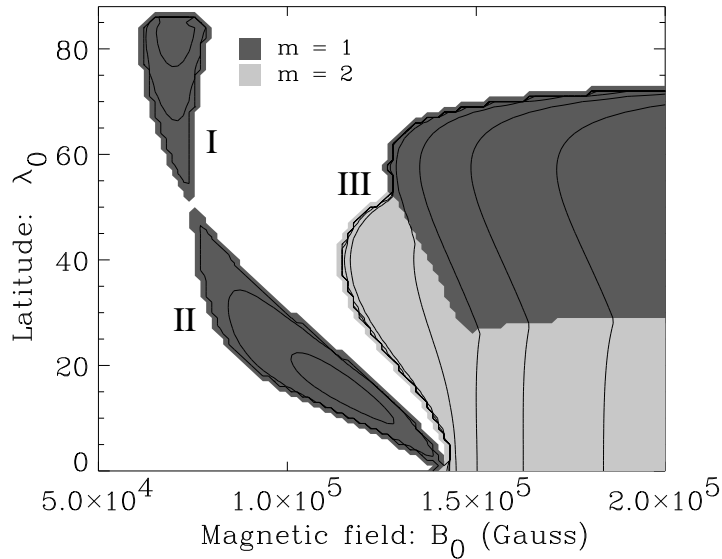


Fig. 47 Stability diagram for toroidal magnetic flux tubes located in the overshoot layer immediately beneath the core-envelope interface. The plot shows contours of growth rates in the latitude-field strength plane. The gray scale encodes the azimuthal wavenumber of the mode with largest growth rate, and regions left in white are stable. Dynamo action is associated with the regions with growth rates ~ 1 yr, here labeled I and II. Region III is associated with the rapid destabilisation, buoyant rise and emergence of magnetic flux, without significant dynamo action. Diagram kindly provided by A. Ferriz-Mas.

{fig:flstab}

The effects of meridional circulation in this class of dynamo models has yet to be investigated; this should be particularly interesting, since both analytic calculations and numerical simulations suggest a *positive* α -effect in the Northern-hemisphere, which should then produce *poleward* propagation of the dynamo wave at low latitude. Meridional circulation could then perhaps produce equatorward propagation of the dynamo magnetic field even with a positive α -effect, as it does in true mean-field models (cf. §0.26.11).

As an interesting aside, note on Fig. 47 how flux tubes located at high latitudes are always stable; this is due to the stabilizing effect of magnetic tension associated with high curvature of the toroidal flux ropes. Even if flux ropes were to form there, they may not necessarily show up at the surface as sunspots. This should be kept in mind when comparing time-latitudes diagrams produced by this or that dynamo model to the sunspot butterfly diagram; the two may not map onto one another as well as often implicitly assumed.

0.29 Global MHD simulations

{pch_sec:globalMHD}

After this grand tour of (relatively) simple solar cycle models, it is worth briefly looking at the theoretical “real thing”, namely full MHD simulations of thermally-driven convection in a thick, stratified rotating spherical shell, across which a solar heat flux is forced to flow. In the solar and stellar dynamo context, the MHD equations (76)—(80) are solved numerically under either one of two physical approximations. Under the *Boussinesq approximation* the fluid density ρ is considered constant, except where it multiplies gravity on the RHS of the momentum equation, thus retaining the effect of thermal buoyancy. In this case eq. (76) reverts to the form appropriate for an incompressible fluid, i.e., $\nabla \cdot \mathbf{u} = 0$. The *anelastic approximation* retains the possibility that ρ be a function of position, and is thus better applicable to stratified environment, but still precludes any temporal variation of density other than associated with thermal dilation. The mass conservation equation is then replaced by $\nabla \cdot (\rho \mathbf{u}) = 0$. In both cases the primary practical advantage is to filter out sound waves and the magnetosonic wave modes, which it turns allow the use of a much larger time step than in fully compressible MHD. Both approximations still capture the Alfvén wave mode, and the anelastic approximation also allows the propagation of gravity waves in stably stratified environments.

We focus in what follows on a specific set of simulations carried out in the anelastic regime³⁰. The simulation domain includes most the convection zone (here $0.718 \lesssim t/R_{\odot} \lesssim 0.96$), as well as a stably stratified fluid layer underneath ($0.61 \lesssim r/R_{\odot} \lesssim 0.718$). The background stratification is solar-like, and covers 4 scale heights in density.

Figure 48 shows time series of the kinetic and magnetic energies in a typical simulation, starting from a static, unmagnetized configuration ($\mathbf{u} = 0$, $\mathbf{B} = 0$) with small random seed magnetic field and velocity perturbation introduced at $t = 0$. Thermal convection sets in very rapidly, and leads to a rapid growth of kinetic energy across the convectively unstable part of the simulation domain in the first few solar days ($1 \text{ sd} \equiv 30$ Earth days). Once convection has stabilized, small-scale dynamo action powered by this turbulent flow commences, and leads to the exponential growth of magnetic energy. Here his phase of exponential growth lasts up to ~ 15 solar days, after which the Lorentz force starts to backreact on the turbulent flow, leading to a saturation of the magnetic energy reminiscent of α -quenching, cf. Fig. 36), and completed at ~ 20 solar days.

Figure 49 shows snapshots at $t = 100 \text{ sd}$ of the radial flow (top) and magnetic field (middle) components extracted near the top of the simulation domain. The morphological asymmetry between the broad, diffuse upflows and narrow concentrated downflows is quite typical of thermally-driven convec-

³⁰ The results presented here are taken pretty directly from the Ghizaru, Charbonneau & Smolarkiewicz paper listed in the bibliography.

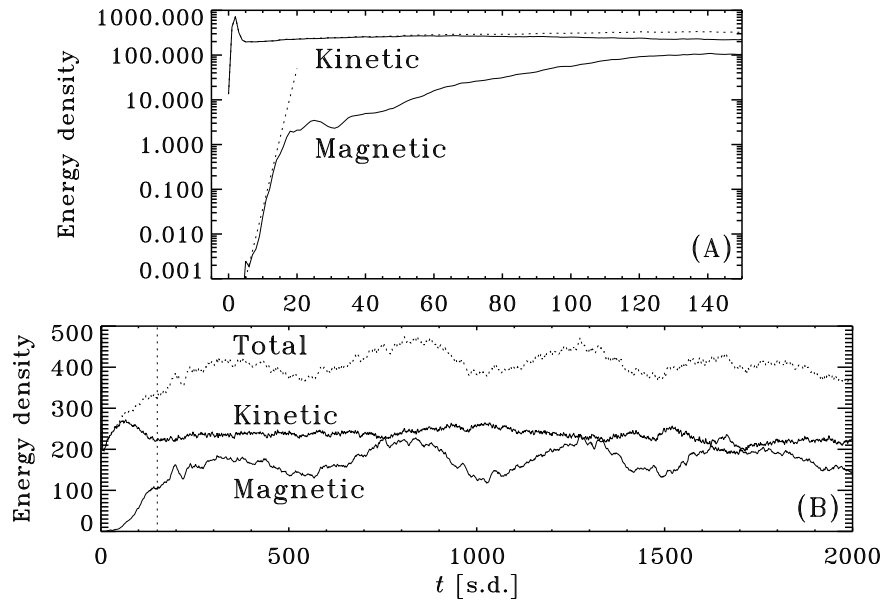


Fig. 48 Time series of kinetic and magnetic energy densities in a 3D anelastic MHD simulation of thermally-driven convection in a thick, stratified rotating spherical shell. Part (A) is a closeup on the first 150 solar days (one solar day \equiv 30 terrestrial days), and part (B) shows the evolution over 2000 solar days (\equiv 165 yr). The phase of exponential growth in the magnetic energy ($5 \lesssim t \lesssim 15$, dotted line segment) begins once the convection has attained a statistically stationary state. The slower growth kicking in at $r \simeq 50$ is associated with the buildup of a large-scale magnetic component, which eventually develops cyclic polarity reversals, leading to the long-timescale modulation of the magnetic energy time series visible on part (B).

{fig:3DMHD1}

tion in a stratified environment. The magnetic field is swept horizontally in the broad areas of upflow and ends up preferentially concentrated in regions where downflow lanes meet, a feature that is typical of MHD convection. As with the CP flow solutions considered earlier, the subsurface magnetic field is spatially and temporally very intermittent, and is characterized by significant magnetic energy but very little net magnetic flux on large spatial scales³¹.

By this time, at least on the basis of these energy time series, one would judge the system to have reached a statistically stationary state. However, integrating further in time reveals variations setting in on longer timescales, associated with the slow buildup a large-scale magnetic field. This is already apparent on Fig. 48. By about 100 solar days, this large-scale component has reached a strength such that it begins to quench the differential rotation having built up in the earlier phases of the simulation through the action of

³¹ Various animations pertaining to this simulation can be viewed on the course Web page.

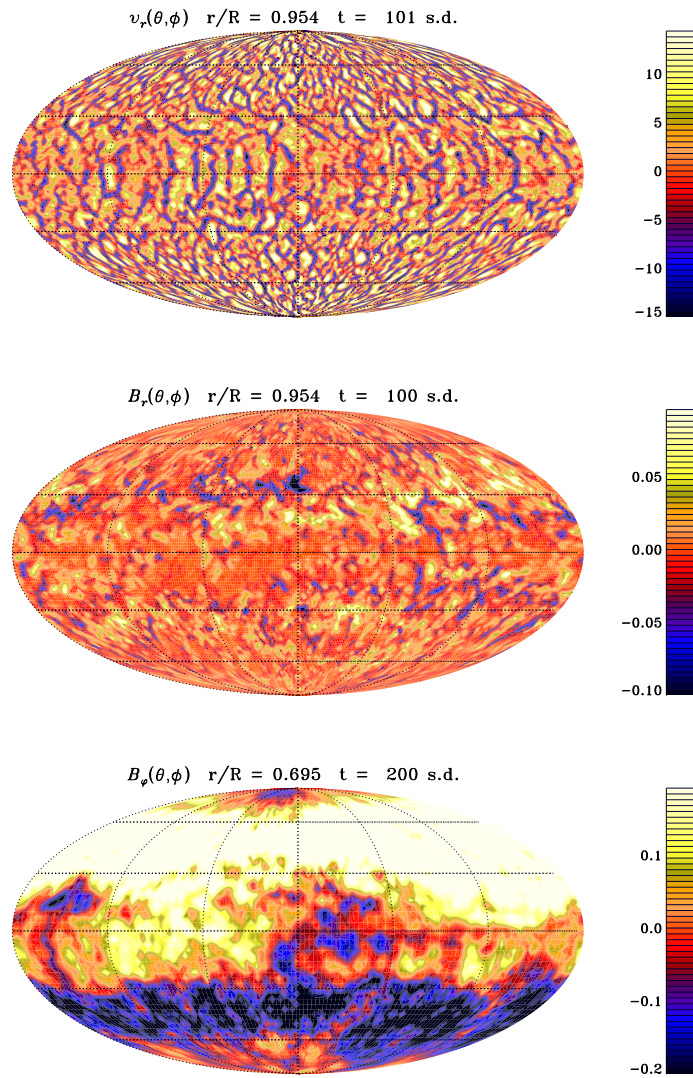


Fig. 49 Mollweide projections of the radial components of the flow velocity (top) and magnetic field (middle) on a spherical surface near the top of the simulation domain. Flow speeds are color-coded in m s^{-1} , and field strengths in Tesla. Note the asymmetry in the upflow/downflow structures, and the relatively small spatial scale of the subsurface magnetic field. The bottom Mollweide projection shows the the zonal magnetic component immediately beneath the base of the convective envelope, where a strong and well-organized axi-symmetric component builds up.

{fig:3DMHD2}

turbulent Reynolds stresses. This leads to a $\sim 20\%$ drop in the kinetic energy density, from ~ 270 to 230 J kg^{-1} at $t \simeq 150$ solar days.

This spatially well-organized magnetic component is particularly prominent at and beneath the base of the convective envelope, where the significant differential rotation, stably stratified environment, and injection of magnetic fields from above by downward turbulent pumping, all conspire to favor the buildup of significant magnetic flux. Indeed the presence of a tachocline-like stable layer underlying the convection zone proper is now believed to be an essential ingredient for the generation of a large-scale magnetic component. The bottom Mollweide projection on Figure 49 shows the zonal magnetic component at a depth slightly below the base of the convective envelope. Here the zonally-averaged toroidal field is seen to be well-organized on the larger scales, and in particular shows a clear antisymmetry with respect to the equatorial plane, in agreement with inferences made on the basis of Hale's polarity laws. Even though the stratification is convectively stable at this depth, convective undershoot from above introduces strong local fluctuations in the magnetic field, without however destroying its large-scale organization.

What is truly remarkable is that this large-scale toroidal field undergoes fairly regular solar-like polarity reversals on decadal timescales. This is shown on Figure 50, in the form of a time-latitude diagram of the zonally-averaged toroidal magnetic component at the core-envelope interface (part A), and a time-latitude diagram of the zonally-averaged surface radial magnetic component (part B). This simulation spans 255 years, in the course of which 8 polarity reversals have taken place, with a mean (half-) period of almost exactly 30 yr here. Examination of Fig. 50A reveals a tendency for equatorward migration in the course of each cycle. Figure 50B also shows the existence of a well defined dipole moment aligned with the rotational axis, reversing polarity approximately in phase with the deep-seated toroidal component. Note also how, despite significant fluctuations in the amplitude and timing of the cycle in each hemisphere, in general both hemispheres remain well-synchronized throughout the whole simulation. This is all extremely solar-like!

What kind of dynamo could this be? To answer this question we need to look in more detail at the flow fields. Figure 51 shows the zonally averaged kinetic helicity and angular velocity, plotted in a meridional $[r, \theta]$ plane. Typical for such simulations, the angular velocity profile is not a very good representation of the helioseismically-inferred profile in the real sun, although it does show some solar-like features, notably equatorial acceleration, deceleration at high latitudes, and disappearance of this latitudinal gradient within a thin radial shear layer coinciding approximately with the core-envelope interface. The tendency for alignment of Ω -isocontours with the rotation axis is a reflection of the Taylor-Proudman theorem, which states that in rotation-dominated systems (Coriolis term dominating over inertial and viscous terms on the RHS of eq. (77)), the flow velocity cannot vary in the direction parallel to the rotation axis. Here the pole-to-equator angular velocity contrast is about one third of what is observed on the Sun. However, an unmagnetized,

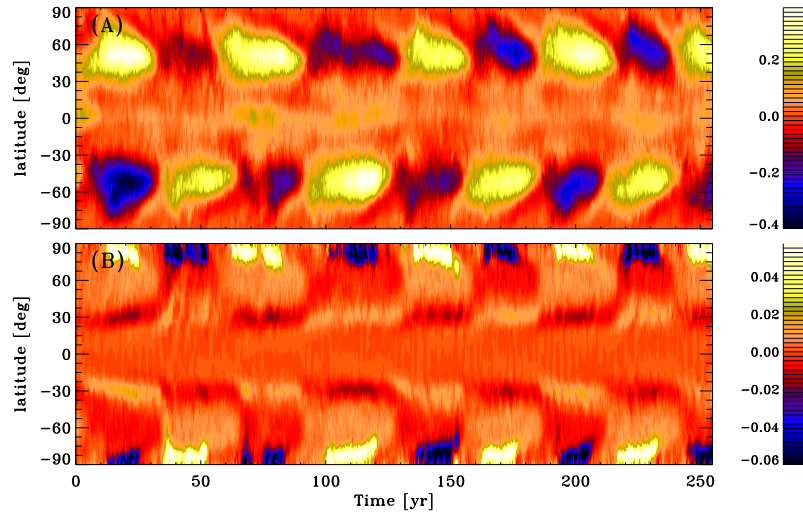


Fig. 50 (A) Time-latitude diagram of the zonally-averaged toroidal magnetic field at a depth corresponding to the top of the stable layer, immediately beneath the base of the convection zone. Note the antisymmetry of the large-scale field about the equatorial plane, the regular polarity reversals fairly synchronous across hemispheres, and the hint of equatorward migration of the toroidal field in the course of each cycle. Part B shows the corresponding time-latitude diagram for the surface radial component. The latter reveals a well-defined axisymmetric dipole moment, oscillating essentially in phase with the deep-seated toroidal component. The color scale codes the magnetic field strength, measured in tesla. Compare to Fig. 32.

{fig:3DMHD3}

purely hydrodynamical cousin simulation develops about the right contrast, which suggests that magnetically-mediated reduction of the large-scale flows is an important dynamo amplitude-limiting mechanism in this simulation, an inference supported by the fact that significant torsional oscillations are also present, varying on the same ~ 30 yr period as the large-scale magnetic field. The kinetic helicity is predominantly negative in the Northern solar hemisphere across the bulk of the convective envelope, as expected from cyclonic convection, with a sign change near its base produced with the rapid decrease of the turbulent intensity with depth.

The combination of a well-defined mean axisymmetric differential rotation and clear-cut hemispheric pattern of kinetic helicity suggests that this simulation may be operating as the $\alpha\Omega$ dynamos considered earlier (§0.26). The production of a positive dipole moment from a positive toroidal field in the Northern hemisphere is what one would associate in mean-field theory with a positive α -effect (more precisely, a positive $\alpha_{\phi\phi}$ tensor component), which, if eq. (213) were to hold, is what would result from a negative kinetic

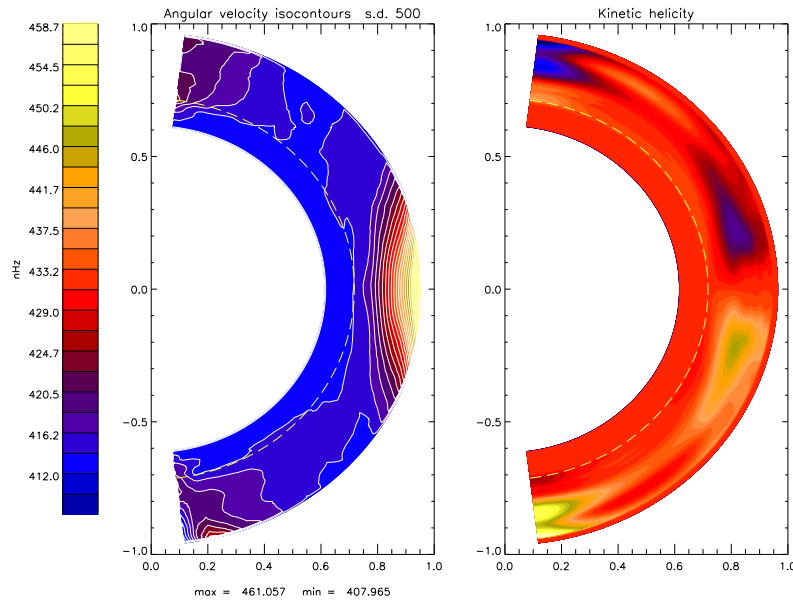


Fig. 51 Zonal and 10-s.d. temporal averages of the angular velocity (left), and kinetic helicity (right), both plotted in meridional $[r, \theta]$ planes at simulation time $t = 385$ s.d. The rotation axis is oriented vertically. Note the rapid disappearance of differential rotation as one crosses below the core-envelope interface, and the hemispheric anti-symmetry of the kinetic helicity pattern.

{fig:3DMHD4}

helicity in the N-hemisphere, as on Figure 51B. Explicit calculations of the ϕ -component of the mean (zonally-averaged) electromotive force $\langle \mathbf{u} \times \mathbf{b} \rangle$ indicates a well-defined spatial pattern, of like sign in both hemispheres but changing sign from one cycle to the next, again exactly what one would expect from an $\alpha\Omega$ turbulent dynamo. On the other hand, the poloidal component of the electromotive force has a magnitude comparable to the shearing term arising from differential rotation, which would then suggest the $\alpha^2\Omega$ mode of dynamo action.

Although such simulations are just beginning to yield solar-like regular cyclic global magnetic polarity reversals, they remain extremely demanding computationally, and are still a long way from producing anything resembling a toroidal flux rope, let alone a sunspot. This is why the much simpler cycle models described in this chapter remain at this writing the favored modelling framework within which to investigate the observed characteristics of solar and stellar cycles, and in particular the origin of fluctuations in their amplitude and duration on long timescales. This is the topic to which we now turn.

Bibliography:

The observational literature on the solar magnetic field is immense, and many good review papers are available. One I particularly like is:

Solanki, S., Inhester, B., and Schüssler, M., *Rep. Prog. Phys.*, **69**, 563–668 (2006).

Hale's original papers on sunspots are still well worth reading. The two key papers are:

Hale, G.E., *The Astrophysical Journal*, **28**, 315-343 (2908),

Hale, G.E., Ellerman, F., Nicholson, S.B., and Joy, A.H., *The Astrophysical Journal*, **49**, 153-178 (1919).

The study of rising toroidal flux ropes a proxy for the emergence of the solar internal toroidal field in the form of sunspot pairs is a topic that has generated a voluminous literature. Among the many noteworthy contributions in this field , we recommend the following as starting points:

Moreno-Insertis, F., *Astron. Astrophys.*, **166**, 291-305 (1986),

Choudhuri, A.R., & Gilman, P.A., *Astrophys. J.*, **316**, 788-800 (1987),

Fan, Y., Fisher, G.W., & DeLuca, E.E., *Astrophys. J.*, **405**, 390-401 (1993),

D'Silva, S., & Choudhuri, A.R., *Astron. Astrophys.*, **272**, 621-633 (1993),

Caligari, P., Moreno-Insertis, F., & Schüssler, M., *Astrophys. J.*, **441**, 886-902 (1995).

The thin flux tube approximation used in most of these calculations is due to

Spruit, H.C., *Astron. Astrophys.*, **98**, 155-160 (1981).

On the storage and stability of toroidal flux ropes below the solar convective envelope, see

Ferriz-Mas, A., & Schüssler, M., *Astrophys. J.*, **433**, 852-866 (1994),

Ferriz-Mas, A., *Astrophys. J.*, **458**, 802-816 (1996).

Considerable effort is currently being put into doing away with the thin flux tube approximation, in order to see which of the above results remains robust once the flux tube is no longer treated as a one-dimensional object. This is a rapidly moving field, so for the latest see the following recent on-line review:

Fan, Y., *Liv. Rev. Sol. Phys.*, **6** (2009),

<http://solarphysics.livingreviews.org/Articles/lrsp-2009-4/>

The following three recent review papers jointly offer a good overview of dynamo models of the solar cycle:

Charbonneau, P., *Liv. Rev. Sol. Phys.*, **7** (2010),

<http://solarphysics.livingreviews.org/Articles/lrsp-2010-3/>

- Ossendrijver, M. A. J. H., *Astron. Astrophys. Rev.*, **11**, 287-367 (2003).
 Hoyng, P., “The field, the mean and the meaning”, in *Advances in nonlinear dynamos*, The Fluid Mechanics of Astrophysics and Geophysics, **8**, eds. A. Ferriz-Mas & M. M. Jiménez, 1–36 (2003).

Mean-field electrodynamics grew out of the original pioneering efforts of,

- Parker, E.N., *Astrophys. J.*, **122**, 293-314 (1955).
 Braginskii, S.I., *Sov. Phys. JETP*, **20**, 726; 1462 (1964),
 Steenbeck, M., & Krause, F., *Astr. Nach.*, **291**, 49-84 (1969),

but the following three monographs are a better starting point for those wishing to dig deeper into the subject:

- Moffatt, H.K. 1978, *Magnetic Field Generation in Electrically Conducting Fluids*, (Cambridge: Cambridge Univ. Press),
 Parker, E.N. 1979, *Cosmical Magnetic Fields*, (Oxford: Clarendon Press),
 ch. 18,
 Krause, F., & Rädler, K.-H. 1980, *Mean-Field Magnetohydrodynamics and Dynamo Theory*, (Oxford: Pergamon Press).

On empirical estimates of the α -effect from numerical simulations of MHD turbulence, start with:

- Pouquet, A., Frisch, U., and Léorat, J., *J. Fluid Mech.*, **77**, 321-354 (1976),
 Ossendrijver, M., Stix, M., and Brandenburg, A., *Astron. Astrophys.*, **376**,
 731-726 (2001),
 Käpylä, P.J., Korpi, M.J., Ossendrijver, M., and Stix, M., *Astron. Astrophys.*, **455**, 401-412 (2006),
 Hubbard, A., Del Sordo, F., Käpylä, P.J., and Brandenburg, A., *Month. Not. R. Astron. Soc.*, **398**, 1891-1899.

The technical literature on dynamo models of the solar cycle is truly immense. There are many hundreds of noteworthy papers out there! Those included below are just meant to be good entry points for those wishing to pursue in greater depth topics covered in this chapter. For a good overview of mean-field solar cycle models and their evolution in time, see

- Lerche, I., & Parker, E.N., *Astrophys. J.*, **176**, 213-223 (1972),
 Yoshimura, H., *Astrophys. J.*, **201**, 740-748 (1975),
 Ivanova, T.S., & Ruzmaikin, A.A., *Sov. Astron.*, **20**, 227 (1976),
 Stix, M., *Astron. Astrophys.*, **47**, 243-254 (1976),
 Rüdiger, G., and Brandenburg, A., *Astron. Astrophys.*, **296**, 557-566
 (1995),
 Moss, D., and Brooke, J. M., *Month. Not. R. Astron. Soc.*, **315**, 521-533
 (2000).

On the impact of meridional circulation on dynamo waves, see

- Bullard, E.C., *Vistas Astron.*, **1**, 685-691 (1955),

- Choudhuri, A.R., Schüssler, M., & Dikpati, M., *Astron. Astrophys.*, **309**, L29-L32 (1995),
 Küker, M., Rüdiger, G., and Schulz, M., *Astron. Astrophys.*, **374**, 301-308 (2001),
 Roberts, P.H., & Stix, M., *Astron. Astrophys.*, **18**, 453-466 (1972).

The meridional circulation profile described in §0.26.11 is the creation of

- van Ballegooijen, A.A., & Choudhuri, A.R., *Astrophys. J.*, **333**, 965-977 (1988).

On α -quenching, standard versus catastrophic and related dynamical issues:

- Blackman, E.G., and Field, G.B., *Astrophys. J.*, **534**, 984-988 (2000),
 Cattaneo, F., and Hughes, D., *Phys. Rev. E*, **54**, R4532-R4535 (1996),
 Durney, B.R., De Young, D.S., and Roxburgh, I.W., *Solar Phys.*, **145**, 207-225 (1993),
 Rüdiger, G., and Kitchatinov, L.L., *Astron. Astrophys.*, **269**, 581-588 (1993).
 Cattaneo, F., and Hughes, D.W., *Month. Not. R. Astron. Soc.*, **395**, L48-L51 (2009).

On interface dynamos, see

- Charbonneau, P., and MacGregor, K.B., *Astrophys. J.*, **473**, L59-L62 (1996),
 MacGregor, K.B., and Charbonneau, P., *Astrophys. J.*, **486**, 484-501 (1997),
 Parker, E.N., *Astrophys. J.*, **408**, 707-719 (1993),
 Petrovay, K., & Kerekes, A., *Month. Not. R. Astron. Soc.*, **351**, L59-L62 (2004),
 Tobias, S.M., *Astrophys. J.*, **467**, 870 (1996).

and on the energetics of thin layer dynamos:

- Steiner, O., & Ferriz-Mas, A., *Astr. Nach.*, **326**, 190-193 (2005).

What is now referred to as Babcock-Leighton solar-cycle models go back to the following three seminal papers by H. Babcock and R. Leighton:

- Babcock, H.W., *Astrophys. J.*, **133**, 572i-587 (1961),
 Leighton, R.B., *Astrophys. J.*, **140**, 1547-1562 (1964),
 Leighton, R.B., *Astrophys. J.*, **156**, 1-26 (1969).

Although some details of the model are different, the 2D surface simulations described in §0.27.1 basically follow

- Wang, Y.-M., Nash, A.G., & Sheeley, N.R. Jr, *Science*, **245**, 712-718 (1989),
 Wang, Y.-M., & Sheeley, N.R. Jr, *Astrophys. J.*, **375**, 761-770 (1991),

but on this general topic of surface magnetic flux evolution, see also:

- Schrijver, C.J., Title, A.M., van Ballegooijen, A.A., Haagenar, H.J., & Shine, R.A., *Astrophys. J.*, **487**, 424-436 (1997),
 Schrijver, C.J., *Astrophys. J.*, **547**, 475-490 (2001),
 Schrijver, C.J. & Title, A.M., *Astrophys. J.*, **551**, 1099-1106 (2001),
 Schrijver, C.J., DeRosa, M.L., & Title, A.M., *Astrophys. J.*, **577**, 1006-1012 (2002),
 Baumann, I., Schmitt, D., Schüssler, N., and Solanki, S., *Astron. Astrophys.*, **426**, 1075-1091 (2004).

The formulation of the Babcock-Leighton solar cycle model of §0.27 is identical to

- Charbonneau, P., St-Jean, C., & Zacharias, P., *Astrophys. J.*, **619**, 613-622 (2005).

For different modeling approaches, see

- Wang, Y.-M., & Sheeley, N.R. Jr, & Nash, A.G., *Astrophys. J.*, **383**, 431-442 (1991),
 Durney, B.R. 1995, *Solar Phys.*, **160**, 213-235 (1995),
 Dikpati, M., & Charbonneau, P., *Astrophys. J.*, **518**, 508-520 (1999),
 Nandy, D., and Choudhuri, A. R., *Astrophys. J.*, **551**, 576-585 (2001),
 Guerrero, G., & de Gouveia Dal Pino, E.M., *Astron. Astrophys.*, **485**, 267-273 (2008),
 Muñoz-Jaramillo, A., Nandy, D., Martens, P.C.H., and Yeates, A.R., *Astrophys. J. Lett.*, **720**, L20-L25 (2010),

On the “tachocline α -effect” dynamo model described in §0.28.1, and associated stability analyses, begin with:

- Dikpati, M., and Gilman, P. A., *Astrophys. J.*, **559**, 428-442 (2001),
 Dikpati, M., Gilman, P. A., and Rempel, M., *Astrophys. J.*, **596**, 680-697 (2003),
 Gilman, P. A., and Fox, P. A., *Astrophys. J.*, **484**, 439-454 (1997).

And for the “flux tube α -effect” dynamo model of §0.28.2, and associated stability analyses, try first:

- Ferriz-Mas, A., Schmitt, D., and Schüssler, M., *Astron. Astrophys.*, **289**, 949-956 (1994),
 Ossendrijver, M. A. J. H., *Astron. Astrophys.*, **359**, 1205-1210 (2000).

On the numerical simulations of global 3D MHD convection in thick, rotating stratified spherical shells, begin with

- Brun, A.S., Miesch, M.S., & Toomre, J., *Astrophys. J.*, **614**, 1073-1098 (2004),
 Browning, M.K., Miesch, M.S., Brun, A.S., & Toomre, J., *Astrophys. J. Lett.*, **648**, 157-160 (2006),
 Brown, B.P., Browning, M.K., Brun, A.S., Miesch, M.S., & Toomre, J., *Astrophys. J.*, **711**, 424-438 (2010),

as well as the following two recent review articles:

Miesch, M.S., *Liv. Rev. Sol. Phys.*, **2** (2005),

<http://solarphysics.livingreviews.org/Articles/lrsp-2005-1/>

Miesch, M.S., & Toomre, J., *Ann. Rev. Fluid Mech.*, **41**, 317-345 (2009).

See also the fascinating results presented in

Cline, K.S., Brummell, N.H., & Cattaneo, F., *Astrophys. J.*, **599**, 1449-1468 (2003),

Käpylä, P.J., Korpi, M.J., Brandenburg, A., Mitra, D., Tavakol, R., 2010, *Astron. Nach.*, **331**, 73–ff.

The production of solar-like magnetic cycles in such simulations is a very recent breakthrough. The simulation results presented in §0.29 are taken from

Ghizaru, M., Charbonneau, P., & Smolarkiewicz, P.K., *Astrophys. J. Lett.*, **715**, L133-137 (2010).

An analysis of the mode of dynamo action therein is presented in

Racine, É., Charbonneau, P., Ghizaru, M., Bouchat, A., & Smolarkiewicz, P.K., *Astrophys. J.*, submitted (2010).

These simulations were computed with the MHD version, developed at the Université de Montréal, of the general purpose hydrodynamical simulation code EULAG; on the latter, see

Prusa, J.M., Smolarkiewicz, P., & Wyszogrodzki, A.A., *Comp. Fluids*, **37**, 1193-1207 (2008),

as well as the EULAG Web Page:

<http://www.mmm.ucar.edu/eulag/>.

Chapter 4

Fluctuations, intermittency and predictivity

{pch_chap:fluct}

*It is nice to know that the computer understands the problem,
but I would like to understand it too.*

Attributed to E.P. Wigner

Given that the basic physical mechanism(s) underlying the operation of the solar cycle are not yet agreed upon, attempting to understand the origin of the observed *fluctuations* of the solar cycle may appear to be a futile undertaking. Nonetheless, work along these lines continues at full steam in part because of the high stakes involved: the solar radiative output and frequencies of all eruptive phenomena relevant to space weather are strongly modulated by the amplitude of the solar cycle; varying levels of solar activity may contribute significantly to climate change; and certain aspects of the observed fluctuations may actually hold important clues as to the physical nature of the dynamo process.

We first briefly review some classical solar cycle fluctuation patterns, as inferred from the sunspot number time series (§0.30). With an eye on reproducing these patterns, we then study the response of some of the basic dynamo models considered in the preceding chapter to stochastic forcing (§0.31), dynamical nonlinearities (§0.32), and time delays (§0.33). We then examine how the interaction of some of these modulation mechanisms can lead to intermittency (§0.34), and close with a brief survey of the current status of model-based solar cycle prediction schemes (§0.35).

0.30 Observed patterns of solar cycle variations

{pch_sec:obsfluct}

0.30.1 Pre-telescopic and early telescopic sunspot observations

{pch_ssec:early}

Until the beginning of the twentieth century, the story of the solar activity cycle is coincident with the story of *sunspots*. Appearing as dark blemishes on the bright solar disk, only the largest sunspots can be visible to the naked-eye under suitable viewing conditions, for example when the sun is partially obscured by clouds or mist, particularly at sunrise or sunset. Numerous such sightings exist in the historical records, starting with Theophrastus (374-287 B.C.) in the fourth century B.C. However, the most extensive pre-telescopic records are found in the far east, especially in the official records of the Chinese imperial courts, starting in 165 B.C. In december 1128 the monks authoring the Worcester Chronicles also left us with the first known sunspot drawing of what must have been two exceptionally large sunspots, since their umbrae and penumbrae could be visually distinguished. A fascinating pre-telescopic sunspot sighting is certainly that of 28 May 1607 by none other than Johannes Kepler (1571-1630). Kepler had been observing the sun for over a month using his *camera obscura* projection technique, basically a pinhole camera. He was hoping to detect a transit of Mercury across the solar disk, as predicted by extant planetary ephemerides. On May 28, through a short-lived break in cloud cover, he did notice a small black spot on the solar disk, and concluded that he was indeed seeing Mercury in transit. It did not take long before he came to realize his mistake.

At the end of the first decade of the seventeenth century, the “discovery” of sunspots was definitely in the air. Within less than a year of one another, four observers turned the newly invented astronomical telescope toward the Sun, and independently noted the existence of sunspots. They were Johann Goldsmid (1587-1616, a.k.a. Fabricius) in Holland, Thomas Harriot (1560-1621) in England, Galileo Galilei (1564-1642) in Italy, and the Jesuit Christoph Scheiner (1575-1650) in Germany. Fabricius was the first to publish his results in 1611, and to correctly interpret the apparent motion of sunspots in terms of axial rotation of the Sun. Like Harriot, Fabricius and his father (the then-well-known astronomer David Fabricius) first observed sunspots directly through their telescope shortly after sunrise or before sunset. The harrowing account of their observations is worth quoting: (excerpt from the translation in the paper by W.M. Mitchell cited at the end of this chapter):

“... Having adjusted the telescope, we allowed the sun’s rays to enter it, at first from the edge only, gradually approaching the center, until our eyes were accustomed to the force of the rays and we could observe the whole body of the sun. We then saw more distinctly and surely the things I have described [sunspots]. Meanwhile clouds interfered, and also the sun hastening to the meridian destroyed our hopes of longer observations; for indeed it was to be feared that an indiscreet examination of a lower sun would cause great injury to the eyes, for even the weaker rays of the setting or rising sun often inflame the eye with a

strange redness, which may last for two days, not without affecting the appearance of objects.”

Galileo and Scheiner, however, were the most active in using sunspots to attempt to infer physical properties of the Sun. To Galileo belongs the credit of making a convincing case that sunspots are indeed features of the solar surface, as opposed to intra-Mercurial planets, as advocated initially by Scheiner. But the Jesuit astronomer did detect and measure, after years of careful observations of sunspots’ apparent motion on the solar disk, the slight inclination of the Sun’s equatorial plane with respect to the ecliptic. But as for the physical nature of sunspots, even the universally opiated Galileo remained unusually cautious, tentatively granting them the status of clouds in the solar atmosphere.

In the three subsequent centuries sunspots became dark mountains peaking through luminous cloud layers, then orifices allowing a view of the dark solar surface, then giant cyclones until, as we already discussed in §0.25, their magnetic nature was firmly established by Hale and collaborators. But this did not prevent sunspot observers to make some striking discoveries in the three centuries separating Hale from Galileo.

0.30.2 The sunspot cycle

{pch_ssec:schwabe}

Early sunspot observers noted the curious fact that sunspots rarely appear outside of a latitudinal band of about $\pm 30^\circ$ centered about the solar equator, but otherwise failed to discover any clear pattern in the appearance and disappearance of sunspots. This fell to the German amateur astronomer Samuel Heinrich Schwabe (1789-1875), who in 1843, after a 17-years telescopic hunt for intra-mercurial planets, announced the existence of a decadal periodicity in the average number of sunspots visible on the Sun. Much impressed by Schwabe’s discovery, the Swiss astronomer Rudolf Wolf (1816-1893) launched in a life-long quest for sunspot data and drawings from previous centuries, with the aim of tracking the sunspot cycle all the way back to the beginning of the telescopic era. Faced with the daunting task of comparing sunspot observations carried out by many different astronomers using various instruments and observing techniques, Wolf defined a *relative sunspot number* (r) as follows:

$$r = k(f + 10g) , \tag{265} \quad \{\text{eq:wolfssn}\}$$

where g is the number of sunspot groups visible on the solar disk, f is the number of individual sunspots (including those distinguishable within groups), and k is a correction factor that varies from one observer to another (with $k = 1$ for Wolf’s own observations). This definition is still in used today, but r is now usually called the Wolf (or Zürich) sunspot number. Wolf succeeded in reliably reconstructing the variations in sunspot number as far as the the

1755–1766 cycle, which has since been known conventionally as “Cycle 1”, with all subsequent cycles numbered consecutively thereafter; at this writing (December 2010), we are just coming out of the unusually extended minimal activity phase delineating cycle 23 from the upcoming cycle 24.

Figure 52 shows two time series of the relative sunspot number. The first (thin black line) is the monthly-averaged value of r as a function of time, and the thick red line is a 13-month running mean of the same. The amplitude, duration and even shape of sunspots cycles can vary substantially from one cycle to the next. The period, in particular, ranges from 9 (cycle 2) to 14 years (cycle 4). Moreover, we have already seen (§0.25) that from a physical —

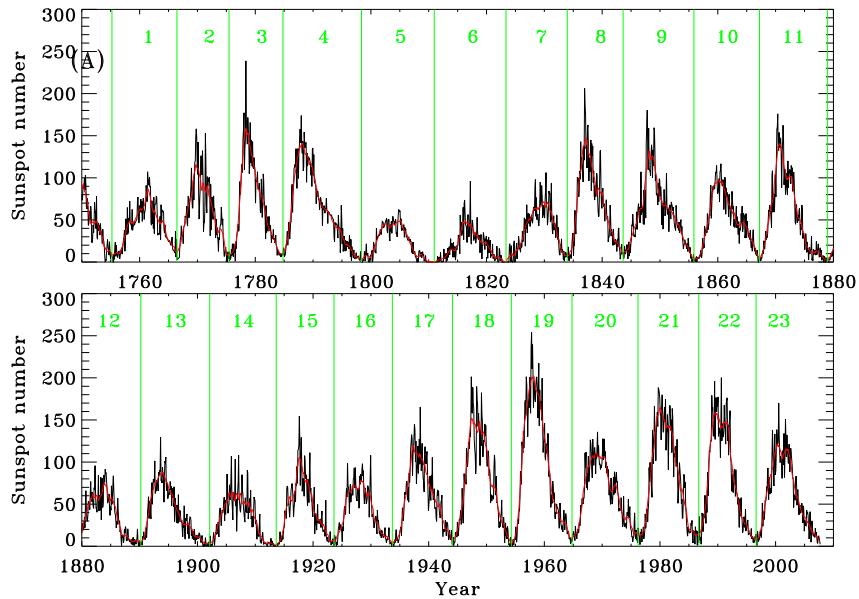


Fig. 52 fig:ssn Two time series of the celebrated Wolf Sunspot Number. The thin black line is the monthly-averaged sunspot number, and the thick red line a 13-month running mean thereof. On the basis of the latter, one can calculate a mean cycle peak amplitude of $SSN=115$ with standard deviation ± 40 , and a mean period 10.8 yr with standard deviation ± 1.6 yr. Cycles are defined from one minimum to the next, and are numbered following Wolf’s convention. These and other related data are publicly available at the Solar Influences Data Analysis Center in Brussels, Belgium (<http://sidc.oma.be>).

rather than botanical— standpoint, the full period of the underlying magnetic cycle is twice that of the sunspot cycle. Yet, because the near totality of phenomena defining solar activity are unaffected by the magnetic polarity of the sun’s large-scale magnetic field, and also perhaps because astronomers

are creatures of tradition, even a century after Hale’s discovery of the sunspot polarity law it remains customary to speak of the “11 year solar cycle”.

0.30.3 The butterfly diagram

{ssec:butterfly}

To the striking cyclic pattern uncovered by Schwabe was soon added an equally striking *spatial* regularity. In 1858, Gustav Spörer (1822-1895) and Richard Carrington (1826-1875) independently pointed out that sunspots are observed at relatively high ($\sim 40^\circ$) heliocentric latitudes at the beginning of a sunspot cycle, but are seen at lower and lower latitudes as the cycle proceeds, until at the end of the cycle they are observed mostly near the equator, at which time spots announcing the onset of the next cycle begin to appear again at $\sim 40^\circ$ latitude. This is illustrated on Figure 53, in the form of a *butterfly diagram* for the time period 1875–2010. The construction of a sunspot butterfly diagram, first carried out by the husband-and-wife team of Annie and E. Walter Maunder in 1904, proceeds as follows: one begins by laying a coordinate grid on, for example, a solar white light or calcium image, with, as in the case of geographic coordinates on Earth, the rotation axis defining the North—South direction. The visible solar disk is then divided in latitudinal strips of constant projected area, and for each such strip the percentage of the area covered by sunspots and/or active regions is calculated and color coded. This defines a one-dimensional (vertical) array describing the average sunspot coverage at one time. By repeating this procedure at constant time intervals and stacking the arrays one besides the other, one obtains a two-dimensional image of average sunspot coverage as a function of heliospheric latitude (vertical axis) and time (horizontal axis).

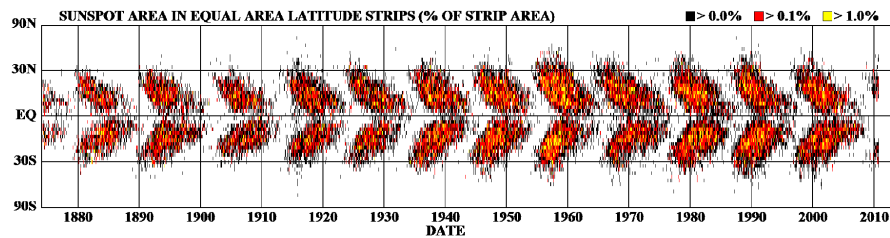


Fig. 53 $F_{10.7}$ sunspot butterfly diagram, showing the equatorward migration of sunspot latitudes in the course of each cycle. The sunspot number peaks about midway though the equatorward migration. Public domain data and graphics courtesy of David Hathaway, NASA/MSFC. [<http://solarscience.msfc.nasa.gov/images/bfly.gif>]

The absence of sunspots at high latitudes ($\gtrsim 40^\circ$) at any time during the cycle, and the equatorward drift of the sunspot distribution as the cycle proceeds from maximum to minimum, are both particularly striking on such

a diagram. Note how the latitudinal distribution of sunspots is never exactly the same, and how for certain cycles there exists a significant North–South asymmetry in the hemispheric distributions. Note also how, at solar minima, spots from each new cycle begin to appear at mid-latitudes while spots from the preceding cycle can still be seen near the equator, and how sunspots are almost never observed within a few degrees in latitude of the equator. Sunspot maximum (1991, 1980, 1969,...) occurs about midway along each butterfly, when sunspot coverage is maximal at about 15 degrees latitude. Note also how fairly good overall hemispheric synchrony is maintained, despite significant amplitude variations and occasional large lag in the beginning of the cycle in one hemisphere with respect to other, for example at the onset of the 1965-1976 cycle.

{pch_ssec:wald}

0.30.4 The Waldmeier and Gnevyshev-Ohl Rules

The sunspot numbers is our longest direct record of solar activity, and thus remains a favored dataset for the analysis and modelling of solar cycle fluctuations. Starting with Wolf himself, the sunspot number time series (monthly, monthly smoothed, yearly, etc) have been analyzed in every possible manner known to statistics, nonlinear dynamics, and numerology³². Many otherwise serious and respectable people engaged in this type of work seem to forget that the definition of the sunspot number is largely arbitrary, and its quantitative relationship to the real dynamical quantity, the sun’s magnetic internal field, uncertain at best.

Nonetheless, some of the various patterns so uncovered do appear to be robust, in that they do not depend too much on the manner the analysis is being carried out, and are also found in other indicators of solar activity; for example the so-called *Gleissberg cycle* refers to a ~ 80 yr modulation of the overall envelope of cycle amplitudes. Some sunspot number patterns have even proven resilient enough to be upgraded to the status of empirical “Rules”, two of the more convincing ones being the so-called Waldmeier Rule, and Gnevyshev-Ohl Rule.

The *Waldmeier Rule* refers to an anticorrelation observed between cycle amplitude and rise time (or duration). Starting for example from the time series of smoothed monthly sunspot number (red line on Figure 52), it is straightforward to assign to each cycle n a peak amplitude A_n and a duration T_n , the latter being simply the time interval between the two minima bracketing a given cycle. Similarly, the rise time is the time interval between a minimum and the subsequent maximum. Figure 54A shows a correlation plot of cycle rise time and amplitude, which is characterized by a linear cor-

³² Two colleagues, David J. Thomson and Werner Mende, both world-renowned experts in time series analysis, have independently remarked to me that the sunspot number time series are quite possibly the “natural” time series having produced the largest number of research journal pages per byte of actual data!

relation coefficient of $r = -0.7$, definitely large enough to merit attention. A similar, though weaker anticorrelation exists between cycle amplitude and duration, a consequence of a similarly weak correlation ($r = +0.4$) existing between cycle rise time and duration (see Fig. 54B). The latter correlation is taken to be indicative of some level of self-similarity in the temporal unfolding of sunspot cycles. The amplitude-duration anticorrelations are intriguing, because one might have (naively) expected that high amplitude cycles should take longer to build up and also last longer, but in fact the opposite seems to hold.

The *Gnevyshev-Ohl Rule* is another intriguing pattern, and is illustrated on Fig. 54C. Cycle peak amplitude A_n are plotted as solid dots, versus cycle number n , following Wolf's numbering convention. Compute now a 1-2-1 running mean of cycle amplitude, i.e.,

$$\langle A_n \rangle = \frac{1}{4}(A_{n-1} + 2A_n + A_{n+1}), \quad n = 2, 3, \dots \quad (266) \quad \{\text{eq:121}\}$$

The resulting time series for $\langle A_n \rangle$ is plotted as a thick blue line on Fig. 54C; notice now how most odd-numbered cycles (orange) lie *above* the running mean curve, while even-numbered cycles (red) usually lie below. In fact, from cycle 9 to 21 inclusive, the pattern has held true without interruption. Remembering that the physical underlying magnetic cycle has a period of twice that of the sunspot cycle, the Gnevyshev-Ohl Rule could then indicate a preference for one magnetic polarity over the other. Such a preference could result from the interaction of a cyclic, dynamo-generated large-scale magnetic field with a steady, large-scale magnetic field contained within the sun's radiative interior, presumably of fossil origin. We shall see shortly that there exist other purely dynamo-based explanations for this pattern.

0.30.5 The magnetic activity cycle

{ssec:activcycle}

Because a fraction the solar magnetic field extends into the corona, and because it is dynamically significant there, the equilibrium structure of the corona ends up being defined by a balance between three primary forces: gravity, plasma pressure, and the Lorentz force. As the photospheric magnetic field inexorably evolves as a result of advection by flows and magnetic flux emergence, this equilibrium is eventually lost, leading to rapid and often spectacular disruptions of coronal structures. The associated phenomena are grouped under the general name of *solar activity*, and include phenomena as diverse as flares and coronal mass ejections. The sun's magnetic field is in fact the primary energy source for the majority of such coronal transients.

As discussed in §0.25, sunspots are effectively tracers of the deep-seated solar magnetic field. One could then expect a strong correlation between measures of solar activity and sunspot numbers. This expectation is indeed borne out, as shown on Figure 55. The sunspot cycle is found to modulate the sun's

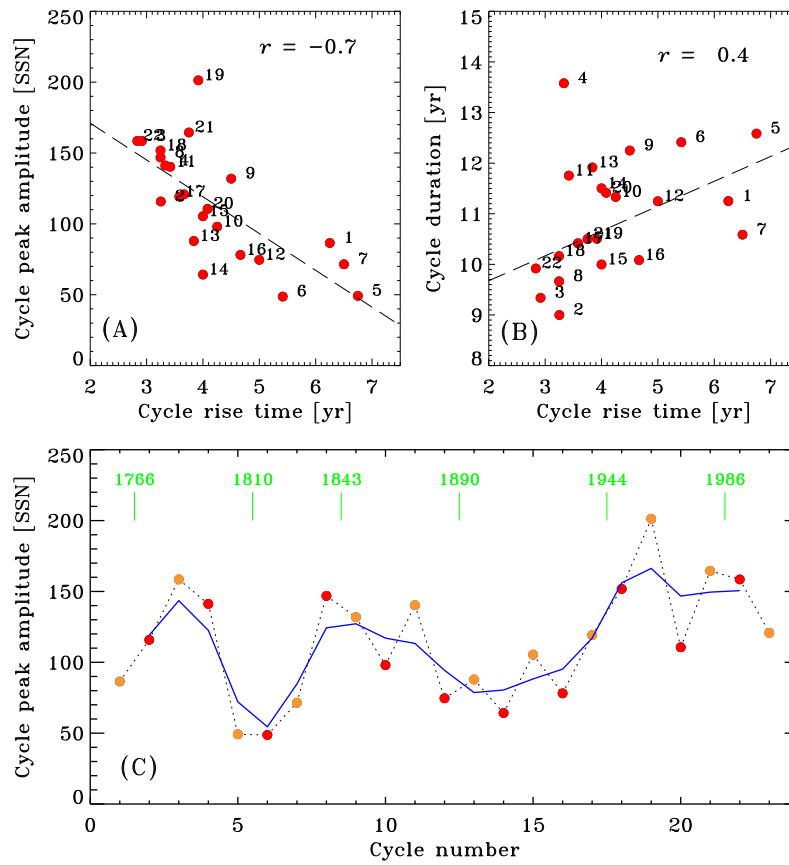


Fig. 54 ^{fig:wald} A sample of solar cycle fluctuation patterns. (A) The anticorrelation between cycle rise time and amplitude, known as the Waldmeier Rule. A similar anticorrelation, although weaker, characterizes cycle amplitudes and durations; this is because a correspondingly weak correlation is found between cycle rise time and duration, as shown in (B). The Gnevyshev-Ohl Rule is illustrated in (C). Under Wolf's numbering convention, the odd-numbered cycles (orange dots) are more often found above the running mean (blue line) than even-numbered cycles (red dots), a pattern that held true uninterrupted from cycle 9 to 21 inclusively.

radiative emission from the radio to the X-ray domain of the electromagnetic spectrum, as well as the frequency of geoeffective eruptive events such as flares and coronal mass ejections. This is (in part) why understanding the origin of fluctuations in the amplitude of the solar cycle remains such an active area of research.

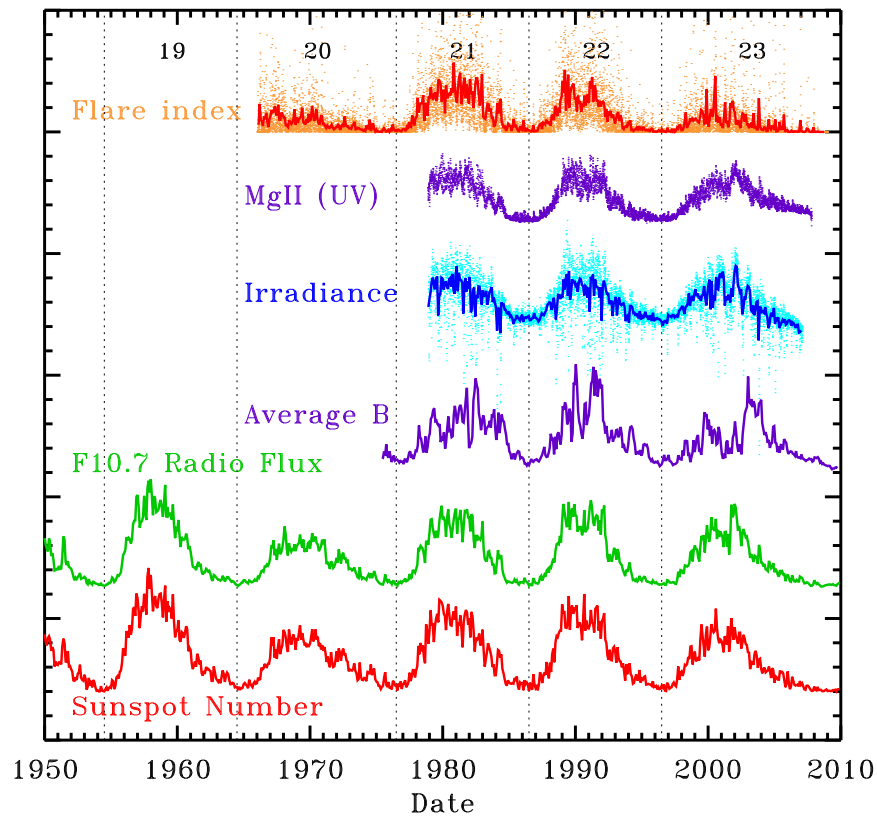


Fig. 55 fig:solaractivity The solar activity cycle, as measured via various proxies. from bottom to top: the sunspot number, the F10.7 radio flux, the disk-averaged line-of-sight magnetic field, the total solar irradiance, the MgII index (a good proxy of ultraviolet emission), and the solar flare index. Plot constructed from public-domain data archived at NOAA (USA), available at <http://www.ngdc.noaa.gov>.

{pch_ssec:maunder}

0.30.6 The Maunder Minimum

One final, peculiar feature associated with the sunspot cycle needs to be discussed, because of its implications for dynamo modelling. The historical reconstructions begun by Wolf have been pushed as far back as the invention of the telescope in the opening decade of the seventeenth century, which marks the beginning of regular sunspot monitoring by astronomers. One such full reconstruction, starting in 1610, is shown on Figure 56 (bottom panel). While observations are a tad patchy from 1610 to 1640, coverage is actually quite good beyond this date. The lack of sunspots in the period 1645-1715 is there-

fore not due to lack of data, but represents a phase of strongly suppressed solar activity now known as the *Maunder Minimum*. The documented occurrence of exceptionally cold winters throughout Europe during those years may be causally related to reduced solar activity, although this remains a topic of controversy.

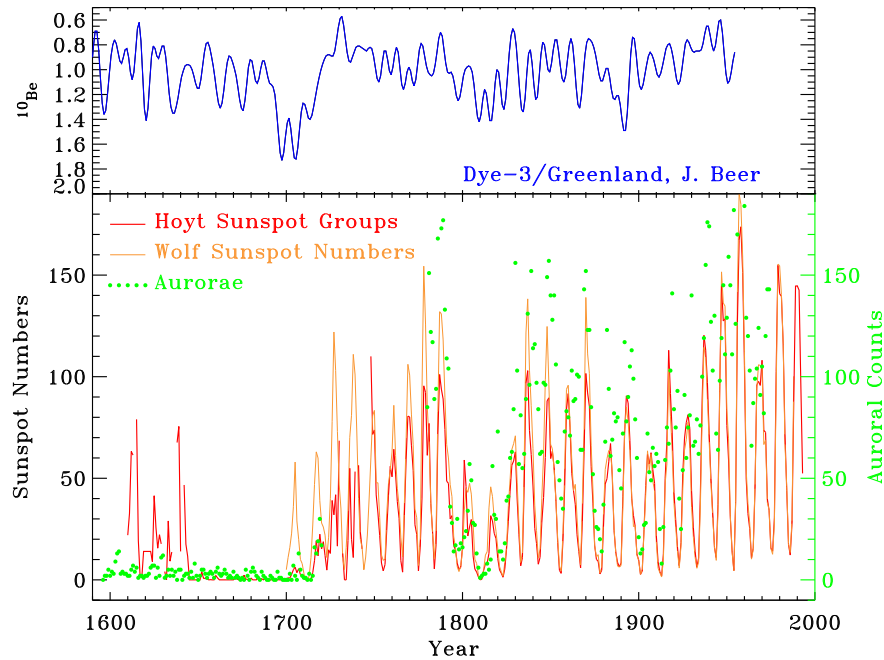


Fig. 56 ^{F1.16} The Maunder minimum, as seen through cosmogenic radioisotopes (top panel) and sunspot and auroral counts (bottom panel). The thick red line is the so-called Group Sunspot Number, a reconstruction similar to Wolf's (thin orange line) but deemed more reliable in the eighteenth century because it relies exclusively on the more easily observable sunspot groups. Beryllium 10 data courtesy of J. Beer, EAWAG/Zürich.

That this is not just a matter of failing to form sunspots is confirmed by historical reconstructions of auroral counts, which are also strongly reduced during the Maunder Minimum (cf. Fig. 56). On the other hand, cosmogenic radioisotopes such as ^{10}Be , whose production frequency is known to be modulated by the frequency of solar eruptive phenomena and general strength of the interplanetary magnetic field, continue to show a cyclic pattern throughout the Maunder minimum (Fig. 56, top panel), indicating that the cycle had actually not come to a complete standstill. The Maunder Minimum remains a real puzzle in many ways.

The cosmogenic isotope record also indicates that similar episodes of markedly reduced solar activity occurred in 1282-1342 (Wolf minimum) and

1416-1534 (Spörer minimum), and that solar activity was significantly above its modern average in the time period 1100-1250 (dubbed Medieval Maximum by Min/Max aficionados). Some recent such reconstructions (see bibliography) have in fact identified 27 grand minima in the past 11,000 years. No convincing periodicity or other temporal pattern has yet been identified in the occurrence of these grand minima. The 1798—1823 *Dalton Minimum*, spanning the low and oddly-shaped cycles 5 and 6, is sometimes categorized as a “failed Grand Minima”, although supporting arguments tend to be of a botanical flavor.

0.30.7 From large-scale magnetic fields to sunspot number

{pch_sec:proxy}

Pondering over the Maunder Minimum puzzle leads naturally into one major difficulty that plagues *any* and *all* dynamo models and MHD simulations discussed in the preceding chapter, when trying to reproduce this or that solar cycle fluctuation patterns seen in the sunspot number: what is the *quantitative* relationship between the internal large-scale magnetic field and the number of sunspots emerging at the solar surface? The process through which the dynamo-generated magnetic field produces toroidal flux ropes in the tachocline is not understood, but the few extant calculations attempting to simulate this formation process indicate that it is much more than a mere matter of toroidal field strength either. The destabilization and rise of these toroidal flux ropes is also not just a matter of field strength, as the stability diagram of Fig. 47 already shows quite well. Once the flux rope emerges, it is not at all clear that the number of sunspots is uniquely and proportionally related to the magnetic field strength or flux in the rope; a bipolar magnetic region made up of two monolithic sunspots would contribute $10 + 2$ to the sunspot number, as defined by eq. (265); with the trailing component of the bipolar region broken up into 10 small spots (say), as often observed, one gets instead $10 + 11$ to the sunspot number; this is a difference by nearly a factor of two, for the the same magnetic flux!

The fact remains that the sunspot number does correlate well with other more “physical” measures of the solar magnetic field, such as photospheric magnetic flux, active region magnetic flux, and the F10.7 solar radio flux (see Fig. 55). Until strong evidence to the contrary is presented, it is probably reasonable to assume that a more strongly magnetized sun will produce more sunspots, but it would be really surprising if that relationship were nicely and conveniently linear over a wide range of overall activity levels

We are used to thinking of sunspot numbers as a proxy for the solar internal magnetic field; but starting from a dynamo solution for the solar large-scale magnetic field, we must now construct a proxy for the sunspot number! Consider the following equally “reasonable” simple proxies: the total magnetic energy, the magnetic energy within the tachocline, and the net toroidal magnetic flux in the tachocline:

$$\text{SSN}_1(t) = \int_V \mathbf{B}^2(r, \theta, t) dV , \quad (267) \quad \{\text{eq:ssn1}\}$$

$$\{\text{eq:ssn2}\} \quad \text{SSN}_2(t) = 2\pi \int_0^\pi \int_{r_c-w}^{r_c+w} \mathbf{B}^2(r, \theta, t) r^2 \sin \theta dr d\theta , \quad (268)$$

$$\{\text{eq:ssn3}\} \quad \text{SSN}_3(t) = \left| 2\pi \int_0^\pi \int_{r_c-w}^{r_c+w} B(r, \theta, t) r^2 \sin \theta dr d\theta \right| , \quad (269)$$

Even though these three proxies are closely related, they lead to SSN proxy timeseries that show some significant differences. This is illustrated on Figure 57, for the advection-dominated mean-field $\alpha\Omega$ model of §0.26.11, subjected to stochastic forcing (more on this shortly, in §0.31). The plot covers a period spanning a gradual rise and decline in cycle amplitude, and the time series are normalized so as to yield the same peak amplitude for the first cycle plotted. It is reassuring to see all three proxies correlating rather well, and showing similar long-term trends in cycle amplitude and duration, yet significant differences are also apparent. Close examination of Figure 57 reveals that the timing of cycle maxima and minima differs slightly from one proxy to the other, moreover in a manner dependent on the cycle amplitude. Note also how the relative differences between the three sunspot proxies depend significantly on the cycle amplitude, approaching 40% for some cycle. For the purposes of the foregoing discussion, these differences are inconsequential, except perhaps when considering cycle prediction schemes (§0.35), where the aim is to predict sunspot number as accurately as possible. With this important caveat under the belt, we proceed with our study of solar cycle fluctuations, using the total magnetic energy as a SSN proxy.

`{pch_sec:stoch}`

0.31 Cycle modulation through stochastic forcing

An obvious means of producing amplitude fluctuations in dynamo models is to introduce stochastic forcing in the governing equations. Sources of stochastic “noise” certainly abound in the solar interior; large-scale flows in the convective envelope, such as differential rotation and meridional circulation, are observed to fluctuate, an unavoidable consequence of dynamical forcing by the surrounding, vigorous turbulent flow. This convection is known to produce its own small-scale magnetic field (viz. Fig. 28), and amounts to a form of rapidly varying zero-mean “noise” superimposed on the slowly-evolving mean magnetic field. This can be readily incorporated into dynamo mod-

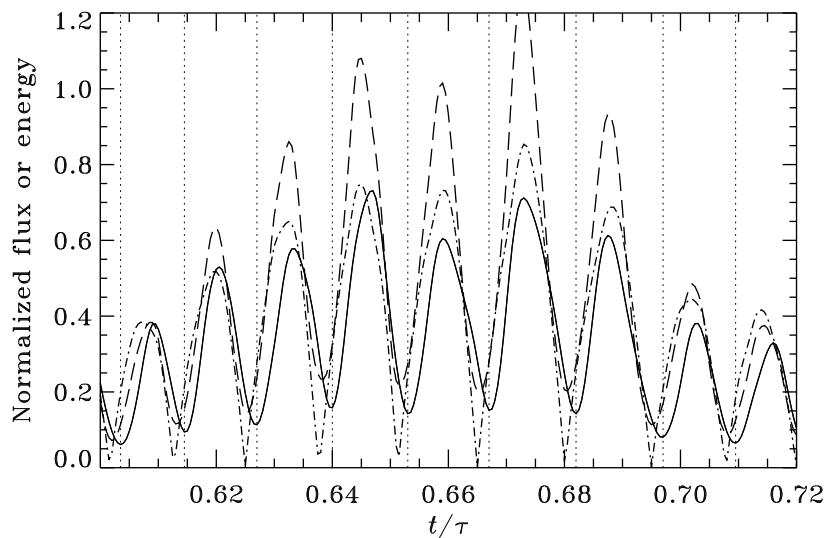


Fig. 57 fig:ssnproxy Time series of the three different sunspot number proxies defined through eqs. (267)—(269), constructed for the advection-dominated mean-field $\alpha\Omega$ solution of Fig. 38, subjected to stochastic fluctuations. The solid line is the total magnetic energy in the domain, the dashed line the magnetic energy within the tachocline, and the dash-dotted line the net toroidal magnetic flux within the tachocline. The three time series have been normalized so that they yield the same peak amplitude for the first cycle displayed here.

els by introducing, on the RHS of the governing equations, an additional zero-mean source term localized at the surface, and varying randomly from node-to-node in latitude and from one time step to the next:

$$A(R, \theta, t) \rightarrow A(R, \theta, t) + \rho \times \delta A, \quad \rho \in [-1, 1], \quad (270) \quad \{\text{eq:pnoise}\}$$

with δA a fixed amplitude, and the random number ρ is uniformly distributed in the interval. Note that the current-free boundary condition at $r/R = 1$ for the toroidal component requires $B(R, \theta, t) = 0$, therefore we only add a perturbation to the poloidal component. This is a classical instance of *additive noise*.

In addition, the azimuthal averaging implicit in all models of the solar cycle considered earlier will yield dynamo coefficients showing significant deviations about their mean values, as a consequence of the spatiotemporally discrete nature of the physical events (e.g., cyclonic updrafts, sunspot emergences, flux rope destabilizations, etc.) whose collective effects add up to produce a mean electromotive force. The impact of these statistical fluctuations about the mean can be modeled in a number of ways. Perhaps the most straightforward is to let the dynamo number fluctuate randomly in time about some pre-set

mean value \bar{C}_α :

$$C_\alpha \rightarrow \bar{C}_\alpha + \rho \times \delta C_\alpha, \quad \rho \in [-1, 1], \quad \text{if}(t \bmod \tau_c) = 0. \quad (271) \quad \{\text{eq:snoise}\}$$

By most statistical estimates, the expected magnitude of these fluctuations is quite large, i.e., many times the mean value, a conclusion also supported by “measurements” of the α -tensor components in numerical simulations. One typically also introduces a *coherence time* (τ_c) during which the dynamo number retains a fixed value. At the end of this time interval, this value is randomly readjusted. Depending on the dynamo model at hand, the coherence time can be related to the lifetime of convective eddies (α -effect-based mean-field models), to the decay time of sunspots (Babcock-Leighton models), or to the growth rate of instabilities (hydrodynamical shear or buoyant MHD instability-based models). Equation (271) represents an instance of *multiplicative noise*, since the fluctuating quantity is multiplying a source term in the governing equations, which is itself a function of the system’s dependent variables.

The effect of stochastic forcing varies according to the type of dynamo model being forced, but some common trends and tendencies nonetheless emerge. In most models stochastic forcing or noise increases both the average amplitudes and durations of cycles. It also introduces long-timescale modulations in the overall cycle amplitudes, “long” in the sense of being much longer than the assumed coherence time for the noise and/or forcing, and often significantly longer than the cycle period itself. Often this can be traced to the production and storage of strong magnetic fields in the low-diffusivity regions of the domain, below the core-envelope interface, where the resistive decay time of these structures can be quite long.

Figures 58 and 59 show some representative results for the advection-dominated mean-field $\alpha\Omega$ model of §0.26.11, and for the Babcock-Leighton model of §0.27.5, respectively. In both cases the total magnetic energy (red line on panels A) is used as a proxy for sunspot number. These two specific stochastically forced solutions were selected because they exhibit a number of solar-like features, including relative ranges of variations in cycle amplitudes ($\pm \sim 40\%$ of the mean) and duration ($\pm \sim 15\%$ of the mean), amplitude modulation patterns spanning many cycles, and shorter-lived Dalton-minimum-like intervals of markedly reduced amplitude.

Both of these solutions (and many of their “cousins” computed with varying amplitude of stochastic forcing) do fairly well at reproducing Gnevyshev-Ohl-like alternating patterns of variations in cycle amplitude about their running mean. This is illustrated on panels (B) of both Figure. A 1-2-1 running mean of cycle amplitudes is first computed according to eq. (266), yielding the thin purple line on panels (A). This is then subtracted from the temporal sequence of cycle amplitudes, to give the “detrended” amplitudes plotted on panels (B). The gray horizontal bars flag the temporal intervals during which a regular alternance of above-and-below the mean is sustained. That

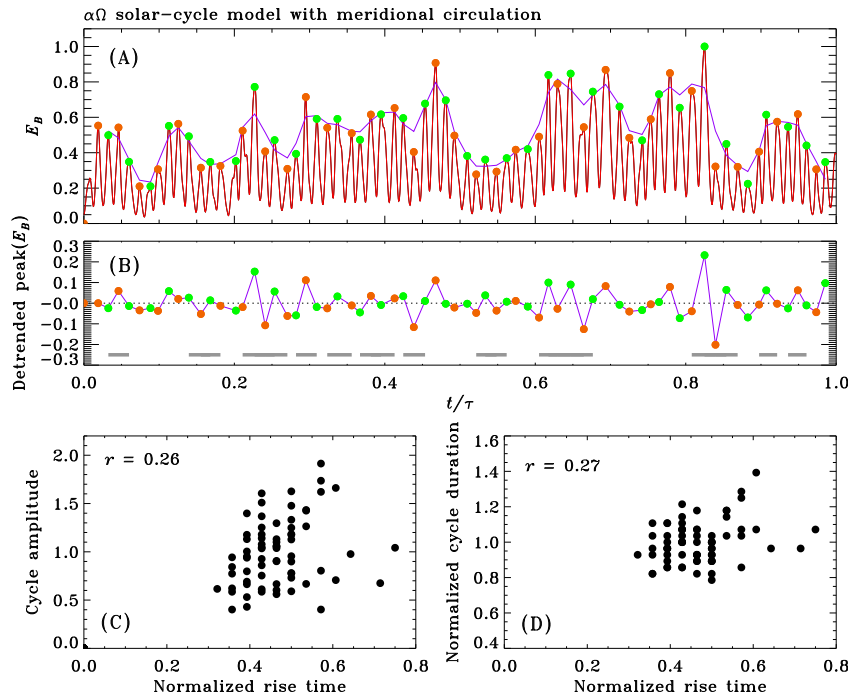


Fig. 58 Impact of stochastic fluctuations of the C_α dynamo number on the behavior of an advection-dominated mean-field solar cycle model including a meridional circulation. This is the solution of Fig. 38, with 100% forcing of the poloidal dynamo number ($\delta C_\alpha / \bar{C}_\alpha = 1$ in eq. (271)). The top panel shows part of a time series for the magnetic energy (red), together with a 1-2-1 running mean of the peak amplitudes (purple), as defined in eq. (266). Subtracting this from the temporal sequence of peak amplitudes yield the “detrended” sequence shown on panel (B), where odd- (even-) numbered cycles are plotted in red (green), and the horizontal gray bars indicate epochs where a Gnevyshev-Ohl-like pattern holds. Panels (C) and (D) are Waldmeier-Rule-like correlation plots between cycle peak, rise time and duration (cf. Fig. 54), with cycle peak and duration normalized to their mean values over the full simulation run.

{fig:aocmstoch}

such sequences should exist is in itself not surprising, in view of the detrending procedure adopted here; purely random numbers would distribute themselves symmetrically about their mean, so that the Gnevyshev-Ohl patterns can materialize only by chance. What is striking here is the distributions of durations for these epochs, which can greatly exceed (especially here in the Babcock-Leighton solution) what one could rightfully expect from Poissonian statistics.

Most stochastically forced models, including the two shown on Figs. 58 and 59, do produce a positive correlation between cycle rise time and duration (cf. panels (C) and Fig. 54C). In the case of the mean-field model of Fig. 58 that

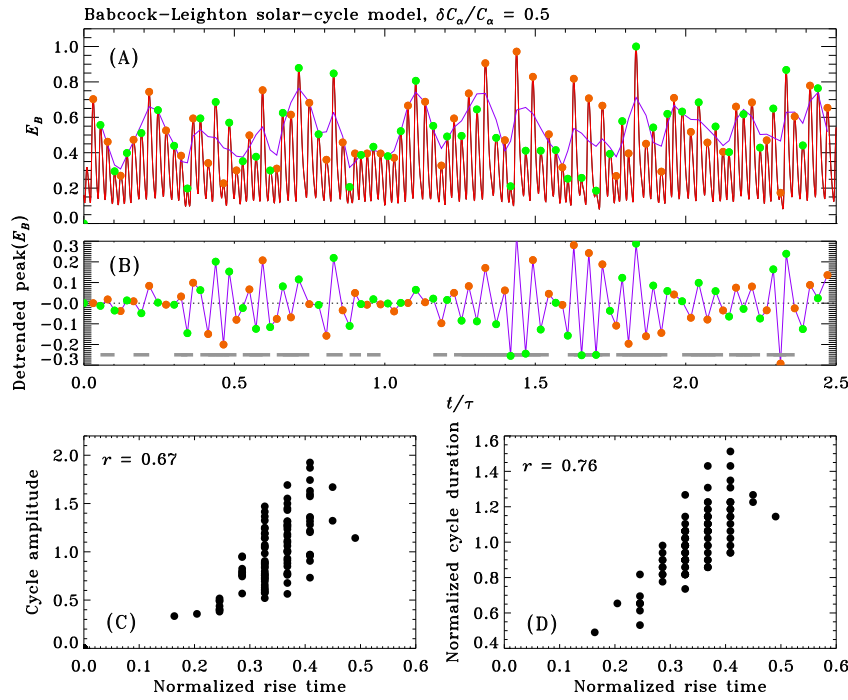


Fig. 59 Identical in format to Fig. 58, but now the parent model is the Babcock-Leighton solution of Fig. 45, with poloidal source term fluctuating at the level $\delta C_\alpha / C_\alpha = 0.5$. The moderately strong positive correlation between cycle amplitudes and rise time is markedly non-solar, but the similar correlation between cycle duration and rise time is in better agreement with solar cycle data.

{fig:BLstoch}

correlation is too weak compared to solar, while for the Babcock-Leighton model it is too strong, but adjustment of the amplitude of stochastic forcing can readily yield a more solar-like correlation.

The situation is nowhere as good with regards to the observed anticorrelation between cycle amplitude and rise time (or duration) embodied in the Waldmeier Rule (viz. Fig. 54A). Whether forced stochastically through the dynamo number or via additive noise in the surface layers, most of the kinematic models considered here end up producing a positive correlation (rather than an anticorrelation) between these two cycle parameters. A Waldmeier-like anticorrelation has been observed in stochastically-forced linear $\alpha\Omega$ model near criticality³³, but this interesting result in general does not carry over to nonlinearly-saturated $\alpha\Omega$ dynamo solutions. It has also been observed in a Babcock-Leighton models subjected to stochastic pertur-

³³ See the paper by Ossendrijver & Hoyng (1996) cited in the bibliography.

bations imposed on the form of the meridional flow profile³⁴, but again it is not clear how generic or robust this actually is. It may well be that the key to the Waldmeier Rule lies at least in part with non-kinematic effects, such as the nonlinear backreaction of the dynamo-generated magnetic field on differential rotation and/or meridional circulation.

0.32 Cycle modulation through the Lorentz force

{pch_sec:Lforce}

The dynamo-generated magnetic field will, in general, produce a Lorentz force that will tend to oppose the driving fluid motions. This is a basic physical effect that should be included in any dynamo model. It is not all trivial to do so, however, since in a turbulent environment both the fluctuating and mean components of the magnetic field can affect both the large-scale flow components, as well as the small-scale turbulent flow providing the Reynolds stresses powering the large-scale flows. One must thus distinguish between two (related) amplitude-limiting mechanisms:

1. Lorentz force associated with the mean magnetic field *directly* affecting large-scale flows (sometimes called the “Malkus-Proctor effect”);
2. Large-scale magnetic field *indirectly* affecting large-scale flows via effects on small-scale turbulence and associated Reynolds and Maxwell stresses (sometimes called “ \mathcal{A} -quenching”)

Introducing magnetic backreaction on differential rotation and/or meridional circulation is a tricky business, because one must then also, in principle, provide a model for the Reynolds stresses powering the large-scale flows in the solar convective envelope as well as a procedure for computing magnetic backreaction on these. This rapidly leads into the unyielding realm of MHD turbulence, although algebraic “ \mathcal{A} -quenching” formulae akin to α -quenching have been proposed based on specific turbulence models. Alternately, one can add an *ad hoc* source term to the RHS of eq. (77), designed in such a way that in the absence of the magnetic field, the desired solar-like large-scale flow is obtained. As a variation on this theme, one can simply divide the large-scale flow into two components, the first (\mathbf{U}) corresponding to some prescribed, steady profile, and the second (\mathbf{U}') to a time-dependent flow field driven by the Lorentz force

$$\mathbf{u}(\mathbf{x}, t, \mathbf{B}) = \mathbf{U}(\mathbf{x}) + \mathbf{U}'(\mathbf{x}, t, \mathbf{B}) , \quad (272) \quad \{\text{eq:Nlflow1}\}$$

with the (non-dimensional) governing equation for \mathbf{U}' including only the Lorentz force and a viscous dissipation term on its RHS. If \mathbf{u} amounts only to differential rotation, then \mathbf{U}' must obey a (nondimensional) differential equation of the form

³⁴ See the paper by Charbonneau & Dikpati (2000) cited in the bibliography.

$$\frac{\partial \mathbf{U}'}{\partial t} = \frac{A}{\rho} (\nabla \times \mathbf{B}) \times \mathbf{B} + P_m \nabla^2 \mathbf{U}' , \quad (273) \quad \{\text{eq:Nlflow2}\}$$

where time has been scaled according to the magnetic diffusion time $\tau = R^2/\eta_e$, as before. Two dimensionless parameters appear in eq. (273). The first (A) is a numerical parameter measuring the influence of the Lorentz force. The second, $P_m = \nu/\eta$, is the *magnetic Prandtl number*. It measures the relative importance of viscous and Ohmic dissipation. When $P_m \ll 1$, large velocity amplitudes in \mathbf{U}' can be produced by the dynamo-generated mean magnetic field. This effectively introduces an additional, long timescale in the model, associated with the evolution of the magnetically-driven flow; the smaller P_m , the longer that timescale.

The majority of studies published thus far and using this approach have only considered the nonlinear magnetic backreaction on differential rotation. This has been shown to lead to a variety of behaviors, including amplitude and parity modulation, periodic or aperiodic, as well as intermittency (more on the latter in §0.34). It has been argued that amplitude modulation in such models can be divided into two main classes:

1. Nonlinear interaction between modes of different parity, with the Lorentz Force-mediated flow variations controlling the transition from one mode to another;
2. Exchange of energy between a single dynamo mode (of some fixed parity) with the flow field. This leads to quasiperiodic modulation of the basic cycle, with the modulation period controlled by the magnetic Prandtl number.

Both types of modulation can co-exist in a given dynamo model, leading to a rich overall dynamical behavior. Figure 60 shows two time-latitude diagrams produced by a nonlinear mean-field interface model³⁵. The model is defined on cartesian slab with a reference differential rotation varying only with depth, and includes backreaction on the differential rotation according to the procedure described above. The model exhibits strong, quasi-periodic modulation of the basic cycle, leading to epochs of strongly reduced amplitude. Note how the dynamo can emerge from such epochs with strong hemispheric asymmetries (top panel), or with a different parity (bottom panel).

The differential rotation can also be suppressed indirectly by magnetic backreaction on the *small-scale* turbulent flow that produces the Reynolds stresses driving the large-scale mean flow. Inclusion of this so-called “ A -quenching” in mean-field dynamo models, alone or in conjunction with other amplitude-limiting nonlinearities, has also been shown to lead to a variety of periodic and aperiodic amplitude modulations, provided the magnetic Prandtl number is small³⁶. This type of models stand or fall with the turbulence model

³⁵ For details on this model see paper by Tobias (1997) cited in the bibliography.

³⁶ See, e.g., the paper by Küker et al. (1999) cited in the bibliography.

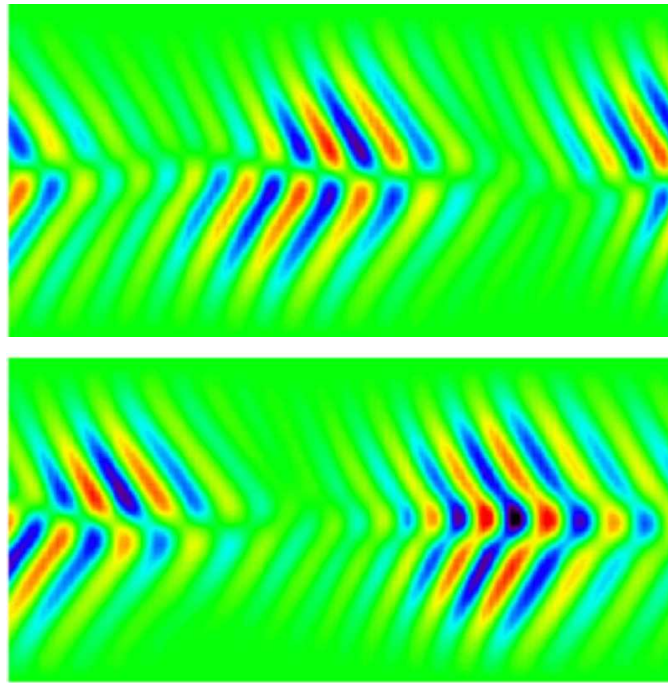


Fig. 60 Amplitude and parity modulation in a dynamo model including magnetic backreaction on the differential rotation. These are the usual time-latitude diagrams for the toroidal magnetic field, now covering both solar hemispheres, and exemplify type I (bottom) and type II (top) modulation arising in a nonlinear, non-kinematic dynamo model defined over a cartesian slab, including magnetic backreaction on the differential rotation (see text). Figure kindly provided by S.M. Tobias.

{fig:tobias}

used to compute the various mean-field coefficients, and it is not yet clear which aspects of the results are truly generic to A -quenching.

The dynamical backreaction of the large-scale magnetic field on meridional circulation has received comparatively little attention. The few calculations published so far³⁷ suggest that diffuse toroidal magnetic fields of strength up to 0.1 T can probably be advected equatorward at the core-envelope interface. That it can indeed do so is crucial models relying on the meridional flow to produce equatorward propagation of magnetic fields as the cycle unfolds. Interestingly, relatively small variations of the meridional persisting over decadal timescales have been shown to produce cycle amplitude variations that are remarkably solar-like³⁸. Such effects are potentially important

³⁷ See papers by Rempel (2001, 2002) cited in the bibliography.

³⁸ See paper by Lopes and Passos (2009) cited in the bibliography.

in all dynamo models of the flux transport variety, and clearly merit further study.

{pch_sec:tdelay}

0.33 Cycle modulation through time delays

It was already noted that in solar cycle models based on the Babcock-Leighton mechanism of poloidal field generation, meridional circulation effectively sets—and even regulates—the cycle period. In doing so, it also introduces a long time delay in the dynamo mechanism, “long” in the sense of being comparable to the cycle period. This delay originates with the time required for circulation to advect the surface poloidal field down to the core-envelope interface, where the toroidal component is produced (A→C on Fig. 44). In contrast, the production of poloidal field from the deep-seated toroidal field (C→D) is a “fast” process, growth rates and buoyant rise times for sunspot-forming toroidal flux ropes being of the order of tens of days. The first, long time delay turns out to have important dynamical consequences.

The long time delay inherent in Babcock-Leighton models of the solar cycle allows a formulation of cycle-to-cycle amplitude variations in terms of a simple one-dimensional iterative map. Working in the kinematic regime, neglecting resistive dissipation, and in view of the conveyor belt argument outlined in §0.27, the toroidal field strength (T_{n+1}) at cycle $n + 1$ is assumed to be linearly proportional to the poloidal field strength (P_n) of cycle n , i.e.,

$$\{\text{eq:map1}\} \quad T_{n+1} = aP_n . \quad (274)$$

Now, because flux eruption is a fast process, the strength of the poloidal field at cycle $n + 1$ is (nonlinearly) proportional to the toroidal field strength of the *current* cycle:

$$\{\text{eq:map2}\} \quad P_{n+1} = f(T_{n+1})T_{n+1} . \quad (275)$$

Here the “Babcock-Leighton” function $f(T_{n+1})$ measures the efficiency of surface poloidal field production from the deep-seated toroidal field. Substitution of eq. (274) into eq. (275) leads immediately to a one-dimensional iterative map:

$$\{\text{eq:map3}\} \quad p_{n+1} = \alpha f(p_n)p_n , \quad (276)$$

where the p_n ’s are normalized amplitudes, and the normalization constants as well as the constant a in eq. (274) have been absorbed into the definition of the map’s parameter α , here operationally equivalent to a dynamo number. In analogy with eq. (261), we adopt here the following nonlinear function

$$\{eq:map4\} \quad f(p) = \frac{1}{4} \left[1 + \operatorname{erf} \left(\frac{p - p_1}{w_1} \right) \right] \left[1 - \operatorname{erf} \left(\frac{p - p_2}{w_2} \right) \right], \quad (277)$$

with $p_1 = 0.6$, $w_1 = 0.2$, $p_2 = 1.0$, and $w_2 = 0.8$. This catches an essential feature of the B-L mechanism, namely the fact that it can only operate in a finite range of toroidal field strength.

A bifurcation diagram for the resulting iterative map is presented on Figure 61A. For a given value of the map parameter α , the diagram gives the locus of the amplitude iterate p_n for successive n values. The “critical dynamo number” above which dynamo action becomes possible, corresponds here to $\alpha = 0.851$ ($p_n = 0$ for smaller α values). For $0.851 \leq \alpha \leq 1.283$, the iterate is stable at some finite value of p_n , which grows gradually with α . This corresponds to a constant amplitude cycle. As α reaches 1.283, period doubling occurs, with the iterate p_n alternating between high and low values (e.g., $p_n = 0.93$ and $p_n = 1.41$ at $\alpha = 1.4$). Further period doubling occurs at $\alpha = 1.488$, then at $\alpha = 1.531$, then again at $\alpha = 1.541$, and ever faster until a point is reached beyond which the amplitude iterate seems to vary without any obvious pattern (although within a bounded range); this is in fact a chaotic regime.

As in any other dynamo model where the source regions for the poloidal and toroidal magnetic field components are spatially segregated, the type of time delay considered here is unavoidable. The Babcock-Leighton model is just a particularly clear-cut example of such a situation. One is then led to anticipate that the map’s rich dynamical behavior should find its counterpart in the original, arguably more realistic spatially-extended, diffusive axisymmetric model that inspired the map formulation. Remarkably, this is indeed the case.

Fig. 61B shows a bifurcation diagram, conceptually equivalent to that shown on part A, but now constructed from a sequence of numerical solutions of the Babcock-Leighton model discussed earlier in §0.27, for increasing values of the dynamo number. Time series of magnetic energy were calculated from the numerical solutions, successive peaks found and their peak amplitude plotted for each individual solution. The sequence of period doubling, leading to a chaotic regime, is strikingly similar to the bifurcation diagram constructed from the corresponding iterative map, down to the narrow multi-periodic windows interspersed in the chaotic domain. This demonstrates that time delay effects are a robust feature, and represent a very powerful source of cycle amplitude fluctuation in Babcock-Leighton models, *even in the kinematic regime*. Although transition to chaos does not always occur through such a classical period doubling sequence, chaos is ubiquitous in this model’s parameter space.

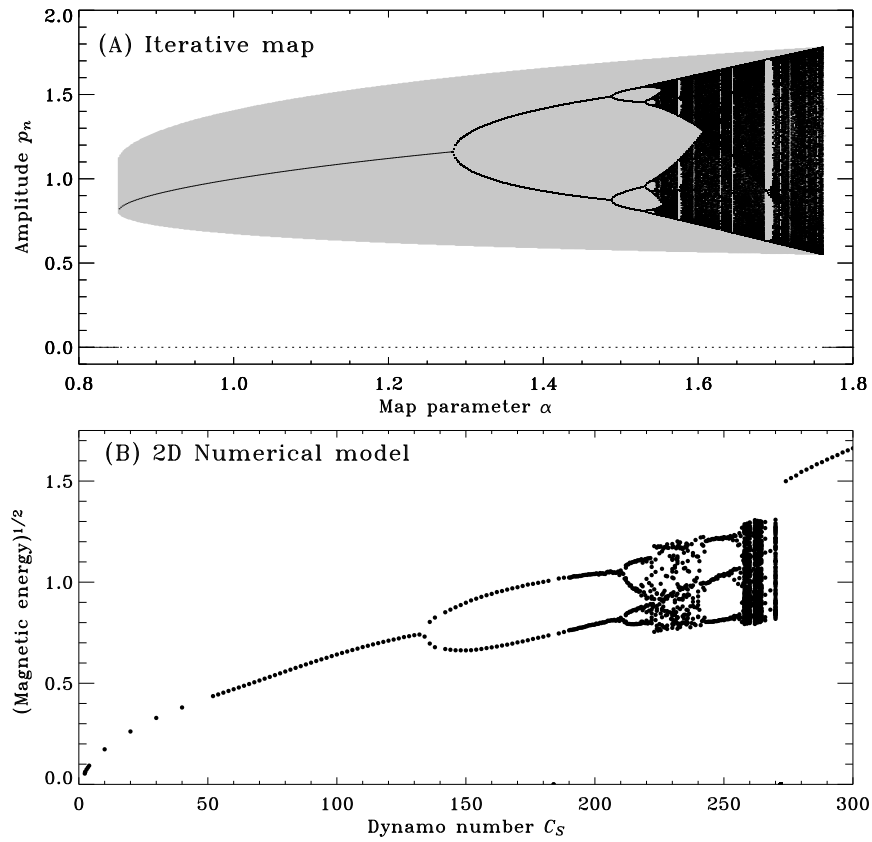


Fig. 61 Two bifurcation diagrams for a kinematic Babcock-Leighton model, where amplitude fluctuations are produced by time-delay feedback. The top diagram is computed using the one-dimensional iterative map given by eqs. (276)–(277), while the bottom diagram is reconstructed from numerical 2D kinematic solutions in spherical geometry, of the type discussed in §0.27. The shaded area in panel (A) maps the attraction basin for the cyclic solutions, with initial conditions located outside of this basin converging to the trivial solution $p_n = 0$.

{fig:bifurc}

0.34 Intermittency

{pch_sec:intermitt}

The term “intermittency” refers to systems undergoing apparently random, rapid switching from quiescent to bursting behaviors, as measured by the magnitude of some suitable system variable. In the context of solar cycle model, intermittency is invoked to explain the existence of Maunder

Minimum-like quiescent epochs of strongly suppressed activity randomly interspersed within periods of “normal” cyclic activity³⁹.

Intermittency has been shown to occur through stochastic fluctuations of the dynamo number in mean-field dynamo models operating at or near criticality⁴⁰. This mechanism for “on-off intermittency” works well, however there is no strong reason to believe that the solar dynamo is running just at criticality, so that is not clear how good an explanation this is of Maunder-type grand minima. Parity modulation driven by stochastic noise can also lead to a form of intermittency in linear or α -quenched models, by exciting the higher-order modes that perturb the normal operation of the otherwise dominant dynamo mode, producing marked hemispheric asymmetries and strongly reduced cycle amplitudes. The transition from active to quiescent (and vice versa) being controlled by stochastic noise, the durations of active and quiescent phases tend to have exponential distributions, which agrees tolerably well with inferences from the radioisotope record⁴¹.

Another way to trigger intermittency in a dynamo model is to let nonlinear dynamical effects, for example a reduction of the differential rotation amplitude, push the effective dynamo number below its critical value; dynamo action then ceases during the subsequent time interval needed to reestablish differential rotation following the diffusive decay of the magnetic field; in the low P_m regime, this time interval can amount to many cycle periods, but P_m must not be too small, otherwise grand minima become too rare. Values $P_m \sim 10^{-2}$ seems to work best. Such intermittency is most readily produced when the dynamo is operating close to criticality⁴².

Intermittency has also been observed in strongly supercritical model including α -quenching as the sole amplitude-limiting nonlinearity. Such solutions can enter grand minima-like epochs of reduced activity when the dynamo-generated magnetic field completely quenches the α -effect. The dynamo cycle restarts when the magnetic field resistively decays back to the level where the α -effect becomes operational once again⁴³.

³⁹ It should be noted, however, that dearth of sunspots does not necessarily means a halted cycle; as noted earlier, flux ropes of strengths inferior to ~ 1 T will not survive their rise through the convective envelope, and the process of flux rope formation from the dynamo-generated mean magnetic field may itself be subjected to a threshold in field strength. The same basic magnetic cycle may well have continued unabated all the way through the Maunder Minimum, but at an amplitude just below one of these thresholds. This idea finds support in the ¹⁰Be radioisotope record, which shows an uninterrupted cyclic signal through the Maunder minimum (see Fig. 56).

⁴⁰ See paper by Ossendrijver & Hoyng (1996) cited in the bibliography for a particularly lucid discussion.

⁴¹ For more on this version of noise-driven intermittency, see the papers by Mininni & Gomez (2004) and Moss *et al.* (2008) cited in the bibliography.

⁴² See, e.g., the papers by Küker *et al.* (1999) and Brooke *et al.* (2002) cited in the bibliography.

⁴³ For representative models exhibiting intermittency of this type, see the paper by Tworkowski *et al.* (1998) cited in the bibliography.

Intermittency can also arise naturally in dynamo models characterized by a lower operating threshold on the magnetic field. These include models where the regeneration of the poloidal field takes place via the MHD instability of toroidal flux tubes (§0.28.2). In such models, the transition from quiescent to active phases requires an external mechanism to push the field strength back above threshold. This can be stochastic noise⁴⁴, or a secondary dynamo process normally overpowered by the “primary” dynamo during active phases. Figures 62 show one representative solution of the latter variety, where intermittency is driven by a weak α -effect-based kinematic dynamo operating in the convective envelope, in conjunction with magnetic flux injection into the underlying region of primary dynamo action by randomly positioned downflows⁴⁵. The top panel shows a sample trace of the toroidal field, and the bottom panel a butterfly diagram constructed near the core-envelope interface in the model.

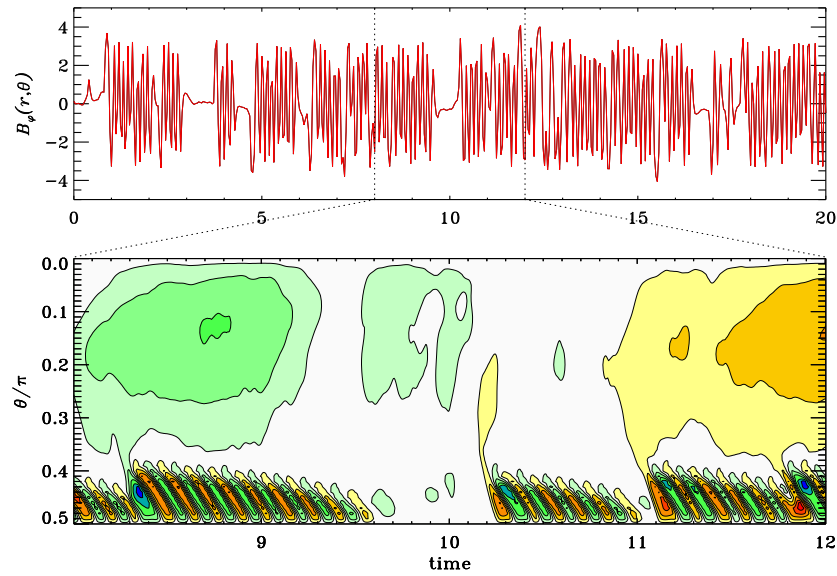


Fig. 62 Intermittency in a dynamo model based on flux tube instabilities (cf. §0.28.2). The top panel shows a trace of the toroidal field, and the bottom panel is a butterfly diagram covering a shorter time span including a quiescent phase at $9.6 \lesssim t \lesssim 10.2$, and a “failed Minimum” at $t \simeq 11$. Figure produced from numerical data kindly provided by M. Ossendrijver.

{fig:mathieu}

Dynamo models exhibiting amplitude modulation through time-delay effects are also liable to show intermittency in the presence of stochastic noise.

⁴⁴ See the paper by Schmitt et al. (1996) cited in the bibliography.

⁴⁵ for more details see paper by Ossendrijver (2000) in bibliography

This intermittency mechanism hinges on the fact that the map's attractor has a finite basin of attraction (indicated by gray shading on Fig. 61A). Stochastic noise acting simultaneously with the map's dynamics can then knock the solution out of this basin of attraction, which then leads to a collapse onto the trivial solution $p_n = 0$, even if the map parameter remains supercritical. Stochastic noise eventually knocks the solution back into the attractor's basin, which signals the onset of a new active phase. This behavior does materialize in the Babcock-Leighton model of §0.27. Figure 63 shows one such representative solution, in the same format as Fig. 62. This is a dynamo solution which, in the absence of noise, operates in the singly-periodic regime. Stochastic noise is added to the vector potential $A\hat{e}_\phi$ in the surface layers, and the dynamo number is also allowed to fluctuate randomly about a pre-set mean value, as described in §0.31. The resulting solution exhibits both amplitude fluctuations and intermittency.

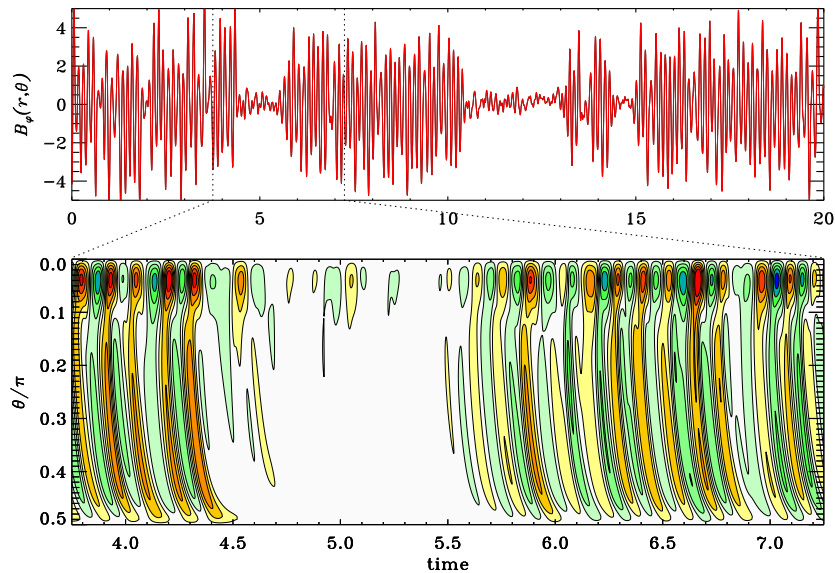


Fig. 63 Intermittency in a dynamo model based on the Babcock-Leighton mechanism (cf. §0.27). The top panel shows a trace of the toroidal field sampled at $(r, \theta) = (0.7, \pi/3)$. The bottom panel is a time-latitude diagram for the toroidal field at the core-envelope interface. Numerical data from the Charbonneau et al. (2004) paper cited in the bibliography.

{fig:BLintermitt}

With its strong polar branch often characteristic of dynamo models with meridional circulation, Fig. 63 is not a particularly good fit to the sunspot butterfly diagram. Yet its fluctuating behavior is solar-like in a number of ways, including epochs of alternating higher-than-average and lower-than-average cycle amplitudes (the Gnevyshev-Ohl rule, cf. Fig. 54), and residual

pseudo-cyclic variations during quiescent phases, as suggested by ^{10}Be data. This latter property is due at least in part to meridional circulation, which continues to advect the (diffusively decaying) magnetic field after the dynamo has fallen below threshold.

{pch_sec:predict}

0.35 Model-based cycle predictions

Over the past decade, the prediction of solar eruptive events and their geomagnetic impacts, known as *space weather*, has become a Very Big Business. Even then, the prediction of the overall level of solar activity is also of interest, as it could be useful, among other things, to the planning of space missions and interplanetary travel. The understanding and prediction of activity levels on timescales decadal and longer is becoming known as *space climate*, and its primary data are the time series of sunspot numbers, and proxies such as the radioisotopes records.

One “hot” prediction problem, lying at the boundary of space weather and space climate, is forecasting the characteristic of the next solar activity cycle, which is usually equated with the timing and amplitude of the cycle as measured in the sunspot number time series (see Fig. 52). It is of course possible to treat this prediction problem as an exercise in time series analysis and forecasting, without any physical input. The SSN time series is just a time series, and it can be extended using a number of techniques coming from statistics (spectral analysis, wavelets, etc) or dynamical system theory (such as attractor reconstruction). To this day, forecasts based on such techniques have not fared significantly better than so-called “climatological” forecasts, which consists in simply “predicting”, e.g., that the next cycle will have the same amplitude as the current cycle, or an amplitude equal to the mean cycle amplitude over the length of the sunspot record, etc. In this section we will focus here instead on prediction schemes based, in one form or another, on dynamo models.

In light of what we have learned thus far, we know we are facing a number of difficulties in trying to use dynamo models to forecast the solar cycle. A basic list of questions that need to be answered (excluding technical details for the time being) should include, at the very least, the following:

1. What type of dynamo powers the solar cycle: $\alpha\Omega$, $\alpha^2\Omega$, interface, Babcock-Leighton, etc.?
2. Which mechanism is driving duration and amplitude fluctuations: stochastic forcing, nonlinear modulation due to the Lorentz force, or time delay, etc.?
3. How do we accurately “predict” sunspot number from a dynamo solution which describes the spatiotemporal evolution of just the diffuse, large-scale magnetic field?

It is a sobering fact that none of these very basic fundamental questions can be answered with confidence at this writing. Nonetheless, we have learned some important things that are useful in the forecasting context. To start with, the dynamo feeds on the existing magnetic field, therefore trying to forecast the next cycle using characteristics of the current cycle (and maybe recent past cycles as well) is definitely justified. This is the physical underpinning of all so-called “precursor methods”, which we’ll first look into.

0.35.1 The solar polar magnetic field as a precursor

{pch_ssec:precforecast}

Temporally extended synoptic magnetograms, such as Fig. 32, suggest that the solar cycle can be divided into sequences of substeps whereby a poloidal field (P) produces a new toroidal component (T), which then leads to the buildup of a new poloidal component, with accompanying polarity reversals; schematically:

$$\dots \rightarrow P(+)\rightarrow T(-)\rightarrow P(-)\rightarrow T(+)\rightarrow P(+)\rightarrow \dots \quad (278) \quad \{\text{eq:schcycle1}\}$$

This suggests that the optimal precursor for the amplitude of the sunspot-generating toroidal component should be sought by moving back up the causal chain by one substep, to the poloidal component produced in the previous sunspot cycle. This is the basis for the set of dynamo-inspired precursor schemes pioneered by A.I. Ohl and brought to maturity by K. Schatten and collaborators now over thirty years ago (see bibliography).

This idea is readily tested using our various dynamo models, as illustrated on Figure 64 for the stochastically forced Babcock-Leighton model of Fig. 59. The top panel shows a short segment of the magnetic energy time series, used again here as a proxy for the sunspot number, together with a time series of the surface polar field strength (in green). The bottom panel shows a time-latitude diagram of the surface radial magnetic component, together with the latitudes of peak toroidal field strength at the core-envelope interface, where sunspots are presumed to originate. The overall spatiotemporal evolution of the surface field, and its phase relationship to the deep-seated toroidal field, are both remarkably solar-like. Examination of the two proxies on the top panels of Figs. 64 reveals that the surface radial field peaks shortly following what one would identify with solar minimum on the basis of our SSN proxy. It is then a simple matter to pair the peak polar field at solar minimum with the SSN proxy of the following cycle, as indicated on Fig. 64 by the purple connecting line segments.

The next step is to correlate the peak poloidal field and peak SSN proxy, in order to ascertain the viability of poloidal field-based precursor schemes. The result is shown on Figure 65A, for three different levels of stochastic forcing, as color-coded, with the solution of Fig. 64 in red. In all cases the time series for the SSN proxies and surface poloidal field strength have been

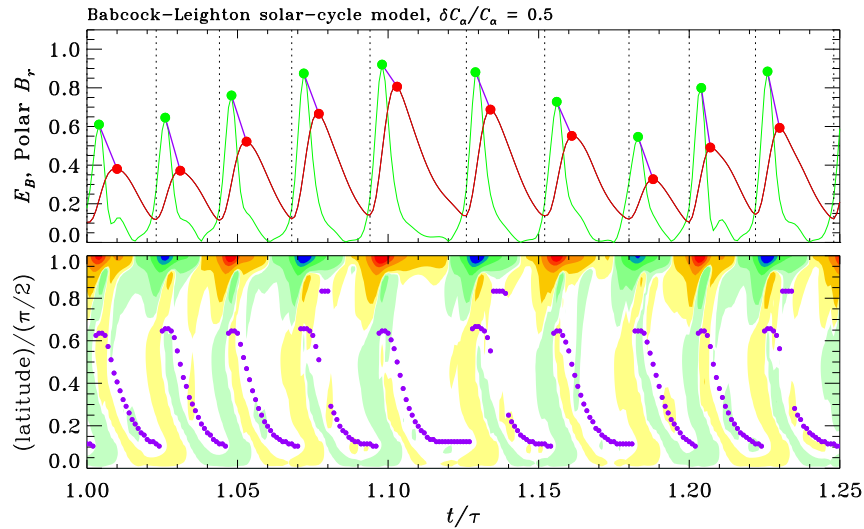


Fig. 64 Portion of a simulation run of a Babcock-Leighton model, with fluctuations at the $\pm 50\%$ level imposed in the magnitude of the surface source term. The unperturbed reference solution is that illustrated on Figure 45. The top panel shows time series of the surface radial magnetic field sampled at the pole (green), together with a time series of the total magnetic energy (red), used here as a proxy for the sunspot number. The purple line segment join the peak poloidal field at (or near) “sunspot minimum” with the peak in the SSN proxy for the following cycle. The bottom panel is a time-latitude diagram of the surface radial field, and the purple dots trace the latitude of peak toroidal field strength at the core-envelope interface as a function of time. Figure taken from the Charbonneau & Barlet (2011) paper cited in the bibliography. Fig: pFCBL1

normalized to the peak values characterizing a parent run without stochastic forcing.

The peak polar field at solar minimum clearly has precursor value, but stochastic forcing rapidly degrades the forecasting accuracy. Consider for example the solution with 100% fluctuation of the dynamo number C_α (green); while a linear correlation coefficient of 0.83 may sound pretty good, the fact remains that for a polar field of 0.8 (say) in the normalized units of Fig. 65 would lead to a SSN forecast covering a very broad range, namely 0.6–1.2 in normalized units, which is not a very accurate forecast at all.

Performing the same analysis on our other solar cycle model reveals that the polar surface field is also a good precursor of cycle amplitude for the advection-dominated mean-field model with meridional circulation of §0.26.11, more robust with respect to high-amplitude stochastic forcing in fact, but no precursor at all for the classical $\alpha\Omega$ model of §0.26.10. This curious situation can be traced to the fact that in the former, the surface polar field does feed back into the dynamo loop, as circulation drags it down back into the tachocline, where it merges with the poloidal field produced there

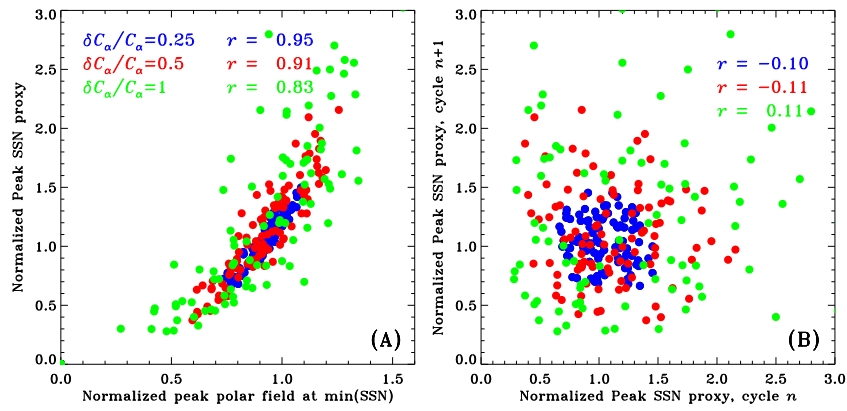


Fig. 65 (A) Correlations between peak SSN and surface poloidal field strength in the stochastically-forced Babcock-Leighton solutions of Fig. 45, for three different levels $\delta C_\alpha / C_\alpha$ of forced stochastic fluctuations in the surface source term, color-coded as indicated. The red dots correspond to the simulation run illustrated on Fig. 64. The amplitudes are normalized to those characterizing the non-fluctuating parent simulation. The linear correlation coefficients r are again given. (B) Similar plot, but this time attempting to correlate the peak SSN peak values for pairs of successive cycles. Fig: BLprecursor

by the α -effect (see Fig. 38). In the circulation-free models, on the other hand, the poloidal field diffuses more or less radially outwards to the surface, with poloidal field of the subsequent cycle being generated completely independently at the base of the envelope (see Fig. 35).

In retrospect, the logic behind Schatten *et al.*'s precursor argument can be understood to hold only for a subset of dynamo models, namely those where some “feedback” of the surface polar field on the dynamo loop takes place. In Babcock-Leighton-type models, the surface field is the sole source of the next cycle’s toroidal field, and so is good precursor. In advection-dominated mean-field models including circulation, the surface poloidal field is a significant source of toroidal field, albeit not the only one. In classical $\alpha\Omega$ mean-field models of the type considered in §0.26.10, where the surface field is only a “passive” manifestation of dynamo action taking place independently in the deep interior, the surface field has no precursor value.

As final point of interest, it is noteworthy that in all solar cycle models considered here, the value for the peak SSN proxy has little or no precursor value in forecasting the next SSN proxy peak. This is illustrated on Fig. 65B for the set of stochastically-forced Babcock-Leighton solutions. This is somewhat surprising, since the peak polar field at the solar minimum separating two successive cycles is here a rather good precursor (cf. Fig. 65A). This situation can be traced to the manner in which stochasticity is introduced in the model. In the case of imposed stochastic fluctuations in the poloidal source

term, the scheme given by eq. (278) must be replaced by something like:

$$\dots \xrightarrow{\text{stoch}} P(+)\text{ } \longrightarrow T(-)\text{ } \xrightarrow{\text{stoch}} P(-)\text{ } \longrightarrow T(+)\text{ } \xrightarrow{\text{stoch}} P(+)\text{ } \longrightarrow \dots \quad (279) \quad \{\text{eq:schcycle2}\}$$

Precursor forecasts based on either component is only possible if the forecast does not go across a “stochastic” arrow. For classical mean-field models, where the shear and α -effect are spatially coincident and operate concurrently in time, the above sequence should instead be schematized as:

$$\{\text{eq:schcycle3}\} \quad \dots \begin{array}{c} \text{stoch} \\ \times \\ \text{stoch} \end{array} P(+)\text{ } \begin{array}{c} \text{stoch} \\ \times \\ \text{stoch} \end{array} P(-)\text{ } \begin{array}{c} \text{stoch} \\ \times \\ \text{stoch} \end{array} P(+)\text{ } \begin{array}{c} \text{stoch} \\ \times \\ \text{stoch} \end{array} \dots \quad (280)$$

which precludes any precursor relationship from previous-cycle magnetic field measurements... unless of course stochastic forcing is very weak or absent, and cycle amplitude modulation is produced primarily by deterministic effects, as briefly considered in §0.32.

`{pch_ssec:modforecast}`

0.35.2 Model-based prediction using solar data

Some recent solar cycle amplitude forecasts have used solar cycle models of the Babcock-Leighton variety (§0.27), in conjunction with input of solar magnetic field observations in a manner usually (and, strictly speaking, incorrectly) described as “data assimilation”. It is particularly instructive to compare the forecast schemes (and cycle 24 predictions) of Dikpati and collaborators on the one hand, and Choudhuri and collaborators on the other. As detailed in Table 3 below, these two schemes are remarkably similar, differing mostly in their formulation of the poloidal source term, solar data used to drive the model, and manner in which this driving is implemented.

As similar as they may be, except at the level of what one would usually consider modelling details, these two forecasting schemes end up producing cycle 24 amplitude forecasts that stand at opposite ends of the very wide range of cycle 24 forecasts produced by other techniques, as well as opposite ends of the range of past cycle amplitudes. A cycle 24 with SSN=80 would place it amongst the weakest of the past century, while SSN=180 would make it second only to the highest cycle amplitude on record, that for cycle 19 (see Fig. 52).

These model-based forecast have been subjected to strident criticism, for a variety of reasons. One of the most fundamental is the possibility —some would say “certainty”— that the solar dynamo is a nonlinear system operating in the chaotic regime, in which case long-term prediction is severely restricted by the exponential divergence of trajectories of the model in phase space. This criticism probably does not apply to the DdTG scheme, which is really a quasi-linear magnetic flux processing “machine”, rather than a truly nonlinear dynamo model; it probably does not apply either to the CCJ

{pch_tab:pred}

Table 3 Two dynamo-based solar cycle forecasting schemes

Authors	Dikpati, deToma & Gilman	Choudhuri, Chatterjee, & Jiang
Code name	DdTG	CCJ
Reference	GRL 33 , L05102 (2005)	PRL 98 , 131103 (2007)
Dynamo model	kinematic axisymmetric Babcock-Leighton	kinematic axisymmetric Babcock-Leighton
Core-CZ interface	$r/R = 0.7$	$r/R = 0.7$
Magnetic diffusivity	eq. (128), $\Delta\eta = 300$ plus high- η surface layer	eq. (128), $\Delta\eta = 10^4$
Differential rotation	solar-like parameterization eqs. (139)–(140), $w/R = 0.05$ all other parameters same	solar-like parameterization eqs. (139)–(140), $w/R = 0.015$ all other parameters same
Meridional circulation	single cell per quadrant closing at $r/R = 0.71$	single cell per quadrant closing at $r/R = 0.635$
Poloidal source term	data-driven surface forcing plus weak tachocline α -effect	subsurface α -effect plus buoyancy algorithm
Nonlinearity	algebraic α -quenching only in tachocline α -effect	algebraic α -quenching in subsurface α -effect
Solar data	time series of total sunspot area used to (continuously) drive parametric surface forcing of A	DM Index used to reset amplitude of A at “solar minimum”
Calibration interval	Cycles 16–23	Cycles 21–23
Cycle 24 forecast	SSN=155–180	SSN=80

scheme, which uses a simple algebraic amplitude-quenching nonlinearity that is usually not conducive to chaotic modulation, although this remains to be verified in the context of their specific choice of dynamo model. More to the point has been the explicit demonstration that very small changes in some unobservable and poorly constrained input parameters to the dynamo model used for the forecast, or alternate but by all appearances equally reasonable means of carrying out data input into the model, can introduce significant errors already for next-cycle amplitude forecasts⁴⁶.

In the context of Babcock-Leighton models, this model-based approach to forecasting is definitely viable in principle, since the solar surface magnetic field is that which will serve as seed to produce the sunspot-generating toroidal component of the next cycle. The one thing that the two model-based forecasting schemes compiled in Table 3 have demonstrated, beyond any doubts, is that modelling details matter a lot.

Bibliography:

On pre- an early-telescopic observations of sunspots, begin with

Mitchell, W.M., “The history of the discovery of the solar spots”, in *Popular Astronomy*, **24**, 22-ff (1916),

⁴⁶ See, e.g., the papers by Bushby & Tobias (2007) and Cameron & Schüssler (2007) cited in the bibliography

- Shea, W.R., “Galileo, Scheiner, and the interpretation of Sunspots”, *Isis*, **61**, 498-519 (1970),
- Vaquero, J.M., and Vázquez, M., “The Sun recored through history”, Springer (2009),
- Reeves, E., and Van Helden, A. (eds. & trans.), “Galileo Galilei & Christoph Scheiner on sunspots”, University of Chicago Press (2010).

If such historical issues are of interest to you, you can also consult the ever-being-enlarged Web site “Great Moments in the History of Solar Physics”:

<http://www.astro.umontreal.ca/~paulchar/grps>

and click on ”History of Solar Physics” at left. For everything you ever wanted know on the characterisations of the sunspot cycle, start with

Hathaway, D.H., *Liv. Rev. Solar Phys.*, **7**, 1 (2010),
<http://solarphysics.livingreviews.org/Articles/lrsp-2009-4/>

On the Maunder minimum, see

- Eddy, J. A., *Science*, **192**, 1189-1202 (1976),
- Eddy, J. A., *Solar Phys.*, **89**, 195-207 (1983),
- Ribes, J. C., and Nesme-Ribes, E., *Astron. Astrophys.*, **276**, 549-563 (1993).

and on cosmogenic radioisotopes:

- Beer, J., *Sp. Sci. Rev.*, **94**, 53-66 (2000).
- Usoskin, I.G., Solanki, S.K., & Kovaltsov, G.A., *Astron. Astrophys.*, **471**, 301-309 (2007).

On the effects of stochastic forcing on various dynamo models, start with:

- Choudhuri, A. R., *Astron. Astrophys.*, **253**, 277-285 (1992),
- Moss, D., Brandenburg, A., Tavakol, R., and Tuominen, I., *Astron. Astrophys.*, **265**, 843-849 (1992),
- Hoyng, P., *Astron. Astrophys.*, **272**, 321-339 (1993),
- Ossendrijver, M. A. J. H., and Hoyng, P., *Astron. Astrophys.*, **313**, 959-970 (1996),
- Charbonneau, P., and Dikpati, M., *Astrophys. J.*, **543**, 1027-1043 (2000),
- Mininni, P., and Gomez, D.O., *Astrophys. J.*, **573**, 454-463 (2002).

The following offer a good sample of the possible amplitude and parity modulation behaviors in nonlinear (sometimes non-kinematic) mean-field dynamo models:

- Brooke, J. M., Moss, D., and Phillips, A., *Astron. Astrophys.*, **395**, 1013-1022 (2002),
- Küker, M., Arlt, R., and Rüdiger, R., *Astron. Astrophys.*, **343**, 977-982 (1999),

- Sokoloff, D., and Nesme-Ribes, E., *Astron. Astrophys.*, **288**, 293-298 (1994),
 Tobias, S. M., *Astron. Astrophys.*, **322**, 1007-1017 (1997),
 Bushby, P.J., *Month. Not. R. Astron. Soc.*, **371**, 772 (2006),
 Rempel, M., *Astrophys. J.*, **647**, 662 (2006),
 Lopes, I., & Passos, D., *Solar Phys.*, **257**, 1 (2009).

On time delay and its consequences for Babcock-Leighton dynamo models, see:

- Durney, B.R., *Solar Phys.*, **196**, 421-426 (2000),
 Charbonneau, P., *Solar Phys.*, **199**, 385-404 (2001),
 Charbonneau, P., St-Jean-Leblanc, C., and Zacharias, P., *Astrophys. J.*, **619**, 613-622 (2005).

The following offers a few good entry points in the literature on intermittency in various types of dynamo models:

- Schmitt, D., Schüssler, M., and Ferriz-Mas, A., *Astron. Astrophys.*, **311**, L1-L4 (1996),
 Tworkowski, A., Tavakol, R., Brandenburg, A., Brooke, J.M., Moss, D., and Tuominen, I., *Month. Not. R. Astron. Soc.*, **296**, 287-295 (1998),
 Covas, E., and Tavakol, R., *Phys. Rev. E*, **60**, 5435-6645 (1999),
 Ossendrijver, M. A. J. H., *Astron. Astrophys.*, **359**, 364-372 (2000),
 Charbonneau, P., Blais-Laurier, G., and St-Jean-Leblanc, C., *Astrophys. J.*, **616**, L183-L186 (2004),
 Petrovay, K., *Astr. Nach.*, **328**, 777-780 (2007),
 Moss, D., Sokoloff, D., Usoskin, I., and Tutubalin, V., *Solar Phys.*, **250**, 221-234 (2008).

Cycle prediction is a topic that has generated a massive literature, which often stands closer to statistical black magic than physics. Two good recent review papers are:

- Hathaway, D.H., *Sp. Sci. Rev.*, **144**, 401-412 (2009).
 Petrovay, K., *Liv. Rev. Sol. Phys.*, **7** (2010),
<http://solarphysics.livingreviews.org/Articles/lrsp-2009-6/>

On dynamo-inspired precursor schemes, see

- Schatten, K.H., Scherrer, P.H., Svalsgaard, L., & Wilcox, J.M., *Geophys. Res. Lett.*, **5**, 411-414 (1978),
 Svalsgaard, L., Cliver, E.W., & Kamide, Y., *Geophys. Res. Lett.*, **32**, 021664 (2005),
 Charbonneau, P., & Barlet, G., *J. Atmos. Solar-Terrestrial Phys.*, **73**, 198-206 (2011).

On the use of dynamo models for cycle prediction, see

- Dikpati, M., DeToma, G., & Gilman, P.A., *Geophys. Res. Lett.*, **33**, L05102 (2006),
Choudhuri, A.R., Chatterjee, P., & Jiang, J., *Phys. Rev. Lett.*, **98**, 131103 (2007),
Cameron, R., & Schüssler, M., *Astrophys. J.*, **659**, 801-811 (2007),

as well as

- Bushby, P., & Tobias, S.M., *Astrophys. J.*, **661**, 1289-1296 (2007).

This last paper is an illuminating discussion of the fundamental limitations inherent in using nonlinear dynamo models for cycle amplitude forecasting. Keep in mind, however, that the two forecasting models discussed in §0.35.2 (1) are *not* operating in a chaotic regime, and (2) achieve their forecasting not just through direct forward integration, but also through continuous or episodic input of observational data. Finally, the following paper is a very interesting example of true data assimilation in a very simple albeit truly nonlinear dynamo model:

- Kitiashvili, I., & Kosovichev, A.J., *Astrophys. J. Lett.*, **688**, 49-52 (2008).

Chapter 5

Stellar dynamos

{pch_chap:dynstar}

*ELWOOD: It's 106 miles to Chicago, we've got a full tank of gas, half a pack of cigarettes, it's dark and we're wearing sunglasses.
JAKE: Hit it!*

*Dan Ackroyd and John Belushi
The Blues Brothers (1980)*

The problem —and the beauty— with the Sun is that it overwhelms us with data. Many of the intricacies we have busied ourselves with in the preceding chapter were directly motivated by the detailed observations and magnetic measurements made possible by the sun's astronomical proximity. The sun remains for sure an exemplar, but with other stars observational constraints are much more sparse, and theoretical considerations take on an enlarged role.

What have we learned in the preceding three chapters about dynamo action in electrically conducting fluids? At the most fundamental level, a top-three list could run as follows:

- We learned in chapter 2 that rotation, and especially differential rotation, is one very powerful mechanism allowing to build a large-scale magnetic field;
- We also learned in chapter 2 that flows with chaotic trajectories, such as arising from strongly turbulent convection, can act as dynamos;
- We learned in chapter 3 that in turbulent flows, the presence of rotation and stratification can break isotropy and reflectional symmetry, and in doing so generate a mean electromotive force that can produce large-scale magnetic fields.

So, offhand we are not in too bad a shape with regards to stellar dynamos. Stars certainly are stratified, and certainly rotate. Thermally-driven convec-

tion is also present across large-part of the HR diagram, but here we start to encounter complications that restrict the use of the “solar exemplar”.

Figure 66 illustrates, in schematic form, the internal structure of main-sequence stars, more specifically the presence or absence of convection zones. A G-star like the Sun has a thick outer convection zone, spanning the outer 30% in radius in the solar case. As one moves down to less massive stars, the relative thickness of the convective envelope increases until, somewhere around spectral type M5, stars become fully convective. Moving instead from the Sun to higher masses, the convective envelope becomes ever thinner, until somewhere around spectral type A0 it essentially vanishes. However, at around the same spectral type Hydrogen burning switches from the p - p chain to the CNO cycle, for which nuclear reaction rates are much more sensitively dependent on temperature. Core energy release becomes strongly depth-dependent, leading to a steep—and convectively unstable—temperature gradient. This produces a small convective core, which grows in size as one moves up to larger masses. In a “typical” B-star of solar metallicity, the convective core spans the inner 25% or so in radius of the star. In main-

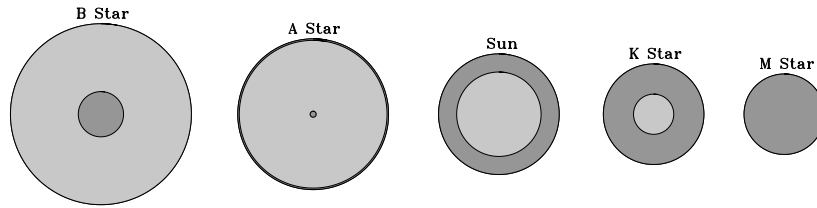


Fig. 66 fig:msstruct Schematic representation of the radiative/convective internal structure of main-sequence stars. The thickness of the outer convection zone for the A-star is here greatly exaggerated; drawn to scale it would be thinner than the black circle delineating the stellar surface on this drawing. Relative stellar sizes are also not to scale; a main-sequence M0 star has a radius some 12 times smaller than its B0 counterpart.

sequence O and B stars, the presence of a turbulent convective core combined with high rotation then makes dynamo action more than likely. However, as we shall see in §0.36 below, the challenge is actually to bring the magnetic field produced in the core to the surface.

Intermediate mass main-sequence stars rotate but lack a convective zone of substantial size. Interestingly, only the most slowly rotating of these stars show evidence of strong magnetic fields, for which the fossil hypothesis remains the favored explanation. This is reviewed in §0.37, along with some dynamo-based alternatives.

Until strong evidence to the contrary is brought to the fore, we are allowed to assume that late-type stars with a thick convective envelopes overlying a radiative core host a solar-type dynamo. This is buttressed by the observation

of solar-like cyclic activity in many such stars. It then becomes natural to look into the way(s) the various types of solar-cycle models considered in the preceding chapters can be “scaled” to other solar-type stars, of varying masses, rotation rates, etc. Some of these issues are discussed in §0.38 below.

With fully convective stars (§0.39), we encounter potential deviations from a solar-type dynamo mechanism; without a tachocline and radiative core to store and amplify toroidal flux ropes, the Babcock-Leighton mechanism becomes problematic. Mean-field models based on the turbulent α -effect remain viable, but the dynamo behavior becomes dependent on the presence and strength of differential rotation, about which we really don’t know very much in stars other than the sun.

Moving off the main-sequence, stellar magnetic field remain ubiquitous, but dynamo modelling is comparatively undeveloped. Observational evidence points towards turbulent dynamo action in pre-main-sequence stars, as briefly reviews in §0.40. On the other hand, the very strong magnetic fields observed in the many types of compact objects marking the endpoint of stellar evolution appear, for the most part, to be remnant of earlier evolutionary phases, rather than being produced by ongoing dynamo action (§0.41).

0.36 Early-type stars

{sec:mstardyn}

0.36.1 Mean-field models

{ssec:mfmsdyn}

We first consider dynamo action in massive stars, beginning with a few simple, representative solutions obtained in the framework of mean-field theory⁴⁷. Within the core, thermally-driven turbulent fluid motions are assumed to give rise to an α -effect and turbulent diffusivity, which both vanish for $r \gtrsim r_c$ (under the assumption that the radiative envelope is turbulence free). In the spirit of the other mean-field models discussed earlier, we consider kinematic dynamos with parametric profiles for α and η :

$$\alpha(r, \theta) = \frac{1}{2} \left[1 + \operatorname{erf} \left(\frac{r - r_c}{w} \right) \right] \operatorname{erf} \left(\frac{2r}{w} \right) \cos(\theta) , \quad (281) \quad \{\text{E1.15a}\}$$

$$\eta(r) = \eta_e + \frac{\eta_c - \eta_e}{2} \left[1 - \operatorname{erf} \left(\frac{r - r_c}{w} \right) \right] , \quad (282) \quad \{\text{E1.15b}\}$$

where $\operatorname{erf}(x)$ is the error function. Equations (281) once again represent “minimal” assumptions on the spatial dependency of the α -effect: it changes sign across the equator ($\theta = \pi/2$), vanishes at $r = 0$, rises to a maximum value

⁴⁷ The content of this section is based on the paper by P. Charbonneau & K.B. MacGregor, *ApJ*, **559**, 1094 (2001)

within the convective core, and falls again to zero for $r \gtrsim r_c$, the transition occurring across a spherical layer of thickness $\sim 2w$. All dynamo solutions discussed below are obtained as eigenvalue problems, as in §0.26.8, solving now the α^2 form or the axisymmetric kinematic mean-field dynamo equations. Remember that such linear solutions leave the absolute scale of the magnetic field unspecified. However, an interesting physical quantity accessible from linear models is the ratio of the surface field strength to the field strength in the dynamo region, here the convective core. In what follows we use towards this purpose the ratio (Σ) of the r.m.s. surface poloidal field to the r.m.s. poloidal field at the core-envelope interface r_c :

$$\{\text{E1.16b}\} \quad \Sigma = \left(\frac{R^2 \int |\nabla \times A|_{r=R}^2 \sin \theta d\theta}{r_c^2 \int |\nabla \times A|_{r=r_c}^2 \sin \theta d\theta} \right)^{1/2}. \quad (283)$$

Figure 67 shows a series of typical linear α^2 solution with increasing diffusivity contrasts between the core and envelope. Linear mean-field dynamo of the α^2 type with a time-independent scalar functional $\alpha(r)$ usually produce *steady* magnetic fields, i.e., the solution eigenvalue is purely real ($\omega = 0$ in eq. (243)). The solution plotted on Figure 67A is dipole-like (i.e., antisymmetric), and is the fastest growing solution for our model with constant η , at the adopted value for C_α ⁴⁸. The poloidal and toroidal magnetic components have comparable strengths, which is again typical of α^2 mean-field models with scalar α -effect. Here the growth rate of the eigenmode is about 20 yr in dimensional units, leaving no doubt that ample time is available to amplify a weak seed magnetic field in the core of a massive star. Note also on Figure 67 how the dynamo-generated magnetic field becomes trapped within and in the immediate vicinity of the convective core for even moderately large values of magnetic diffusivity contrast between core and envelope.

In the presence of significant differential rotation, core dynamo action can produce polarity reversals and wave-like propagation of the magnetic field, much like in the $\alpha\Omega$ solar cycle models considered earlier. Figure 68 illustrates a half-cycle of a representative $\alpha^2\Omega$ solution, constructed by imposing a radial gradient of angular velocity across a thin shear layer coinciding with the core-envelope interface:

$$\{\text{E3.1}\} \quad \Omega(r, \theta) = \Omega_c + \frac{\Omega_e - \Omega_c}{2} \left[1 + \operatorname{erf} \left(\frac{r - r_c}{w} \right) \right]. \quad (284)$$

⁴⁸ The α^2 form of the mean-field dynamo equations also admits growing solutions than are non-axisymmetric even though the α -effect profile exhibits axisymmetry with respect to the rotation axis. Growth rates for non-axisymmetric modes are often comparable to those of their axisymmetric counterparts. Motivated largely by the challenge posed by planetary magnetic fields, α^2 models can and have been constructed where non-axisymmetric modes are the fastest growing, and dominate in the moderately supercritical nonlinear regime.

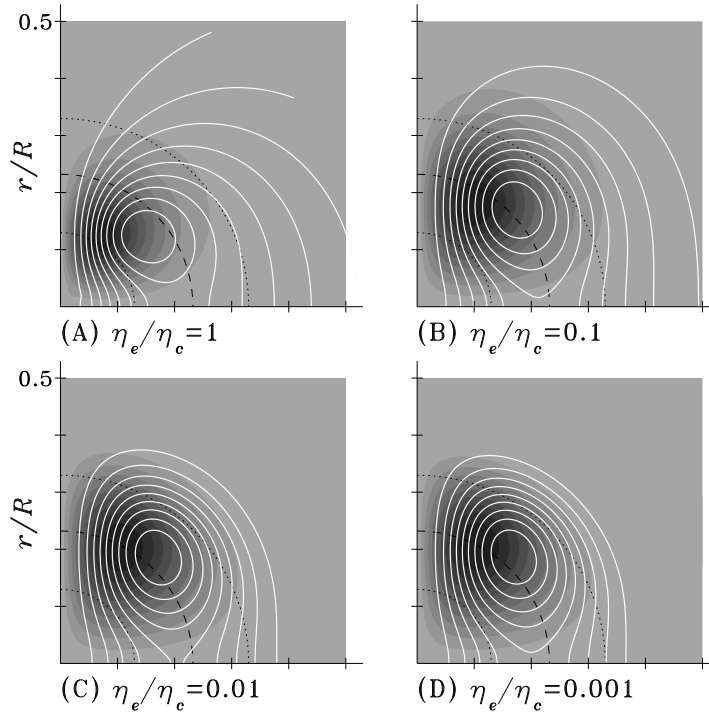


Fig. 67 Four antisymmetric steady α^2 dynamo solutions, computed using varying magnetic diffusivity ratios between the core and envelope. The solutions are plotted in a meridional quadrant, with the symmetry axis coinciding with the left quadrant boundary. Poloidal fieldlines are plotted superimposed on a gray scale representation for the toroidal field (light to dark is weaker to stronger field). The dashed line marks the core-envelope interface depth r_c , and the two dotted lines indicates the depths $r_c \pm w$ corresponding to the width of the transition layer between core and envelope. These solutions have a surface-to-core magnetic field strength ratio $\Sigma \simeq 10^{-2}$ at $\eta_e/\eta_c = 1$, down already to 3×10^{-4} at $\eta_e/\eta_c 10^{-1}$ and falling below 10^{-8} for $\eta_e/\eta_c \lesssim 10^{-2}$. (Fig. 67)

The magnetic field distribution is shown at five distinct phases, at constant intervals of $\Delta\varphi = \pi/4$, in a format identical to that of Fig. 67 for each panel (note in particular that the eigenmodes are again plotted only in the inner half of the star). At a given phase the solutions bear some resemblance to the α^2 solutions of Fig. 67C, in that the magnetic field is again trapped in the interior. As before, the toroidal field is concentrated near the core-envelope interface, and in fact here peaks slightly outside $r = r_c$ (dashed circular arc).

The availability of an additional energy source in the toroidal component of the dynamo equations leads to solutions where the toroidal field strength in general exceeds that of the poloidal field, scaling roughly as the ratio C_Ω/C_α in the limit $C_\Omega \gg C_\alpha$. For a given diffusivity ratio η_e/η_c , oscillatory $\alpha^2\Omega$ solutions have a smaller surface-to-core field strength ratio Σ than α^2 models,

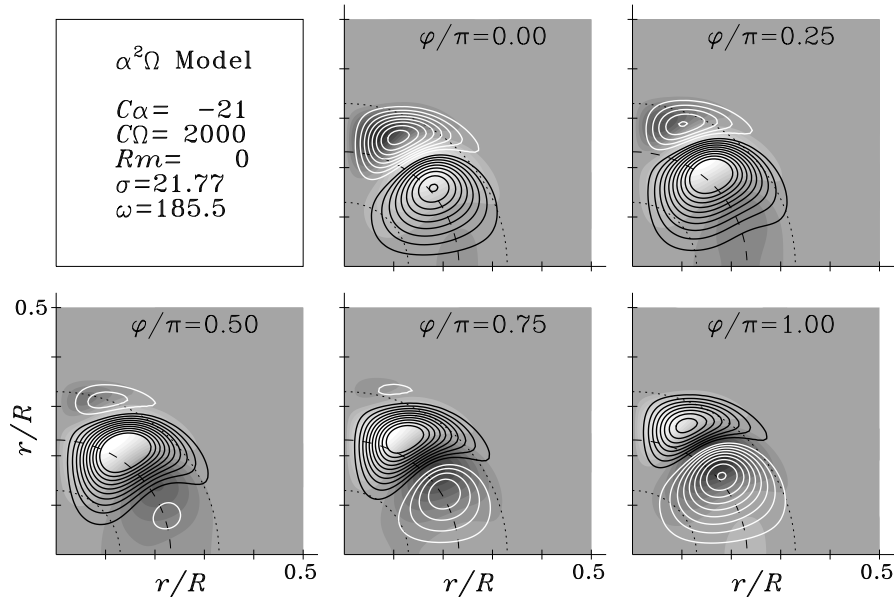


Fig. 68 A representative $\alpha^2\Omega$ solution. As this is an oscillatory solution, the eigenfunction is plotted at five equally spaced phase intervals ($\Delta\varphi = \pi/4$), covering half an oscillation cycle. The format in each panel is similar to Fig. 67. White (black) lines indicate fieldlines oriented in a clockwise (counterclockwise) direction. Note the wave-like propagation of the magnetic field from low to high latitudes. This symmetric solution has $C_\alpha = -21$, $C_\Omega = 2000$, $w/R = 0.1$, $\eta_e/\eta_c = 10^{-2}$, and is characterized by a growth rate $\sigma = 21.8\tau^{-1}$ and frequency $\omega = 186\tau^{-1}$. For $\eta_c = 10^{13}\text{ cm}^2\text{ s}^{-1}$, this corresponds to a dynamo period of about 7 yr, quite short compared to any other relevant timescales. ^{Fig 3}

a direct consequence of the oscillatory nature of the field, which restricts the radial extent of the eigenfunction above the core-envelope interface to a distance comparable to the electromagnetic skin depth, which is very much smaller than the stellar radius for $\eta_e/\eta_c \ll 1$.

{ssec:coresims}

0.36.2 Numerical simulations of core dynamo action

It is interesting to compare and contrast core dynamo action, as modeled via mean-field electrodynamics, to what is produced by three-dimensional MHD simulations⁴⁹. Such simulations do yield vigorous core dynamo action, with the magnetic energy approaching equipartition with the turbulent fluid motions. However, most of the magnetic energy is contained in small spatial scales, with the axisymmetric large-scale component accounting for only a

⁴⁹ The content of this section is based primarily on the paper by A.S. Brun, M.K. Browning, & J. Toomre, ApJ, 629, 461 (2005).

few percent of the total magnetic energy. The simulations generate a highly time-variable differential rotation that contributes significantly to the induction of a toroidal component by shearing of the poloidal fields. This is most pronounced in the vicinity of the core-envelope boundary, where a persistent system of magnetic field bands approximately aligned in the azimuthal direction are produced.

These simulations could be said to behave like $\alpha^2\Omega$ mean-field dynamo, but the analogy is only superficial because significant differences exist, most notably perhaps the absence of a well-defined, persistent mean-field-aligned turbulent electromotive force. Except in the innermost portion of the convective core, the mean kinetic helicity is negative in the Northern hemisphere, but in contrast the mean magnetic helicity does not show a well-defined, persistent hemispheric pattern, again a departure from mean-field expectations.

One important similarity with the mean-field models considered in §0.36.1 is the trapping of the magnetic field within or in the immediate vicinity of the convective core. This is shown on Figure 69, which depicts two temporal and azimuthal averages at different epochs in a representative simulation. The weak axisymmetric toroidal field present in the inner portion of the radiative envelope is produced primarily by the shearing effect of the differential rotation, which is imprinted from the core to the lower envelope by the relatively high viscous forces characterizing this simulation.

0.36.3 Getting the magnetic field to the surface

{ssec:totop}

Whatever the mode of core dynamo action, a universal feature is the “trapping” of the magnetic field within the core and in the lower part of the radiative envelope, a direct consequence of the difficulty experienced by a magnetic field to diffusively penetrate a good electrical conductor. This is long-recognized property of stellar core dynamos, and represents a rather formidable obstacle to be bypassed if the magnetic fields generated by dynamo action in the convective core are to become observable at the stellar surface. For O and B main-sequence stars, estimates for the diffusion time yield values largely in excess of the main-sequence lifetime. Introducing thermally-driven meridional circulation in the radiative envelope, expected to be a significant internal flow in rapidly rotating stars, does accelerate the transport of the deep field to the surface, but also impedes dynamo action. Another possibility is that the dynamo-generated magnetic field manages to produce toroidal flux ropes that subsequently rise buoyantly to the surface. The analogy with the sun becomes even more compelling if a rotational shear layer does exist at the boundary between the inner convective core and outer radiative envelope. However, and unlike in the solar case, here the toroidal flux ropes are rising through a *stably* stratified environment, and so lose their buoyant force as they rise, because they cool faster than the surrounding stratification. Calculations performed in the thin flux tube approximation

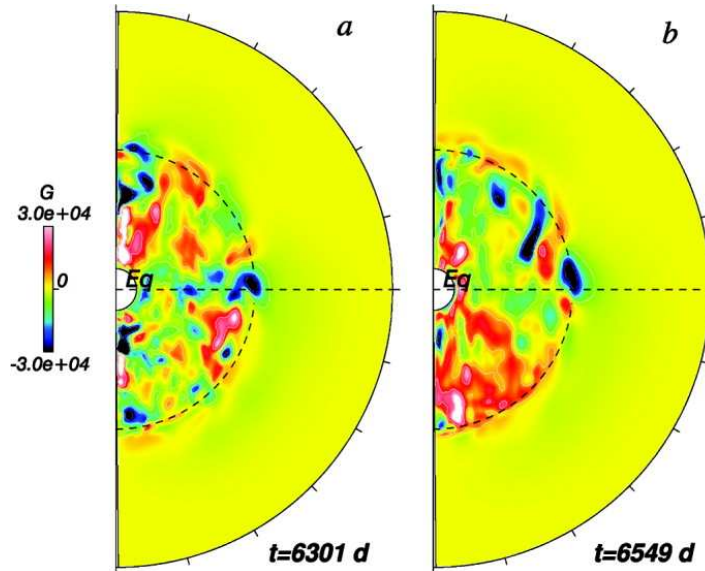


Fig. 69 Temporal+Azimuthal average of the toroidal magnetic field in a 3D MHD numerical simulation of dynamo action in the core of a $2 M_{\odot}$ early A-star. The magnetic field reaches strengths in the tens of kG, evolves rapidly, and is structured on a broad range of spatial scales, but remains confined to the convective core (dashed circular arc; the simulation domain only covers the inner 30% in radius of the star). Figure taken from Brun *et al.* 2005, ApJ, 629, 461 (Figure 18, p. 478). fig-core-sim

suggest that such toroidal flux ropes, assuming they do form, could rise perhaps halfway across the radiative envelope, but are unlikely to make it all the way to the surface through buoyancy alone. References listed in the bibliography should provide helpful entry points into the literature to those interested in further pursuing this aspect of massive star magnetism.

{ssec:altcore} 0.36.4 Alternative to core dynamo action

Dynamo-based explanations for the magnetic fields of early-type main-sequence stars certainly exist. One intriguing possibility that clearly requires serious modelling is that dynamo action in the outer layers of massive stars could take place in convection zones associated with a peak in iron opacities. Recent years have also witnessed renewed interest in the possibility that dynamo action could take place in the radiative envelope of intermediate- and high-mass main-sequence stars, through turbulence associated with one or more global instabilities of the magnetic field. This idea has attracted attention outside of the dynamo circles because the associated turbulent transport would also cause enhanced chemical mixing, known to be required to properly fit evolutionary tracks of massive stars, but whose origin remains mysterious.

Introduction of simple parametrizations for the associated chemical and angular momentum mixing in models of evolving massive stars has shown that it is difficult to maintain sufficient differential rotation for the instability to operate, while keeping chemical mixing at the required level. The interested reader will find entry point in this vast literature in the bibliography at the end of this chapter.

This instability-driven dynamo mechanism probably cannot explain the strong magnetic fields observed in the very slowly rotating Ap/Bp stars (more on these presently), again because significant internal differential rotation is unlikely in the radiative envelope of these stars.

0.37 A-type stars

{sec:Adyn}

0.37.1 Observational overview

Extant observations suggest a true dichotomy with regards to stellar magnetism in intermediate-mass stars: most A and B stars (around 95%) on or near the main-sequence have no measurable (as yet) magnetic field, but nearly all those who do combine strong, large-scale magnetic fields, steady on decadal timescales at least, with slow rotation and pronounced photospheric abundance anomalies. As we will see later in this course, the presence of a strong, large-scale photospheric magnetic field (of whatever origin) favors angular momentum loss, and therefore slow rotation; and a strong magnetic field and low rotation favor atmospheric stability, giving full leeway for chemical separation to operate and alter photospheric abundances.

In the most slowly rotating, strongly magnetized Ap stars, the mean surface magnetic field strength (“mean” in the sense of being averaged over the stellar surface) can be detected by Zeeman splitting, as in sunspots. Figure 70 below shows a striking example of such splitting. In more rapidly rotating stars magnetic Doppler imaging become a possibility; this relies on the varying shapes of spectral lines formed as magnetic structures cross the visible part of the stellar disk. “Imaging” remains indirect, in the sense that the stellar surface is of course not resolved spatially, but the availability of many spectral lines, with some appropriate regularization scheme, allows this inverse problem to be solved. Figure 71 shows a particularly well-studied exemplar, namely the chemically peculiar star 49Cam. The field strength is high, the magnetic topology quite complex, with the idea of a strongly inclined dipole, historically the common interpretation for Ap stars magnetic fields, being a rather rough approximation here.

It is an intriguing fact that the few chemically-normal, (relatively) rapidly rotating early-type stars on which magnetic fields have been detected all sit in the early-B range of spectral types and belong to the β Cep sub-class (and include the prototype star β Cep itself). However, indirect evidence for

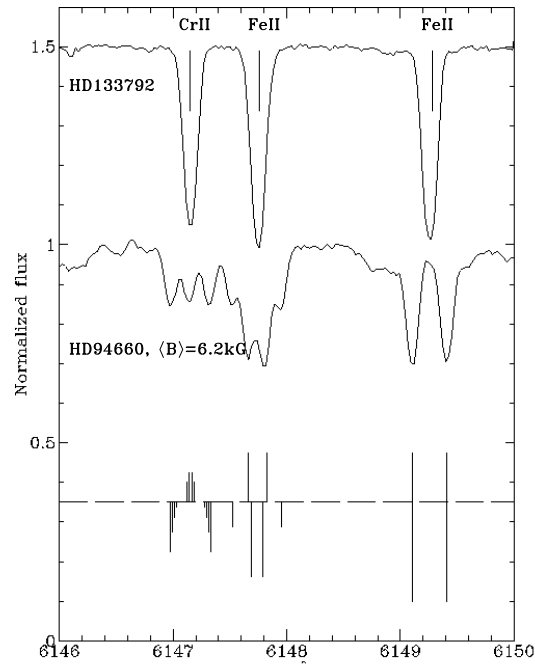


Fig. 70 ^{fig:modB} Zeeman splitting of magnetically-sensitive absorption line in the spectrum of the Ap star HD94660. The inferred mean field strength for this star is 0.62T. The top trace is that of a typical unmagnetized star of similar spectral type. The horizontal axis is the wavelength, measured in Å. The bottom trace shows the multiplet structure of the three spectral lines. Figure reproduced from the Mathys *et al.* paper cited in the bibliography, with a few labels added.

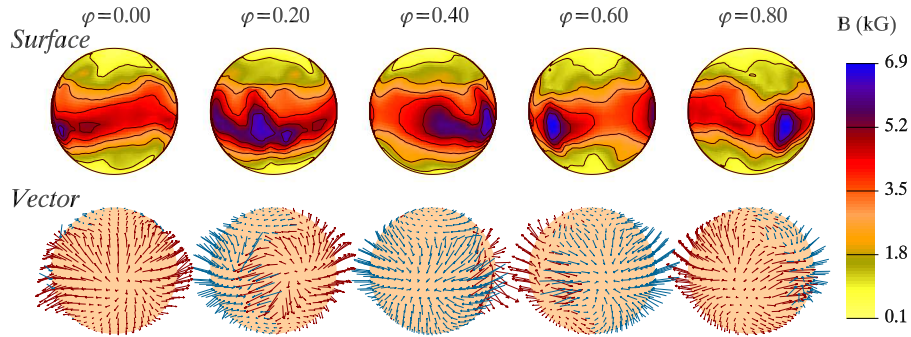


Fig. 71 ^{fig:49cam} The surface magnetic field on the Ap star 49Cam, as reconstructed for various rotational phases (φ) by magnetic Doppler imaging. The top row shows the net field strength, and the bottom row the orientation of the surface magnetic field vector. Plot courtesy of J. Silvester and G. Wade, RMC/Kingston.

photospheric magnetism in O and B star has been accumulating steadily, be it as emission of hard radiation above and beyond what shock dissipation can provide, channelling of stellar winds, and spectral variability. Ongoing spectropolarimetric campaigns targeting massive stars are now providing more and more data for theoreticians/modellers to chew on in upcoming years.

0.37.2 The fossil field hypothesis

Stars with spectral types ranging from late-B to early-F stand out as the least likely to support dynamo action, because they lack a convective region of substantial size. This squares well with various lines of observations; in particular, main-sequence A-stars are amongst the most “magnetically quiet” stars in the HR diagram. As we just discussed, a subset of late-B and A stars, namely the slowly-rotating, chemically peculiar Ap/Bp stars, do show strong magnetic fields, but those show no sign of anything even mildly analogous to solar activity. The single pattern of temporal evolution noted is a decrease, by factors of 2-3, in the overall strength of the surface field, most prominent in the early stages of main-sequence evolution. This seems compatible with the idea of diffusive decay of residual higher-degree eigenmodes, and slow decreases associated with flux conservation as the stars slowly expand in the course of their main-sequence evolution (cf. §0.18).

0.37.3 Dynamical stability of large-scale magnetic fields

The study of the purely resistive decay of large-scale magnetic field in stellar interiors carried out in §0.18 precluded, by its very design, the development of flows propelled by potential hydromagnetic instabilities. Investigations into the latter have shown that under typical stellar interior conditions, large-scale magnetic fields embedded in stably-stratified radiative interiors are indeed susceptible to the development of instabilities with growth rates much smaller than any relevant evolutionary timescales. Even simple field configurations, such as low-order multipole purely toroidal or purely poloidal fields are found to be unstable, with rotation possibly providing a stabilizing influence at high rotation rates (see references in the bibliography at the end of this chapter). Although these (linear) stability analyses rely on a number of strong simplifying assumptions, they lead to a picture whereby the most likely stable global configurations are magnetic fields comprised of a mixture of large-scale poloidal and toroidal component with comparable strengths. Remarkably, this has recently been confirmed by full MHD simulations. Such configurations would establish themselves on very short timescales, after which they would undergo resistive decay on the magnetic diffusion timescale. Short-lived unstable phase early in their evolution notwithstanding, this overall picture remains generally consistent with the fossil field *Ansatz* for Ap/Bp stars.

0.37.4 The transition to solar-like dynamo activity

On the main-sequence, as one move down from late-A to late-F spectral types, solar-type surface convection sets in, with the convection zone rapidly gaining in depth as below spectral type F5. How and when solar-type dynamo action sets in is a relatively unexplored question that clearly deserves further attention, from both the observational and modelling standpoints.

{sec:solarstars}

0.38 Solar-type stars

0.38.1 Observational overview

The disk of solar-type stars other than the sun cannot be spatially resolved, and so direct observation of starspots is not possible, although rotational modulations of the luminosity associated with starspot darkening most certainly is. Direct measurements of magnetic polarisation of starlight is difficult as well, unless the field has a strong large-scale component, otherwise the polarisation associated with regions of opposite polarities —e.g., starspot pairs— cancel out when integrated over the solar disk. Most evidence for the presence of magnetic fields on such stars is thus indirect, yet extremely compelling, as it covers a wide range of phenomena visible on the sun, such as spectral line variability, rotational modulation of luminosity due to the passage of large starspots, flares, radio bursts, and variability in magnetically-sensitive spectral lines on a wide range of timescales. Such indirect observational evidence for magnetic fields has been found on *every* late-type main-sequence star observed with sufficient sensitivity. The sun is, indeed, a typical solar-type stars!

One magnetic field-related stellar observable that is particularly noteworthy is the emission in the cores of the H and K lines of CaII (396.8nm and 393.4nm, respectively). On the Sun, this emission is known to be associated with non-radiative heating of the upper atmosphere, and is known to scale well with the local photospheric magnetic flux. Starting back in 1968 at Mt Wilson Observatory, Olin C. Wilson (1909-94) began measuring the CaII H+K flux in a sample of solar-type stars, a laborious task that was later picked up by a brave group of undeterrable associates and followers, whose collective labor has produced a 40 year long archive of CaII emission time series for no less than 111 stars in the spectral type range F2-M2, on or near the main-sequence.

Figure 72 shows a few sample time series of the so-called Calcium index S , measuring the ratio of core emission intensity in the H and K lines to that of the neighbouring continuum. Some stars show solar-like cycles, others have irregular CaII emission, some show long term trends and others can only be dubbed “flatliners”. Note that the mere presence of detectable CaH+K emis-

sion indicates magnetic activity; the absence of detectable temporal variations in flatliner stars simply means that they stars lack a solar-like well-organized magnetic cycle. From the point of view of dynamo theory and modelling, the

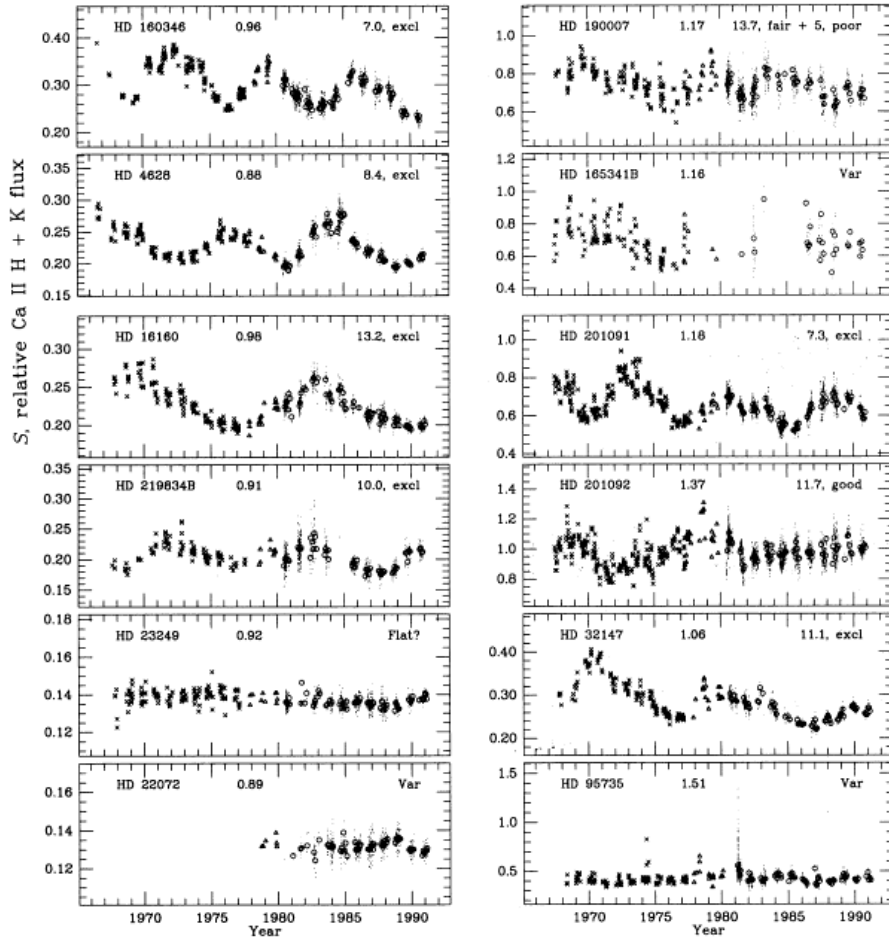


Fig. 72 fig:stellcyc Calcium emission index in a small subsample of the Mt Wilson dataset, showing the variety of CaII emission patterns: cycles, non-cyclic irregular emission, long term trend, and constant emission. On such plots, the sun would have a mean emission level $\langle S_{\odot} \rangle = 0.179$, with a min/max range of about 0.04. Figure cropped from a much larger Figure in Baliunas *et al.* 1995, ApJ,438, 269 [Figure 1g].

following two empirical facts are particularly noteworthy:

1. Magnetic activity, as measured e.g. by the level of CaH+K emission, generally increases with increasing rotation rate (decreasing P_{rot}).
2. Stellar cycle periods increase with increasing rotation periods P_{rot} .

0.38.2 Empirical stellar activity relationships

Among stars showing cycles in CaH+K, at a given spectral type the relationship between cycle period and rotation rate is well represented by a power-law of the form $P_{\text{cyc}} \propto P_{\text{rot}}^n$, with n varying between 0.75 and 1.75 depending on spectral type. Interestingly, all these data can be described reasonably well by a single power-law fit to the ratio of the rotation period to the convective turnover time, known as the Rossby Number (Ro; the inverse Rossby number is often referred to as the Coriolis number):

$$P_{\text{cyc}} \propto \left(\frac{P_{\text{rot}}}{\tau_c} \right)^n, \quad n = 1.25 \quad (285)$$

Recall from the discussion in §0.26.2 that this ratio is supposed to measure the efficiency of the Coriolis force in breaking the mirror-symmetry of convective turbulence, and thus producing a non-zero α -effect. Recall also that the larger the dynamo number, the more magnetic energy mean-field models can produce (viz. Fig. 36). So, in a rough qualitative sense, observations seem to fit our (naive) theoretical expectations.

In reality, there are of course significant complications to this highly simplified picture. For example, coronal (X-Ray) and chromospheric (Ca H+K) emission is observed to saturate as Ro falls below about 0.1, and even decreases a bit beyond $\text{Ro} \sim 10^{-2}$, although it not clear whether this reflects a saturation of the emission mechanism, of the surface filling factor of magnetized structured, or of magnetic field generation by the large-scale dynamo. The P_{rot} vs Ro relationship becomes a lot tighter if stars for which reliable cycle periods are known are first divided into “active” and “inactive” subgroups on the basis of their overall level of Ca emission. Solar-like cyclic activity is by no means the rule among solar-type stars, with only about 60% of stars showing well-defined cycles, 25% showing aperiodic variations, and the remaining 15% being “flatliners” with low level, constant chromospheric emission. Indeed, the Sun has “twins”, i.e., main-sequence stars of closely similar surface temperature, gravity and rotation rate, which do not show any variability in chromospheric emission. An intriguing possibility is that these stars just happened to have been caught in a Maunder Minimum-like phase of suppressed cyclic activity.

0.38.3 Solar and stellar spin-down

Stellar observations indicate that there is evidently a lot more to dynamo action than just rotation, nonetheless the latter is clearly a key factor. For this reason, in any attempt to secure a coherent picture of dynamo action in solar-type stars, an important global feedback mechanism of dynamo action

must first be considered: angular momentum loss, and its effect on stellar rotation rates.

Although the existence of systematic differences between the average rotation rates of early- and late-type stars was known already for nearly a hundred years, observational evidence for main-sequence spin-down of solar-type stars was established much later. Figure 73 below is a reproduction of a diagram put together by Robert Kraft in 1967, showing the distribution in a HR diagram of projected equatorial rotational velocities ($v \sin i$) measured in a sample of field stars. As one runs down the main sequence, there occurs a sharp drop in $v \sin i$ starting around spectral type F5. Slow rotation is the rule on the cool side of this so-called rotational dividing line, while on the hot side rapid rotation is common. Kraft went on to show that under the assumption of solid-body rotation, in the interval $1.5 \lesssim M/M_{\odot} \lesssim 20$ observed rotation rates are consistent with a power-law dependence between stellar angular momentum (J) and mass (M) of the form $J \propto M^{1.57}$, which abruptly breaks down below F5.

The decrease in the moment of inertia of stars associated with their contraction towards the main-sequence can easily account for ZAMS equatorial rotational velocities of a few hundreds of kilometers per second. As was already understood then, the anomaly in Kraft's diagram lay in fact with the slowly-rotating low-mass stars. A most spectacular illustration that this is due to main-sequence *spin-down* and is associated with magnetism was provided in a short, now classical 1972 paper by Andrew Skumanich. Figure 74, reproduced from this paper, illustrates the simultaneous and gradual decrease of both the average rotation rates and magnetic activity for late-type stars—as measured by emission in the core of the Ca H and K lines—in a few open clusters of known ages. Later observations focusing on young clusters such as α Persei and the Pleiades have revealed that main-sequence spin-down for late-type stars is very swift, with the bulk of it completed in the first few 100 Myr after arrival on the ZAMS.

The key in explaining main-sequence spin-down is the realisation that stars with hot coronae lose mass through thermally-driven winds, and that the presence of a coronal magnetic field—ultimately of dynamo origin—turns out to greatly enhance the loss of angular momentum in the wind. We first examine this issue, using a geometrically simple but dynamically self-consistent MHD wind solution known as the Weber-Davis model.

Working under ideal MHD, we consider steady ($\partial/\partial t = 0$) spherically symmetric ($\partial/\partial \theta = \partial/\partial \phi = 0$) outflow from a star rotating at angular velocity Ω and characterized by a known surface radial component of the magnetic field B_{r0} . The coronal base temperature $T(r_0) \equiv T_0$ is considered given, and the energy equation is summarily dealt with by assuming a polytropic relationship between pressure and density. Outflow solutions are sought only in the equatorial plane, where we also set $B_{\theta} = 0$. This may smell of monopolar magnetic fields, but this is actually a fair representation of the interplanetary magnetic field during solar minimum conditions. Under these assumptions,

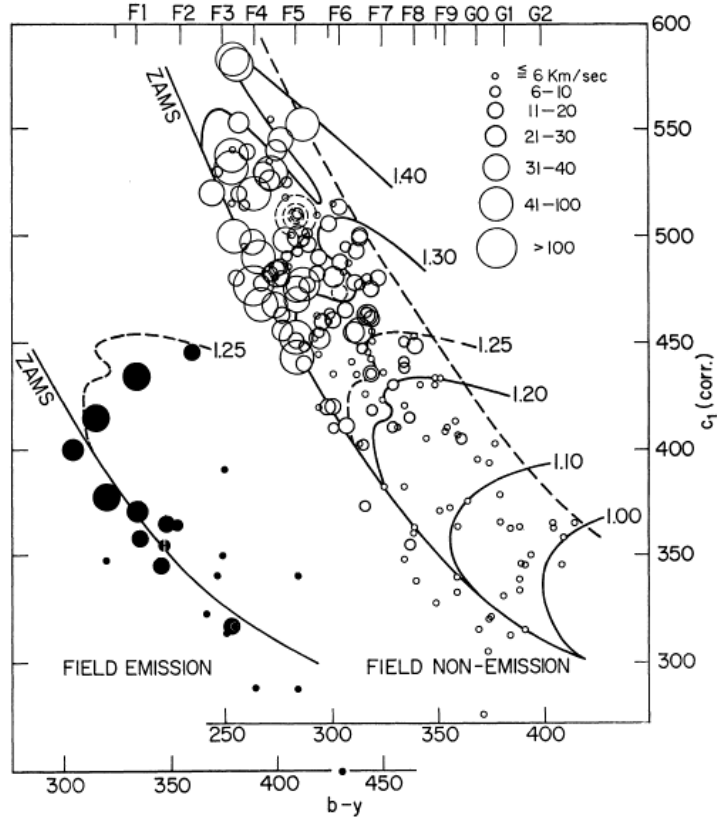


Fig. 73 fig:Kraft Distribution of projected rotational velocities ($v \sin i$) for main-sequence stars, plotted in an observational HR diagram. Luminosity increases vertically upwards, and effective temperature horizontally leftward. Astronomical spectral types are listed along the upper axis. Solid lines are stellar evolutionary tracks, labeled according to mass in solar units. These tracks, particularly for $M/M_{\odot} \gtrsim 1.2$, are now somewhat obsolete. Diagram reproduced from Kraft, R. 1967, ApJ, 150, 551 (Figure 1, p. 558).

mass conservation and the $\nabla \cdot \mathbf{B} = 0$ constraint yield two conservation statements for the mass and magnetic flux across a spherical surface:

$$\{\text{eq:WD1}\} \quad \frac{1}{r^2} \frac{\partial}{\partial r} (r^2 \rho u_r) = 0, \quad \rightarrow \quad r^2 \rho u_r = C_1, \quad (286)$$

$$\{\text{eq:WD4}\} \quad \frac{1}{r^2} \frac{\partial}{\partial r} (r^2 B_r) = 0 \quad \rightarrow \quad r^2 B_r = C_2, \quad (287)$$

where C_1 and C_2 are integration constants, corresponding respectively to the mass and magnetic flux carried by the wind. The ϕ -component of the

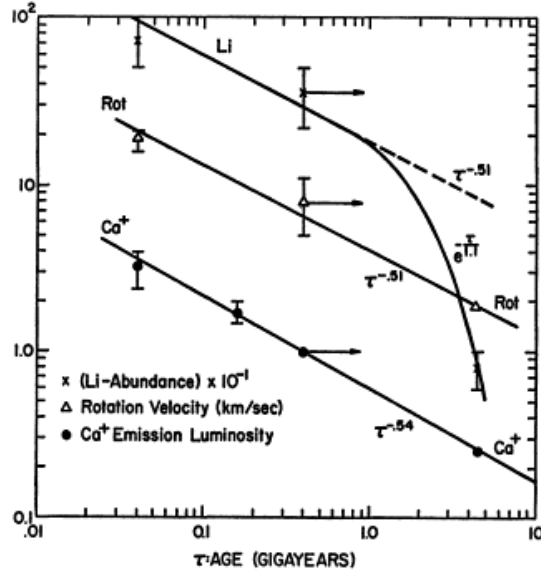


Fig. 74 ^{fig:sku} Main-sequence temporal evolution of rotation rates, Calcium emission and Lithium abundance in solar-type stars. Diagram reproduced from Skumanich, A. 1972, ApJ, 171, 565 (Figure 1, p. 566).

induction equation is also readily integrated to yield:

$$\{eq:WD3\} \quad \frac{1}{r} \frac{\partial}{\partial r} (ru_r B_\phi - ru_\phi B_r) = 0, \quad \rightarrow \quad r(u_r B_\phi - u_\phi B_r) = C_3. \quad (288)$$

To evaluate the integration constant C_3 we transform to a reference frame co-rotating with the star, i.e., $u_\phi \rightarrow u'_\phi + \Omega r$, where the prime indicates evaluation in the co-rotating frame. Note that this (non-relativistic) transformation leaves the radial components of \mathbf{u} and \mathbf{B} unaffected. In that frame \mathbf{B} is stationary, and since we are working under the flux-freezing approximation \mathbf{u} and \mathbf{B} must be parallel: $u'_r/u'_\phi = B'_r/B'_\phi$. Since $B_r = B'_r$, eq. (288) yields $C_3 = -\Omega r^2 B_r$, so that

$$B_\phi = \frac{B_r}{u_r} (u_\phi - \Omega r). \quad (289) \quad \{eq:WD8\}$$

Now, under the geometry and flow symmetry considered here, the ϕ -components of the momentum equation can be brought to the form:

$$\frac{\partial}{\partial r} (ru_\phi) = \frac{B_r}{\mu_0 \rho u_r} \frac{\partial}{\partial r} (rB_\phi); \quad (290) \quad \{eq:WD9\}$$

but in view of eqs. (286) and (287), we have $B_r/\mu_0 \rho u_r = C_2/\mu_0 C_1$, i.e., a constant! Consequently, eq. (290) integrates immediately to

$$ru_\phi - \frac{rB_\phi B_r}{\mu_0 \rho u_r} = L, \quad (291) \quad \{\text{eq:WD10}\}$$

where L is yet another integration constant. It has a well-defined physical meaning, as it corresponds to the total specific angular momentum carried away by the wind, which is made up of two contributions: the specific angular momentum of the expanding fluid (first term on LHS), and the torque density associated with magnetic tension. Using eq. (289) to substitute for B_ϕ , and expressing the magnetic field components in terms of the corresponding Alfvén velocity components (§0.10):

$$\{\text{eq:WD11}\} \quad a_r = \frac{B_r}{\sqrt{\mu_0 \rho}}, \quad a_\phi = \frac{B_\phi}{\sqrt{\mu_0 \rho}}, \quad (292)$$

produces, after some straightforward algebra:

$$\{\text{eq:WD13}\} \quad u_\phi = \Omega r \frac{(u_r^2 L / \Omega r^2) - a_r^2}{u_r^2 - a_r^2}. \quad (293)$$

The denominator of this expression vanishes if the radial flow velocity ever becomes equal to the radial Alfvén speed, unless the numerator also happens to vanish. Regularity of the solution through this critical point then requires that we set

$$\{\text{eq:WD14}\} \quad \boxed{L = \Omega r_A^2}, \quad (294)$$

where r_A is the *Alfvén radius*, defining the spherical shell where $u_r = a_r$. Now, remember that L is the total angular momentum carried away by the wind, *including* the torque density provided by magnetic tension. Equation (294) states that this is equal to the angular momentum that would be carried away by an unmagnetized wind flowing strictly radially, and co-rotating with the solar/stellar surface out to radius r_A .

Equation (294) holds only in the equatorial plane, where the WD solution is computed. The WD model can be “stretched” to the whole sphere by assuming that a whole spherical shell is co-rotating out to r_A ; this means replacing eq. (287) by:

$$\{\text{eq:amloss3}\} \quad L_{\text{sph}} = \frac{2}{3} \Omega r_A^2, \quad (295)$$

where the factor $2/3$ arises from the moment of inertia integral. The angular momentum loss rate then follows directly from multiplication by the mass loss rate:

$$\{\text{eq:lossrate1}\} \quad \frac{dJ}{dt} = \dot{M} \times L_{\text{sph}} = -4\pi \rho_A r_A^2 u_{rA} \left(\frac{2}{3} \Omega r_A^2 \right). \quad (296)$$

At the Alfvén radius we have $u_{rA} = a_{rA}$, with $B_{rA}^2 = 4\pi\rho_A a_{rA}^2$. Moreover, conservation of magnetic flux implies $r_0^2 B_{r0} = r_A^2 B_{rA}$. Putting all this into eq. (296) leads to

$$\frac{dJ}{dt} = -\frac{2}{3} B_{r0}^2 r_0^4 \Omega a_{rA}^{-1}. \quad (297) \quad \{\text{eq:lossrate2}\}$$

Knowing the stellar moment of inertia I and assuming rigid rotation throughout the interior, the spin-down timescale is readily calculated:

$$\tau_{\text{sp}} = I\Omega \left(\frac{dJ}{dt} \right)^{-1}. \quad (298) \quad \{\text{eq:spd2}\}$$

Now, for rotating magnetized winds that are mostly thermally driven, a_{rA} is of the order of the sound speed ($c_s = \sqrt{kT/\mu m_p} \sim 10^5 \text{ m s}^{-1}$ for a coronal temperature of $\sim 10^6 \text{ K}$) to within a factor of two or so. If the coronal temperature is held fixed, this means that the angular momentum loss rate is only a function of the rotation rate and surface magnetic field strength. We encountered in earlier chapters various lines of argument indicating that the dynamo-generated magnetic field strength should increase with increasing rotation rate, an expectation also buttressed by observations of chromospheric activity in solar-type stars of varying rotation rates. If one assumes $B_{r0} \propto \Omega$, and for a fixed moment of inertia on the main-sequence (a very good approximation, for a change...), eq. (296) then lead to

$$\frac{d\Omega}{dt} \propto -\Omega^3. \quad (299) \quad \{\text{eq:sku1}\}$$

This already indicates that faster rotating stars spin down a lot faster than their more slowly rotating cousins, which provides a natural explanation for the convergence of rotation rates observed at a given spectral types when looking at stellar rotation in progressively older clusters. Now, eq. (299) readily integrates to.

$$\frac{1}{\Omega^2(t)} - \frac{1}{\Omega^2(t_0)} \propto t - t_0, \quad (300) \quad \{\text{eq:sku2}\}$$

where t_0 is the time of arrival on the ZAMS (or shortly thereabouts). In the asymptotic limit $t \gg t_0$, $\Omega \ll \Omega(t_0)$, this becomes

$$\Omega(t) \propto t^{-1/2}, \quad (301) \quad \{\text{eq:sku3}\}$$

which, how about that, is precisely the power-law relationship inferred observationally by Skumanich (cf. Fig. 74). For the sun, with $I \simeq 10^{54} \text{ CGS}$, $\Omega = 2.6 \times 10^{-6} \text{ rad s}^{-1}$ and $B_0 \sim 2 \times 10^{-4} \text{ T}$ one finds a leisurely spin-down timescale of about 5 Gyr; but in a “young sun” with a rotation period of 2 day and $B_0 = 3 \times 10^{-3} \text{ T}$, this drops to a few 10^7 yr , indicating that rapidly

rotating young solar-type stars spin down swiftly after arriving on the ZAMS.

0.38.4 Modelling dynamo action in solar-type stars

The above discussion indicates that one could expect dynamo action to be far more vigorous in young, rapidly rotating solar-type stars, and the good fit of our spin-down model prediction with Skumanich's $t^{-1/2}$ relationship even suggests a linear increase of magnetic field strength with rotation rate, at least up to ~ 10 times the present solar rotation if coronal X-Ray emission can be taken as proxy of dynamo efficiency. Can such trends be convincingly recovered from the various solar dynamo models introduced in chapter 3? In practice, we are facing a number of difficulties in carrying out such an "extrapolation" to stars other than the sun, with convection zones of greater or lesser depths, differing luminosity, and a range of rotation rates. At the very least we need to be able to specify:

1. How the form and magnitude of differential rotation and meridional circulation change with rotation rate and luminosity, the latter determining the magnitude of convective velocities, and thus the magnitude of the turbulent Reynolds stresses powering the large-scale flows important for dynamo action;
2. How the α -effect and turbulent diffusivities vary in stars with different rotation rates and convection zone properties;
3. How the process of sunspot formation (essential in Babcock-Leighton models) vary with varying convection zone depth, rotation, etc.

It is quite sobering to reflect upon the fact that we currently do not have theories or models allowing us to provide firm, quantitative and robust answers to *any* of these questions. Moreover, the preceding two chapters should have made it clear that even in the sun, we don't really know for sure what is the mechanism responsible for the regeneration of the poloidal magnetic component. How then can we hope to go about modelling stellar dynamos with anything resembling confidence? At this point in time I would argue that we cannot. However, the problem can be turned around, in that stellar observations can perhaps be used to distinguish between different classes of dynamo models. The possibility hinges on features like the distinct dependency of the cycle period on model parameters in various models. For the simple α -quenched mean-field dynamo solutions discussed in §0.26.9, the (dimensionless) cycle period is, to a first approximation, independent on the dynamo numbers (see Fig. 34B), so that the physical period scales primarily as

$$P_{\text{cyc}} \propto \eta^{-1}, \quad [\alpha\text{-quenched } \alpha\Omega \text{ model}] \quad (302)$$

where η is the assumed turbulent diffusivity. Mean-field models including more complex form of nonlinearities produce produce more complex parametric dependencies, but a strong dependency on η always emerges. This quantity, in turn, is expected to increase with increasing convective velocities (cf. eq. (213)), and therefore with increasing luminosity. On the other hand, in Babcock-Leighton dynamo models the cycle period is found to be controlled primarily by the turnover time of the meridional flow cell, with a much weaker dependency on the assumed value for the turbulent diffusivity. For the specific “solar” model described in §0.27, the cycle period is found to vary as:

$$P_{\text{cyc}} \propto u_0^{-0.89} s_0^{-0.13} \eta^{-0.22}, \quad [\text{Babcock} - \text{Leighton}] \quad (303)$$

where u_0 is the surface meridional flow speed (see Fig. 38, top left), and s_0 is the parameter measuring the magnitude of the Babcock-Leighton source term in eq. (261)⁵⁰. Unfortunately, using this relationship in conjunction with observed stellar cycle data requires one to specify how the meridional flow speed varies with rotation, which currently remains highly uncertain on the theoretical and simulation fronts. But this is a very promising avenue.

All of these model-based cycle period formulae have been obtained in kinematic dynamo models, and how well they hold in the dynamical nonlinear regime is entirely unknown. With regular cycles now reproduced in global 3D MHD simulations of solar convection (cf. §0.29), these dynamical effects may become ripe for investigation in the near future.

0.39 Fully convective stars

{sec:fullconvdyn}

We now move to the bottom end of the main-sequence, where stars become fully convective around spectral type M5. Observationally, no obvious discontinuity is observed in X-Ray or Ca H+K emission as one moves into spectral types M, and indeed some of the more active (single) flare stars are fast rotators of very late spectral type. Yet with such fully convective stars we certainly encounter potential deviations from a solar-type dynamo mechanism; without a stably stratified tachocline and radiative core to store and amplify toroidal flux ropes, the Babcock-Leighton mechanism (§0.27), the tachocline α -effect and the flux-tube α -effect (§0.28) all becomes problematic. Mean-field models based on the turbulent α -effect remain viable in principle, but the dynamo behavior becomes dependent on the presence and

⁵⁰ Note however that the above relation was calibrated in a relatively narrow range of parameters: $2 \leq u_0 \leq 30 \text{ m s}^{-1}$, $0.03 \leq s_0 \leq 1 \text{ m s}^{-1}$, $2 \times 10^6 \leq \eta \leq 5 \times 10^7 \text{ m}^2\text{s}^{-1}$, and is only expected to hold in the so-called advection-dominated regime; see the paper by Dikpati & Charbonneau (1999) cited in the bibliography.

strength of differential rotation, about which we really don't know very much in stars other than the sun.

One might have expected dynamo action in fully convective stars —either in late-M main-sequence or pre-main sequence TTauri stars— to be a mere variant on core dynamo action in massive stars, but in fact a number of significant differences come into play, related to the physical conditions at the boundary of the convecting sphere. Full-sphere MHD simulations⁵¹ of a “M-star in a box” by Dobler, Stix & Brandenburg (2006) are particularly interesting in this respect. They indicate that vigorous dynamo action does occur, with the magnetic energy at $\sim 20\%$ of equipartition with the turbulent fluid motions at low to moderate rotation, and reaching equipartition at high rotation rates. The simulations are characterized by a very well-defined, persistent spatial pattern of mean kinetic helicity, again negative in N-hemisphere (see Fig. 75A). This leads to the production of a well-defined large-scale magnetic component, with energy content going from some 20% of the total magnetic energy at low rotation, up to $\sim 50\%$ at high rotation rates. The large-scale field has poloidal and toroidal components of comparable strength, typical of mean-field α^2 dynamos, and is often dominated by a well-defined quadrupolar component (see Fig. 75B). Because the convecting sphere cannot exchange angular momentum across its outer boundary (here the stellar surface), differential rotation is much weaker than in the massive star core dynamo simulations reviewed in §0.36.2), and is concentrated in the vicinity of the rotation axis, as shown on Fig. 75C).

Much like in the core dynamo simulations briefly described in §0.36.2, these simulations only reach moderate values of the viscous and magnetic Reynolds numbers, many orders of magnitude below what one would expect under stellar interior conditions. Nonetheless the important bottom line, once again, is that production of magnetic field through dynamo action here also appears inescapable.

{ssec:prepost}

0.40 Pre- and post-main-sequence stars

As with main-sequence late-type stars, abundant evidence for magnetic fields in pre- and post-main sequence stars of spectral types later than F has now been accumulating, mostly again in the form of stellar analogs to well-observed solar phenomena: X-Ray and EUV emission, flaring, spectral variability, rotational modulation by starspots, and so on. More recently magnetic Doppler imaging has been used to reconstruct the surface magnetic field of some pre-main-sequence stars in the TTauri evolutionary phase. Whether TTauri or giants, all these stars have low surface temperature and thick convection zones, so observations of magnetic activity indicators similar to what

⁵¹ The content of this section is based primarily on the paper by W. Dobler, M. Stix & A. Brandenburg, *ApJ*, **638**, 336 (2006)

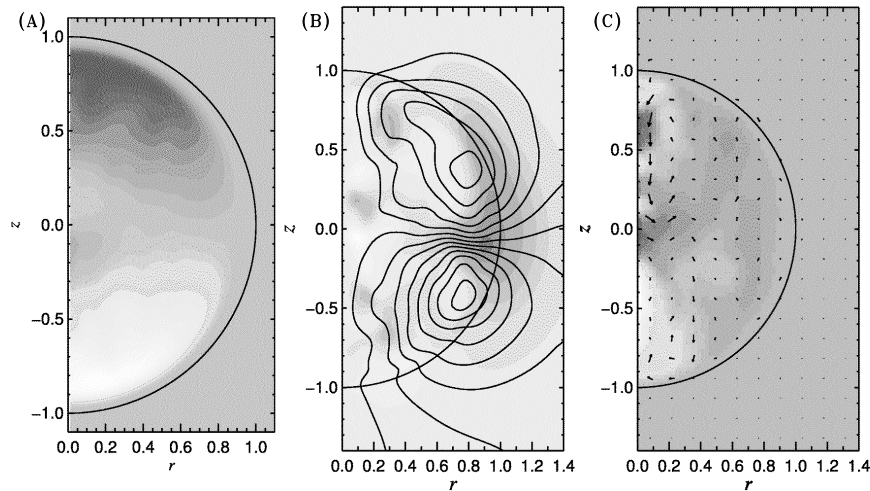


Fig. 75 Temporal+Azimuthal average of (A) kinetic helicity, (B) toroidal and poloidal magnetic field, and (C) large-scale flows the toroidal magnetic field in a complete MHD simulation (including overall structure) of a fully convective star including central heat source and surface cooling. The gray scale codes kinetic helicity in part (A), the toroidal magnetic component in (B), and the zonal flow component in (C), gray to black (white) coding negative (positive) values. Figure adapted from Dobler *et al.* 2006, ApJ, 638, 336 (Figure 5, p. 341, and Figure 10, p. 343). Fig:DSB

is observed in late-type main-sequence stars points once again to the importance of convection zones of significant radial extent below the photosphere. Indeed, there seems to exist a rather clear-cut, slightly inclined dividing line bisecting the upper part of the HR diagram (main-sequence and up in luminosity), on the right side (low T_{eff}) of which evidence of magnetic activity is ubiquitous. Things get messy again with very cool supergiants, with signs of magnetic activity disappearing across various not quite coincident dividing lines, depending on the indicator chosen (X-Ray emission, non-thermal emission lines, etc).

With classical TTauri stars, additional complications also come from the presence of an accretion disk, itself most likely magnetized, perhaps the site of magnetic field generation by dynamo action, and perhaps even magnetically coupled to its central star. Such a coupling has been invoked to explain the (relatively) low rotation rates of TTauri stars, which after all are contracting and accreting large amounts of mass—and angular momentum—from their disk, and should therefore spin up far more than is observed. Indeed, without angular momentum loss mediated by magnetic fields in the early stages of star formation, it is quite likely that stars could simply not eliminate enough angular momentum to form at all!

In hot post-main sequence stars, the observational situation is not well documented or understood. It is a remarkable fact that magnetic fields have been

detected in *all* sdO and sdB hot subdwarfs for which a serious attempt has been made. The evolutionary status of these objects is not well-understood, but they most likely represent what used to be the inner core of giants prior to the episode of strong mass loss that accompanies the transition to the horizontal branch. Detection of ~ 0.1 T-strength magnetic fields in such stars is strong evidence for the existence of magnetic fields in the deep interior of their main-sequence progenitors.

{ssec:compact}

0.41 Compact objects

Magnetic fields in isolated white dwarfs have been detected through circular polarisation measurements in the wings of strong spectral lines, usually Balmer lines in the so-called “DA” white dwarfs showing Hydrogen lines in their photospheres. Actual Zeeman splitting is only detected in the most strongly magnetized objects (\gtrsim a few 10^2 T). Inferred field strengths range from a few T up to a whopping 10^4 T, with a few objects possibly approaching 10^5 T, and the overall incidence of magnetism standing at 10–15 percent. However, these techniques are only sensitive to large-scale magnetic fields, still producing a net polarisation signal when integrated over the stellar disk, and so the true incidence of magnetism in white dwarfs may actually be significantly higher. At around 22,000 K on the white dwarf cooling track, there appears a subclass of Carbon-rich objects, the so-called “hot DQ” white dwarfs, in which the Carbon enrichment believed to be due to the onset of envelope convection and associated mixing. Interestingly, the incidence of magnetism in this subclass of white dwarfs is some four times larger than across the overall white dwarf population. Could this sudden increase in the incidence of magnetism be physically associated with the onset of envelope convection through a dynamo mechanism? This interesting question remains to be explored.

Inferred magnetic field strengths in neutron stars range from 10^4 to 10^{11} T. Neutron stars magnetic fields are of course most readily detected via the pulsar phenomenon, arising from the misalignment of the magnetic axis with respect to the rotation axis of the (very rapidly rotating) neutron star. It is quite striking that the highest strengths of large-scale magnetic fields in main-sequence stars (a few T in Ap stars), in white dwarfs ($\sim 10^5$ T) and in the most strongly magnetized neutron stars ($\sim 10^{11}$ T) all amount to similar surface magnetic fluxes, lending support to the idea that these high field strengths can be understood from simple flux-freezing arguments (§0.12). There is also observational evidence that actual magnetic field evolution is taking place as pulsars age, but this remains very slippery territory, both from the modelling and observational points of view.

Observationally, very little is known about black holes except that there seems to be one at the center of our galaxy, so you won’t be surprised to

hear that even less is known about black hole magnetic fields. One should perhaps just point out that solutions to the field equations of general relativity for rotating, electrically charged black holes do exist, which is a good start towards magnetic fields production. Evidence to date is limited to energetic phenomena interpreted in terms of magnetic channelling of material onto the black hole. But beyond that, at the present time there is mostly religious fervor.

1 Galaxies and beyond

{sec:galax}

Magnetic fields in the diffuse, low-density interstellar gas is most readily detected through synchrotron radiation emitted by relativistic charged particles spiralling along magnetic fieldlines. This technique is succesfull not only within the Milky Way, but also for other galaxies. Other means of detection, for the time being limited to the Milky Way, include the polarisation of optical starlight by elongated (i.e., non-spherical) dust grains aligning themselves perpendicularly to magnetic fieldlines; these aligned dust grains also sometimes emit detectable polarized infrared radiation. Finally, for relatively strong fields Zeeman splitting of spectral lines in the radio domain has also been measured. As with stars, magnetic fields seem to be ubiquitous features in pretty much all galaxies. Indirect evidence for the existence of extragalactic magnetic fields also exists, with an upper limit of $\sim 10^{-3}$ nT on the mean field strength over length scales of order 100Mpc and larger. These fields could be primordial in origin, or could have been ejected in intergalactic space by galactic winds.

The galactic magnetic field in the solar neighbourhood has a strength of about 0.6nT, up to a few nT near galactic center. This is indeed typical of spiral galaxies, which show field strengths in the range 0.5–1.5nT, up to some 3nT in high density regions of spiral arms. The strongest large-scale galactic magnetic fields so far measured have strength reaching 1nT, and have been found in starburst galaxies. While this may seem quite low values, such field strengths have important consequences for star formation, the distribution of cosmic rays, and equilibrating the interstellar medium against gravity.

Given that most stars appear to be magnetized to some degrees, and that many stars tend lose mass (some by blowing up!), it is perhaps not surprising to detect magnetic field in the galactic interstellar medium. What is surprising is that this magnetic field tends to be organized on large spatial scales, commensurate in fact with galactic dimensions. Such large-scale, spatially well-organized magnetic fields are most likely produced by a dynamo mechanism, not at all dissimilar to that responsible for the presence of magnetic fields in many stars, including the Sun. Simple kinematic models, much like those considered in §0.26, have been built, in which differential rotation in the disk and interstellar turbulence driven by supernovae explosion and/or

magnetic buoyancy instabilities could jointly act as a $\alpha\Omega$ dynamo on galactic scales. Additional, indirect evidence for well-organized large-scale magnetic fields in galaxies include the collimation of jets, and energetic phenomena often encountered in quasars and AGN; at the present time, the most convincing physical models for such phenomena all involve magnetic fields at some level.

These galactic magnetic fields can also provide the seed required to start up the dynamo processes in the sun and stars, as per the linearity of the MHD induction equation, which brings us back almost to the beginning of our grand tour of solar and stellar dynamos. The next lap is yours to take. Have fun with it!

Bibliography:

The following conference proceedings gives an excellent sampling of the current state-of-the-art in the observation of stellar magnetic fields:

Neiner, C., & Zahn, J.-P. (eds.), *Stellar Magnetism*, EAS Publication Series, **39** (2009).

The first three chapters therein, authored by John Landstreet (pps. 1-53), provide an outstanding overview of the subject, in just a little over 50 pages.

The linear α^2 and $\alpha^2\Omega$ models for core dynamo action in massive stars presented in §0.36 are taken pretty directly from

Charbonneau, P., & MacGregor, K.B., *Astrophys. J.*, **559**, 1094-1107 (2001).

This paper also addresses the effects of thermally-driven meridional circulation in the radiative envelope, including its deleterious effect on core dynamo action. The difficulty in bringing the magnetic fields produced by core dynamo action to the surface of a star with a thick radiative envelope was cogently demonstrated a while ago already by

Levy, E.H., & Rose, W.K., *Astrophys. J.*, **193**, 419-427 (1974),
Schüssler, M., & Pähler, A., *Astron. Astrophys.*, **68**, 57-62 (1978).

The discussion of core dynamo action in massive stars (§0.36.2) is based on the following paper:

Brun, A.S., Browning, M.K., & J. Toomre, J., *Astrophys. J.*, **629**, 461-481 (2005),

from which Figure 69 was taken. For calculations of buoyantly rising thin flux tubes in the radiative envelope of massive stars, see

MacGregor, K.B., & Cassinelli, J., *Astrophys. J.*, **586**, 480-494 (2003),
MacDonald, J., & Mullan, D.J., *Month. Not. R. Astron. Soc.*, **348**, 702-716 (2004).

On the inference of large-scale magnetic fields on slowly rotating chemically peculiar stars, see

Donati, J.-F., & Landstreet, J.D., *Astron. Astrophys. Rev.*, **47**, 333-370 (2009).

On the stability of large-scale magnetic field in stellar radiative interiors, start with:

Pitts, E., & Tayler, R.J., *Month. Not. R. Astron. Soc.*, **216**, 139-154 (1985),
Braithwaite, J., & Nordlund, Å, *Astron. Astrophys.*, **450**, 1077-1095 (2006).

On the associated possible dynamo effect:

Spruit, H.C., *Astron. Astrophys.*, **381**, 923-932 (2002),
Arlt, R., Hollerbach, R., & Rüdiger, G., *Astron. Astrophys.*, **401**, 1087-1094 (2003),
Braithwaite, J., *Astron. Astrophys.*, **449**, 451-460 (2006),

and for consequences on the structural, rotational and chemical evolution of massive stars:

Maeder, A., & Meynet, G., *Astron. Astrophys.*, **411**, 543-552 (2003),
Maeder, A., & Meynet, G., *Astron. Astrophys.*, **440**, 1041-1049 (2005).

The potential observational consequences of iron opacity-driven outer convection zones in massive stars (including a brief discussion of the possibility of dynamo action) are discussed in

Cantiello, M., Langer, N., and 7 co-authors, *Astron. Astrophys.*, **499**, 279-290 (2009).

Observations of solar-like activity in late-type stars are reviewed in in the following two recent online review articles:

Berdyugina, S.V., *Liv. Rev. Solar Phys.*, **2**, (2005),
<http://solarphysics.livingreviews.org/Articles/lrsp-2005-8/>
Hall, J.C., *Liv. Rev. Solar Phys.*, **5**, (2008),
<http://solarphysics.livingreviews.org/Articles/lrsp-2008-2/>

Figure 72 was taken from the following paper, still today one of the more cogent exposition of the Mt Wilson CaII project and data:

Baliunas, S.L., Donahue, R.A., Soon, W.H., and 24 co-authors, *Astrophys. J.*, **438**, 269-287 (1995).

On the confrontation of such observations with various types of dynamo models, start with

Noyes, R.W., Weiss, N.O., & Vaughan, A.H., *Astrophys. J.*, **287**, 769-773 (1984),
Saar, S.H., & Brandenburg, A., *Astrophys. J.*, **524**, 295-310 (1999),
Moss, D., *Month. Not. R. Astron. Soc.*, **352**, L17-L20 (2004).

Jouve, L., Brown, B.P., and Brun, A.S., *Astron. Astrophys.*, **509**, A32 (2010).

On the Weber-Davis MHD wind model, see

Weber, E.J., & Davis, L. Jr., *Astrophys. J.*, **148**, 217-227 (1967),
 Belcher, J.W., & MacGregor, K.B., *Astrophys. J.*, **210**, 498-507 (1976),

as well as

Keppens, R., & Goedbloed, J.P., *Astron. Astrophys.*, **343**, 251-260 (1999),

for a geometrically more realistic version. The idea that magnetized outflows can lead to stellar angular momentum loss can be traced to

Schatzman, E., *Ann. Astrophys.*, **25**, 18 (1962).

In this context, another important pioneering paper is

Mestel, L., *Month. Not. R. Astron. Soc.*, **138**, 359 (1968).

The theoretical derivation of Skumanich's square root relation in the context of $\alpha\Omega$ dynamo theory is due to

Durney, B., in *Solar Wind*, eds. C.P. Sonett, P.J. Coleman, & L.M. Wilcox (Washington: NASA), p. 282 (1972).

There is a huge technical literature available on the observation, evolution and consequences of stellar rotation. The following conference proceedings volume should make a good starting point to those interested in digging further into the various ramifications of this topic:

Eenens, P. Maeder, A., (ed.), *Stellar rotation*, IAU Symposium No. 215.

The discussion of dynamo action in fully convective stars (§0.39) is based primarily on the following paper:

Dobler, W., Stix, M., & Brandenburg, A., *Astrophys. J.*, **638**, 336-347 (2006),

but on this topic do not miss:

Durney, B.R., DeYoung, D.S., & Roxburgh, I.A., *Solar Phys.*, **145**, 207-225 (1993),

Browning, M.K., *Astrophys. J.*, **676**, 1262-1280 (2008).

On magnetic field detection in white dwarfs and hot subdwarfs, see

Liebert, J., Bergeron, P., & Holberg, J.B., *Astron. J.*, **125**, 348-353 (2003),
 Kawka, A., Vennes, S., Schmidt, G.D., Wickramasinghe, D.T., & Koch, R., *Astrophys. J.*, **654**, 499-520 (2007),

O'Toole, S.J., Jordan, S., Friedrich, S., & Heber, U., *Astron. Astrophys.*, **437**, 227-234 (2005).

I found the following online article a good starting point to learn about galactic magnetic fields:

Beck, R., *Scholarpedia*, **2**(8), 2411 (2007)

http://www.scholarpedia.org/articles/Galactic_magnetic_fields.

On galactic dynamo models, as well as alternatives explanations for galactic magnetic fields, the following review article provides an excellent entry point into the technical literature:

Kulsrud, R.M., & Zweibel, E.G., *Rep. Prog. Phys.*, **71**, 046901 (2008).

Appendix A: Useful identities and theorems from vector calculus

A.1 Vector Identities

$$\mathbf{A} \cdot (\mathbf{B} \times \mathbf{C}) = \mathbf{C} \cdot (\mathbf{A} \times \mathbf{B}) = \mathbf{B} \cdot (\mathbf{C} \times \mathbf{A})$$

$$\mathbf{A} \times (\mathbf{B} \times \mathbf{C}) = \mathbf{B}(\mathbf{A} \cdot \mathbf{C}) - \mathbf{C}(\mathbf{A} \cdot \mathbf{B})$$

$$(\mathbf{A} \times \mathbf{B}) \times \mathbf{C} = \mathbf{B}(\mathbf{A} \cdot \mathbf{C}) - \mathbf{A}(\mathbf{B} \cdot \mathbf{C})$$

$$\nabla \times \nabla f = 0$$

$$\nabla \cdot (\nabla \times \mathbf{A}) = 0$$

$$\nabla \cdot (f\mathbf{A}) = (\nabla f) \cdot \mathbf{A} + f(\nabla \cdot \mathbf{A})$$

$$\nabla \times (f\mathbf{A}) = (\nabla f) \times \mathbf{A} + f(\nabla \times \mathbf{A})$$

$$\nabla \cdot (\mathbf{A} \times \mathbf{B}) = \mathbf{B} \cdot (\nabla \times \mathbf{A}) - \mathbf{A} \cdot (\nabla \times \mathbf{B})$$

$$\nabla(\mathbf{A} \cdot \mathbf{B}) = (\mathbf{B} \cdot \nabla)\mathbf{A} + (\mathbf{A} \cdot \nabla)\mathbf{B} + \mathbf{B} \times (\nabla \times \mathbf{A}) + \mathbf{A} \times (\nabla \times \mathbf{B})$$

$$\nabla \cdot (\mathbf{A}\mathbf{B}) = (\mathbf{A} \cdot \nabla)\mathbf{B} + (\mathbf{B} \cdot \nabla)\mathbf{A}$$

$$\nabla \times (\mathbf{A} \times \mathbf{B}) = (\mathbf{B} \cdot \nabla)\mathbf{A} - (\mathbf{A} \cdot \nabla)\mathbf{B} - \mathbf{B}(\nabla \cdot \mathbf{A}) + \mathbf{A}(\nabla \cdot \mathbf{B})$$

$$\nabla \times (\nabla \times \mathbf{A}) = \nabla(\nabla \cdot \mathbf{A}) - \nabla^2 \mathbf{A}$$

The Gradient Theorem

For two points \mathbf{a} , \mathbf{b} in a space where a scalar function f with spatial derivatives everywhere well-defined up to first order,

$$\int_{\mathbf{a}}^{\mathbf{b}} (\nabla f) \cdot d\boldsymbol{\ell} = f(\mathbf{b}) - f(\mathbf{a}) ,$$

independently of the integration path between \mathbf{a} and \mathbf{b} .

The Divergence Theorem

For any vector field \mathbf{A} with spatial derivatives of all, its scalar components everywhere well-defined up to first order,

$$\int_V (\nabla \cdot \mathbf{A}) dV = \oint_S \mathbf{A} \cdot \hat{\mathbf{n}} dS ,$$

where the surface S encloses the volume V .

A.4 Stokes' theorem

For any vector field \mathbf{A} with spatial derivatives of all, its scalar components everywhere well-defined up to first order,

$$\int_S (\nabla \times \mathbf{A}) \cdot \hat{\mathbf{n}} dS = \oint_{\gamma} \mathbf{A} \cdot d\boldsymbol{\ell} ,$$

where the contour γ delimits the surface S , and the orientation of the unit normal vector $\hat{\mathbf{n}}$ and direction of contour integration are mutually linked by the right-hand rule.

Appendix B: Coordinate systems and the fluid equations

{pch_app: coord}

This Appendix is adapted in part from Appendix B of the book by Jean-Louis Tassoul entitled *Theory of Rotating Stars* (Princeton University Press, 1978), with a number of additions, including the MHD induction equation, expressions for the operators $\mathbf{u} \cdot \nabla$, $\nabla \times \nabla \times$, and ∇^2 acting on a vector

field. Also, the Note that in sections 1.1.4 and 1.2.4, the quantities in square brackets correspond to the components of the deformation tensor $D_{jk} = (1/2)(\partial_j u_k + \partial_k u_j)$.

B.1 Cylindrical coordinates (s, ϕ, z)

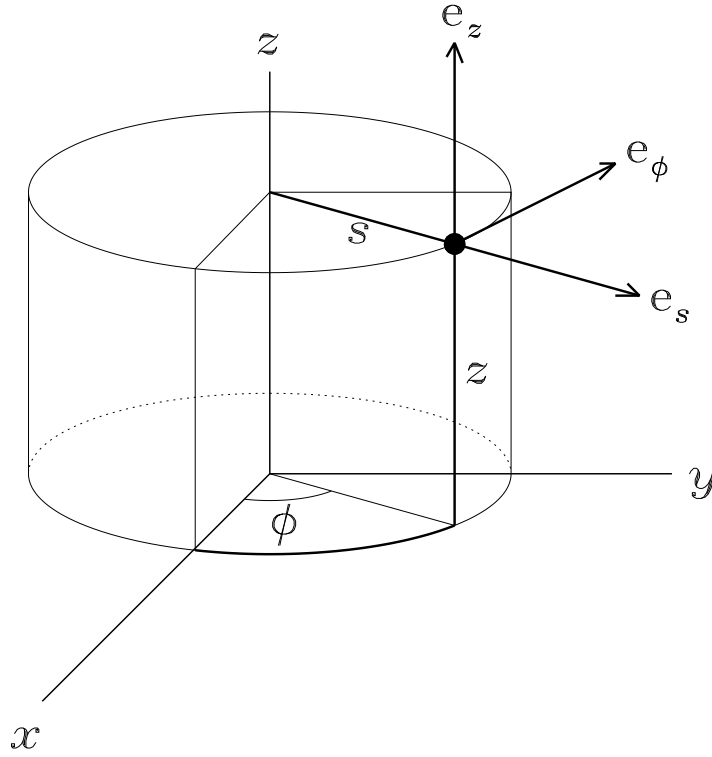


Fig. 76 Geometric definition of cylindrical coordinates. The coordinate ranges are $s \in [0, \infty]$, $\phi \in [0, 2\pi]$, $z \in [-\infty, \infty]$. The cylindrical radius s is measured perpendicularly from the cartesian z -axis. The zero point of the azimuthal angle ϕ is on the cartesian x -axis. The local unit vector triad is oriented such that $\hat{e}_z = \hat{e}_s \times \hat{e}_\phi$.

B.1.1 Conversion to cartesian coordinates

$$x = s \cos \phi, \quad y = s \sin \phi, \quad s = \sqrt{x^2 + y^2}, \quad \phi = \text{atan}(y/x), \quad z = z.$$

$$\hat{\mathbf{e}}_x = \cos \phi \hat{\mathbf{e}}_s - \sin \phi \hat{\mathbf{e}}_\phi, \quad \hat{\mathbf{e}}_y = \sin \phi \hat{\mathbf{e}}_s + \cos \phi \hat{\mathbf{e}}_\phi,$$

$$\hat{\mathbf{e}}_s = \cos \phi \hat{\mathbf{e}}_x + \sin \phi \hat{\mathbf{e}}_y, \quad \hat{\mathbf{e}}_\phi = -\sin \phi \hat{\mathbf{e}}_x + \cos \phi \hat{\mathbf{e}}_y, \quad \hat{\mathbf{e}}_z = \hat{\mathbf{e}}_z.$$

B.1.2 Infinitesimals

$$d\ell = ds \hat{\mathbf{e}}_s + s d\phi \hat{\mathbf{e}}_\phi + dz \hat{\mathbf{e}}_z$$

$$dV = s ds d\phi dz$$

B.1.3 Vector operators

$$\frac{D}{Dt} = \frac{\partial}{\partial t} + u_s \frac{\partial}{\partial s} + \frac{u_\phi}{s} \frac{\partial}{\partial \phi} + u_z \frac{\partial}{\partial z}$$

$$\nabla f = \frac{\partial f}{\partial s} \hat{\mathbf{e}}_s + \frac{1}{s} \frac{\partial f}{\partial \phi} \hat{\mathbf{e}}_\phi + \frac{\partial f}{\partial z} \hat{\mathbf{e}}_z$$

$$(\mathbf{u} \cdot \nabla) \mathbf{A} = \left(\mathbf{u} \cdot \nabla A_s - \frac{u_\phi A_\phi}{s} \right) \hat{\mathbf{e}}_s + \left(\mathbf{u} \cdot \nabla A_\phi + \frac{u_\phi A_s}{s} \right) \hat{\mathbf{e}}_\phi + (\mathbf{u} \cdot \nabla A_z) \hat{\mathbf{e}}_z$$

$$\nabla \cdot \mathbf{A} = \frac{1}{s} \frac{\partial}{\partial s} (s A_s) + \frac{1}{s} \frac{\partial A_\phi}{\partial \phi} + \frac{\partial A_z}{\partial z}$$

$$\nabla \times \mathbf{A} = \left(\frac{1}{s} \frac{\partial A_z}{\partial \phi} - \frac{\partial A_\phi}{\partial z} \right) \hat{\mathbf{e}}_s + \left(\frac{\partial A_s}{\partial z} - \frac{\partial A_z}{\partial s} \right) \hat{\mathbf{e}}_\phi + \frac{1}{s} \left(\frac{\partial (s A_\phi)}{\partial s} - \frac{\partial A_s}{\partial \phi} \right) \hat{\mathbf{e}}_z$$

$$\nabla^2 = \frac{1}{s} \frac{\partial}{\partial s} \left(s \frac{\partial}{\partial s} \right) + \frac{1}{s^2} \frac{\partial^2}{\partial \phi^2} + \frac{\partial^2}{\partial z^2}$$

$$\nabla^2 \mathbf{A} = \left(\nabla^2 A_s - \frac{A_s}{s^2} - \frac{2}{s^2} \frac{\partial A_\phi}{\partial \phi} \right) \hat{\mathbf{e}}_s + \left(\nabla^2 A_\phi - \frac{A_\phi}{s^2} + \frac{2}{s^2} \frac{\partial A_s}{\partial \phi} \right) \hat{\mathbf{e}}_\phi + (\nabla^2 A_z) \hat{\mathbf{e}}_z$$

B.1.4 The divergence of a second-order tensor

$$[\nabla \cdot \mathbf{T}]_s = \frac{1}{s} \frac{\partial(sT_{ss})}{\partial s} + \frac{1}{s} \frac{\partial T_{\phi s}}{\partial \phi} + \frac{\partial T_{zs}}{\partial z} - \frac{T_{\phi\phi}}{s}$$

$$[\nabla \cdot \mathbf{T}]_\phi = \frac{1}{s} \frac{\partial s T_{(s\phi)}}{\partial s} + \frac{1}{s} \frac{\partial T_{\phi\phi}}{\partial \phi} + \frac{\partial T_{z\phi}}{\partial z} + \frac{T_{\phi s}}{s}$$

$$[\nabla \cdot \mathbf{T}]_z = \frac{1}{s} \frac{\partial(sT_{sz})}{\partial s} + \frac{1}{s} \frac{\partial T_{\phi z}}{\partial \phi} + \frac{\partial T_{zz}}{\partial z}$$

B.1.5 Components of the viscous stress tensor

$$\tau_{ss} = 2\mu \left[\frac{\partial u_s}{\partial s} \right] + \left(\mu_\vartheta - \frac{2}{3}\mu \right) \nabla \cdot \mathbf{u}$$

$$\tau_{\phi\phi} = 2\mu \left[\frac{1}{s} \frac{\partial u_\phi}{\partial \phi} + \frac{u_s}{s} \right] + \left(\mu_\vartheta - \frac{2}{3}\mu \right) \nabla \cdot \mathbf{u}$$

$$\tau_{zz} = 2\mu \left[\frac{\partial u_z}{\partial z} \right] + \left(\mu_\vartheta - \frac{2}{3}\mu \right) \nabla \cdot \mathbf{u}$$

$$\tau_{s\phi} = \tau_{\phi s} = 2\mu \left[\frac{1}{2} \left(\frac{1}{s} \frac{\partial u_s}{\partial \phi} + s \frac{\partial}{\partial s} \frac{u_\phi}{s} \right) \right]$$

$$\tau_{\phi z} = \tau_{z\phi} = 2\mu \left[\frac{1}{2} \left(\frac{\partial u_\phi}{\partial z} + \frac{1}{s} \frac{\partial u_z}{\partial \phi} \right) \right]$$

$$\tau_{zs} = \tau_{sz} = 2\mu \left[\frac{1}{2} \left(\frac{\partial u_z}{\partial s} + \frac{\partial u_s}{\partial z} \right) \right]$$

B.1.6 Equations of motion

$$\rho \left(\frac{Du_s}{Dt} - \frac{u_\phi^2}{s} \right) = -\rho \frac{\partial \Phi}{\partial s} - \frac{\partial p}{\partial s} + \frac{B_z}{\mu_0} \left(\frac{\partial B_s}{\partial z} - \frac{\partial B_z}{\partial s} \right) - \frac{B_\phi}{\mu_0 s} \left(\frac{\partial(sB_\phi)}{\partial s} - \frac{\partial B_s}{\partial \phi} \right)$$

$$+ \frac{1}{s} \frac{\partial}{\partial s} (s\tau_{ss}) + \frac{1}{s} \frac{\partial \tau_{s\phi}}{\partial \phi} + \frac{\partial \tau_{sz}}{\partial z} - \frac{\tau_{\phi\phi}}{s}$$

$$\rho \left(\frac{Du_\phi}{Dt} - \frac{u_\phi u_s}{s} \right) = -\frac{\rho}{s} \frac{\partial \Phi}{\partial \phi} - \frac{1}{s} \frac{\partial p}{\partial \phi} + \frac{B_s}{\mu_0 s} \left(\frac{\partial(sB_\phi)}{\partial s} - \frac{\partial B_s}{\partial \phi} \right) - \frac{B_z}{\mu_0} \left(\frac{1}{s} \frac{\partial B_z}{\partial \phi} - \frac{\partial B_\phi}{\partial z} \right)$$

$$+ \frac{1}{s} \frac{\partial}{\partial s} (s\tau_{\phi s}) + \frac{1}{s} \frac{\partial \tau_{\phi\phi}}{\partial \phi} + \frac{\partial \tau_{\phi z}}{\partial z} + \frac{\tau_{s\phi}}{s}$$

$$\rho \frac{Du_z}{Dt} = -\rho \frac{\partial \Phi}{\partial z} - \frac{\partial p}{\partial z} + \frac{B_\phi}{\mu_0} \left(\frac{1}{s} \frac{\partial B_z}{\partial \phi} - \frac{\partial B_\phi}{\partial z} \right) - \frac{B_s}{\mu_0} \left(\frac{\partial B_s}{\partial z} - \frac{\partial B_z}{\partial s} \right) + \frac{1}{s} \frac{\partial}{\partial s} (s\tau_{zs}) + \frac{1}{s} \frac{\partial \tau_{z\phi}}{\partial \phi} + \frac{\partial \tau_{zz}}{\partial z}$$

B.1.7 The energy equation

$$\rho T \frac{Ds}{Dt} = \Phi_u + \Phi_\eta + \frac{1}{s} \frac{\partial}{\partial s} \left[\chi^s \frac{\partial T}{\partial s} \right] + \frac{1}{s^2} \frac{\partial}{\partial \phi} \left[\chi \frac{\partial T}{\partial s} \right] + \frac{\partial}{\partial z} \left[\chi \frac{\partial T}{\partial z} \right]$$

$$\Phi_u = 2\mu(D_{ss}^2 + D_{\phi\phi}^2 + D_{zz}^2 + 2D_{s\phi}^2 + 2D_{\phi z}^2 + 2D_{zs}^2) + (\mu_\vartheta - \frac{2}{3}\mu)(\nabla \cdot \mathbf{u})^2$$

$$\Phi_\eta = \frac{\eta}{\mu_0} \left[\left(\frac{1}{s} \frac{\partial B_z}{\partial \phi} - \frac{\partial B_\phi}{\partial z} \right)^2 + \left(\frac{\partial B_s}{\partial z} - \frac{\partial B_z}{\partial s} \right)^2 + \frac{1}{s^2} \left(\frac{\partial(sB_\phi)}{\partial s} - \frac{\partial B_s}{\partial \phi} \right)^2 \right]$$

B.1.8 The MHD induction equation

$$\frac{\partial B_s}{\partial t} = \frac{1}{s} \frac{\partial}{\partial \phi} (u_s B_\phi - u_\phi B_s) - \frac{\partial}{\partial z} (u_z B_s - u_s B_z)$$

$$- \frac{1}{s^2} \frac{\partial \eta}{\partial \phi} \left(\frac{\partial(sB_\phi)}{\partial s} - \frac{\partial B_s}{\partial \phi} \right) + \frac{\partial \eta}{\partial z} \left(\frac{\partial B_s}{\partial z} - \frac{\partial B_z}{\partial s} \right) + \eta \left(\nabla^2 B_s - \frac{B_s}{s^2} - \frac{2}{s^2} \frac{\partial B_\phi}{\partial \phi} \right)$$

$$\begin{aligned} \frac{\partial B_\phi}{\partial t} &= \frac{\partial}{\partial z} (u_\phi B_z - u_z B_\phi) - \frac{\partial}{\partial s} (u_s B_\phi - u_\phi B_s) \\ -\frac{\partial \eta}{\partial z} \left(\frac{1}{s} \frac{\partial B_z}{\partial \phi} - \frac{\partial B_\phi}{\partial z} \right) &+ \frac{1}{s} \frac{\partial \eta}{\partial s} \left(\frac{\partial (s B_\phi)}{\partial s} - \frac{\partial B_s}{\partial \phi} \right) + \eta \left(\nabla^2 B_\phi - \frac{B_\phi}{s^2} + \frac{2}{s^2} \frac{\partial B_s}{\partial \phi} \right) \\ \frac{\partial B_z}{\partial t} &= \frac{1}{s} \frac{\partial}{\partial s} (s u_z B_s - s u_s B_z) - \frac{1}{s} \frac{\partial}{\partial \phi} (u_\phi B_z - u_z B_\phi) \\ -\frac{\partial \eta}{\partial s} \left(\frac{\partial B_s}{\partial z} - \frac{\partial B_z}{\partial s} \right) &+ \frac{1}{s} \frac{\partial \eta}{\partial \phi} \left(\frac{1}{s} \frac{\partial B_z}{\partial \phi} - \frac{\partial B_\phi}{\partial z} \right) + \eta (\nabla^2 B_z) \end{aligned}$$

B.2 Spherical coordinates (r, θ, ϕ)

B.2.1 Conversion to cartesian coordinates

$$x = r \sin \theta \cos \phi, \quad y = r \sin \theta \sin \phi, \quad z = r \cos \theta.$$

$$r = \sqrt{x^2 + y^2 + z^2}, \quad \theta = \text{atan}(\sqrt{x^2 + y^2}/z), \quad \phi = \text{atan}(y/x).$$

$$\hat{e}_x = \sin \theta \cos \phi \hat{e}_r + \cos \theta \cos \phi \hat{e}_\theta - \sin \phi \hat{e}_\phi,$$

$$\hat{e}_y = \sin \theta \sin \phi \hat{e}_r + \cos \theta \sin \phi \hat{e}_\theta + \cos \phi \hat{e}_\phi,$$

$$\hat{e}_z = \cos \theta \hat{e}_r - \sin \theta \hat{e}_\theta.$$

$$\hat{e}_r = \sin \theta \cos \phi \hat{e}_x + \sin \theta \sin \phi \hat{e}_y + \cos \theta \hat{e}_z,$$

$$\hat{e}_\theta = \cos \theta \cos \phi \hat{e}_x + \cos \theta \sin \phi \hat{e}_y - \sin \theta \hat{e}_z,$$

$$\hat{e}_\phi = -\sin \phi \hat{e}_x + \cos \phi \hat{e}_y.$$

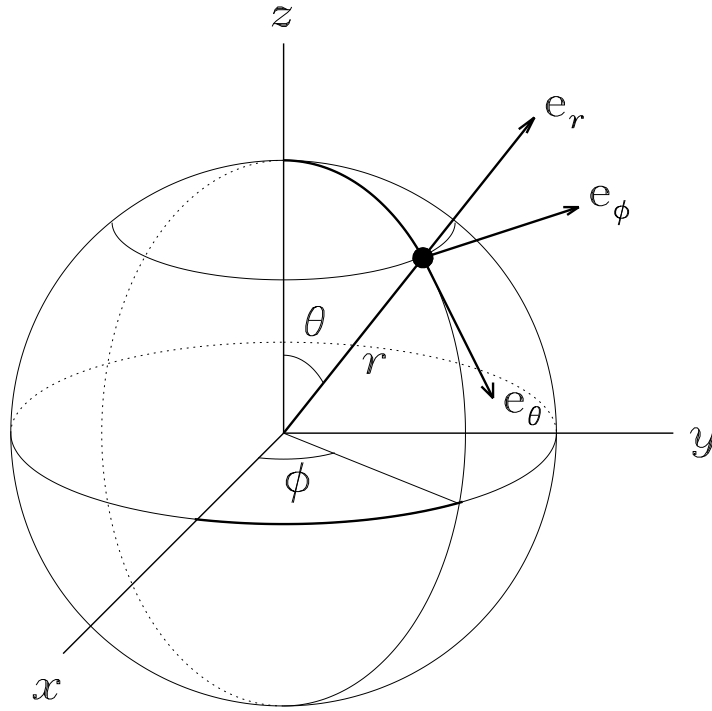


Fig. 77 Geometric definition of polar spherical coordinates. The coordinate ranges are $r \in [0, \infty]$, $\theta \in [0, \pi]$, $\phi \in [0, 2\pi]$. The zero point of the azimuthal angle ϕ is on the cartesian x -axis and the zero point of the polar angle θ (sometimes called *colatitude*) is on the cartesian z -axis. Note that in so-called geographical coordinates, *longitude* $\equiv \phi$, but *latitude* $\equiv \pi/2 - \theta$. The local unit vector triad is oriented such that $\hat{e}_r = \hat{e}_\theta \times \hat{e}_\phi$.

B.2.2 Infinitesimals

$$d\mathbf{l} = dr\hat{e}_r + r d\theta\hat{e}_\theta + r \sin\theta d\phi\hat{e}_\phi$$

$$dV = r^2 \sin\theta dr d\theta d\phi$$

B.2.3 Operators

$$\frac{D}{Dt} = \frac{\partial}{\partial t} + u_r \frac{\partial}{\partial r} + \frac{u_\theta}{r} \frac{\partial}{\partial \theta} + \frac{u_\phi}{r \sin\theta} \frac{\partial}{\partial \phi}$$

$$\begin{aligned}
\nabla f &= \frac{\partial f}{\partial r} \hat{\mathbf{e}}_r + \frac{1}{r} \frac{\partial f}{\partial \theta} \hat{\mathbf{e}}_\theta + \frac{1}{r \sin \theta} \frac{\partial f}{\partial \phi} \hat{\mathbf{e}}_\phi \\
(\mathbf{u} \cdot \nabla) \mathbf{A} &= \left(\mathbf{u} \cdot \nabla A_r - \frac{u_\theta A_\theta}{r} - \frac{u_\phi A_\phi}{r} \right) \hat{\mathbf{e}}_r \\
&+ \left(\mathbf{u} \cdot \nabla A_\theta - \frac{u_\phi A_\phi}{r} \cot \theta + \frac{u_\theta A_r}{r} \right) \hat{\mathbf{e}}_\theta + \left(\mathbf{u} \cdot \nabla A_\phi + \frac{u_\phi A_r}{r} + \frac{u_\theta A_\theta}{r} \cot \theta \right) \hat{\mathbf{e}}_\phi \\
\nabla \cdot \mathbf{A} &= \frac{1}{r^2} \frac{\partial}{\partial r} (r^2 A_r) + \frac{1}{r \sin \theta} \frac{\partial (A_\theta \sin \theta)}{\partial \theta} + \frac{1}{r \sin \theta} \frac{\partial A_\phi}{\partial \phi} \\
\nabla \times \mathbf{A} &= \frac{1}{r \sin \theta} \left(\frac{\partial (A_\phi \sin \theta)}{\partial \theta} - \frac{\partial A_\theta}{\partial \phi} \right) \hat{\mathbf{e}}_r \\
&+ \frac{1}{r \sin \theta} \left(\frac{\partial A_r}{\partial \phi} - \frac{\partial (A_\phi r \sin \theta)}{\partial r} \right) \hat{\mathbf{e}}_\theta + \frac{1}{r} \left(\frac{\partial (r A_\theta)}{\partial r} - \frac{\partial A_r}{\partial \theta} \right) \hat{\mathbf{e}}_\phi \\
\nabla^2 &= \frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial}{\partial r} \right) + \frac{1}{r^2 \sin \theta} \frac{\partial}{\partial \theta} \left(\sin \theta \frac{\partial}{\partial \theta} \right) + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2}{\partial \phi^2} \\
\nabla^2 \mathbf{A} &= \left(\nabla^2 A_r - \frac{2A_r}{r^2} - \frac{2}{r^2 \sin \theta} \frac{\partial A_\theta \sin \theta}{\partial \theta} - \frac{2}{r^2 \sin \theta} \frac{\partial A_\phi}{\partial \phi} \right) \hat{\mathbf{e}}_r \\
&+ \left(\nabla^2 A_\theta + \frac{2}{r^2} \frac{\partial A_r}{\partial \theta} - \frac{A_\theta}{r^2 \sin^2 \theta} - \frac{2 \cos \theta}{r^2 \sin^2 \theta} \frac{\partial A_\phi}{\partial \phi} \right) \hat{\mathbf{e}}_\theta \\
&+ \left(\nabla^2 A_\phi + \frac{2}{r^2 \sin \theta} \frac{\partial A_r}{\partial \phi} + \frac{2 \cos \theta}{r^2 \sin^2 \theta} \frac{\partial A_\theta}{\partial \phi} - \frac{A_\phi}{r^2 \sin^2 \theta} \right) \hat{\mathbf{e}}_\phi
\end{aligned}$$

B.2.4 The divergence of a second-order symmetric tensor

$$[\nabla \cdot \mathbf{T}]_r = \frac{1}{r^2} \frac{\partial (r^2 T_{rr})}{\partial r} + \frac{1}{r \sin \theta} \frac{\partial (T_{\theta r} \sin \theta)}{\partial \theta} + \frac{1}{r \sin \theta} \frac{\partial T_{\phi r}}{\partial \phi} - \frac{T_{\theta\theta} + T_{\phi\phi}}{r}$$

$$[\nabla \cdot \mathbf{T}]_\theta = \frac{1}{r^2} \frac{\partial(r^2 T_{r\theta})}{\partial r} + \frac{1}{r \sin \theta} \frac{\partial(T_{\theta\theta} \sin \theta)}{\partial \theta} + \frac{1}{r \sin \theta} \frac{\partial T_{\phi\theta}}{\partial \phi} + \frac{T_{\theta r}}{r} - \frac{T_{\phi\theta} \cot \theta}{r}$$

$$[\nabla \cdot \mathbf{T}]_\phi = \frac{1}{r^2} \frac{\partial(r^2 T_{r\phi})}{\partial r} + \frac{1}{r \sin \theta} \frac{\partial(T_{\theta\phi} \sin \theta)}{\partial \theta} + \frac{1}{r \sin \theta} \frac{\partial T_{\phi\phi}}{\partial \phi} + \frac{T_{\phi r}}{r} + \frac{T_{\phi\theta} \cot \theta}{r}$$

B.2.5 Components of the viscous stress tensor

$$\tau_{rr} = 2\mu \left[\frac{\partial u_r}{\partial r} \right] + (\mu_\vartheta - \frac{2}{3}\mu) \nabla \cdot \mathbf{u}$$

$$\tau_{\theta\theta} = 2\mu \left[\frac{1}{r} \frac{\partial u_\theta}{\partial \theta} + \frac{u_r}{r} \right] + (\mu_\vartheta - \frac{2}{3}\mu) \nabla \cdot \mathbf{u}$$

$$\tau_{\phi\phi} = 2\mu \left[\frac{1}{r \sin \theta} \frac{\partial u_\phi}{\partial \phi} + \frac{u_r}{r} + \frac{u_\theta \cot \theta}{r} \right] + (\mu_\vartheta - \frac{2}{3}\mu) \nabla \cdot \mathbf{u}$$

$$\tau_{r\theta} = \tau_{\theta r} = 2\mu \left[\frac{1}{2} \left(\frac{1}{r} \frac{\partial u_r}{\partial \theta} + r \frac{\partial}{\partial r} \frac{u_\theta}{r} \right) \right]$$

$$\tau_{\theta\phi} = \tau_{\phi\theta} = 2\mu \left[\frac{1}{2} \left(\frac{1}{r \sin \theta} \frac{\partial u_\theta}{\partial \phi} + \frac{\sin \theta}{r} \frac{\partial}{\partial \theta} \frac{u_\phi}{\sin \theta} \right) \right]$$

$$\tau_{\phi r} = \tau_{r\phi} = 2\mu \left[\frac{1}{2} \left(r \frac{\partial}{\partial r} \frac{u_\phi}{r} + \frac{1}{r \sin \theta} \frac{\partial u_r}{\partial \phi} \right) \right]$$

B.2.6 Equations of motion

$$\rho \left(\frac{D u_r}{D t} - \frac{u_\theta^2 + u_\phi^2}{r} \right) = -\rho \frac{\partial \Phi}{\partial r} - \frac{\partial p}{\partial r} + \frac{B_\phi}{\mu_0 r \sin \theta} \left(\frac{\partial B_r}{\partial \phi} - \frac{\partial}{\partial r} (B_\phi r \sin \theta) \right) - \frac{B_\theta}{\mu_0 r} \left(\frac{\partial (r B_\theta)}{\partial r} - \frac{\partial B_r}{\partial \theta} \right)$$

$$+ \frac{1}{r \sin \theta} \left[\frac{\sin \theta}{r} \frac{\partial}{\partial r} (r^2 \tau_{rr}) + \frac{\partial}{\partial \theta} (\tau_{r\theta} \sin \theta) + \frac{\partial \tau_{r\phi}}{\partial \phi} \right] - \frac{\tau_{\theta\theta} + \tau_{\phi\phi}}{r}$$

$$\begin{aligned}
\rho \left(\frac{Du_\theta}{Dt} + \frac{u_r u_\theta}{r} - \frac{u_\phi^2 \cot \theta}{r} \right) &= -\frac{\rho}{r} \frac{\partial \Phi}{\partial \theta} - \frac{1}{r} \frac{\partial p}{\partial \theta} + \frac{B_r}{\mu_0 r} \left(\frac{\partial(rB_\theta)}{\partial r} - \frac{\partial B_r}{\partial \theta} \right) - \frac{B_\phi}{\mu_0 r \sin \theta} \left(\frac{\partial(B_\phi \sin \theta)}{\partial \theta} - \frac{\partial B_\theta}{\partial \phi} \right) \\
&+ \frac{1}{r \sin \theta} \left[\frac{\sin \theta}{r} \frac{\partial}{\partial r} (r^2 \tau_{\theta r}) + \frac{\partial}{\partial \theta} (\tau_{\theta\theta} \sin \theta) + \frac{\partial \tau_{\theta\phi}}{\partial \phi} \right] + \frac{\tau_{r\theta}}{r} - \frac{\tau_{\phi\phi} \cot \theta}{r} \\
\rho \left(\frac{Du_\phi}{Dt} + \frac{u_r u_\phi}{r} + \frac{u_\theta u_\phi \cot \theta}{r} \right) &= -\frac{\rho}{r \sin \theta} \frac{\partial \Phi}{\partial \phi} - \frac{1}{r \sin \theta} \frac{\partial p}{\partial \phi} \\
&+ \frac{B_\theta}{\mu_0 r \sin \theta} \left(\frac{\partial(B_\phi \sin \theta)}{\partial \theta} - \frac{\partial B_\theta}{\partial \phi} \right) - \frac{B_r}{\mu_0 r \sin \theta} \left(\frac{\partial B_r}{\partial \phi} - \frac{\partial(B_\phi r \sin \theta)}{\partial r} \right) \\
&+ \frac{1}{r \sin \theta} \left[\frac{\sin \theta}{r} \frac{\partial}{\partial r} (r^2 \tau_{\phi r}) + \frac{\partial}{\partial \theta} (\tau_{\phi\theta} \sin \theta) + \frac{\partial \tau_{\phi\phi}}{\partial \phi} \right] + \frac{\tau_{r\phi}}{r} + \frac{\tau_{\theta\phi} \cot \theta}{r}
\end{aligned}$$

B.2.7 The energy equation

$$\begin{aligned}
\rho T \frac{Ds}{Dt} &= \Phi_u + \Phi_\eta + \frac{1}{r^2} \frac{\partial}{\partial r} \left[\chi r^2 \frac{\partial T}{\partial r} \right] + \frac{1}{r^2 \sin \theta} \frac{\partial}{\partial \theta} \left[\chi \sin \theta \frac{\partial T}{\partial \theta} \right] + \frac{1}{r^2 \sin^2 \theta} \frac{\partial}{\partial \phi} \left[\chi \frac{\partial T}{\partial \phi} \right] \\
\Phi_u &= 2\mu(D_{rr}^2 + D_{\theta\theta}^2 + D_{\phi\phi}^2 + 2D_{r\theta}^2 + 2D_{\theta\phi}^2 + 2D_{\phi r}^2) + (\mu_\vartheta - \frac{2}{3}\mu)(\nabla \cdot \mathbf{u})^2 \\
\Phi_\eta &= \frac{\eta}{\mu_0 r^2 \sin^2 \theta} \left[\left(\frac{\partial(B_\phi \sin \theta)}{\partial \theta} - \frac{\partial B_\theta}{\partial \phi} \right)^2 + \left(\frac{\partial B_r}{\partial \phi} - \frac{\partial(B_\phi r \sin \theta)}{\partial r} \right)^2 + \sin^2 \theta \left(\frac{\partial(rB_\theta)}{\partial r} - \frac{\partial B_r}{\partial \theta} \right)^2 \right]
\end{aligned}$$

B.2.8 The MHD induction equation

$$\begin{aligned}
\frac{\partial B_r}{\partial t} &= \frac{1}{r \sin \theta} \left[\frac{\partial}{\partial \theta} (\sin \theta (u_r B_\theta - u_\theta B_r)) - \frac{\partial}{\partial \phi} (u_\phi B_r - u_r B_\phi) \right] - \frac{1}{r^2} \frac{\partial \eta}{\partial \theta} \left(\frac{\partial(rB_\theta)}{\partial r} - \frac{\partial B_r}{\partial \theta} \right) \\
&+ \frac{1}{r^2 \sin^2 \theta} \frac{\partial \eta}{\partial \phi} \left(\frac{\partial B_r}{\partial \phi} - \frac{\partial(B_\phi r \sin \theta)}{\partial r} \right) + \eta \left(\nabla^2 B_r - \frac{2B_r}{r^2} - \frac{2}{r^2 \sin \theta} \frac{\partial(B_\theta \sin \theta)}{\partial \theta} - \frac{2}{r^2 \sin \theta} \frac{\partial B_\phi}{\partial \phi} \right)
\end{aligned}$$

$$\begin{aligned} \frac{\partial B_\theta}{\partial t} &= \frac{1}{r \sin \theta} \frac{\partial}{\partial \phi} (u_\theta B_\phi - u_\phi B_\theta) - \frac{1}{r} \frac{\partial}{\partial r} (r u_r B_\theta - r u_\theta B_r) - \frac{1}{r^2 \sin^2 \theta} \frac{\partial \eta}{\partial \phi} \left(\frac{\partial (B_\phi \sin \theta)}{\partial \theta} - \frac{\partial B_\theta}{\partial \phi} \right) \\ &+ \frac{1}{r} \frac{\partial \eta}{\partial r} \left(\frac{\partial (r B_\theta)}{\partial r} - \frac{\partial B_r}{\partial \theta} \right) + \eta \left(\nabla^2 B_\theta + \frac{2}{r^2} \frac{\partial B_r}{\partial \theta} - \frac{B_\theta}{r^2 \sin^2 \theta} - \frac{2 \cos \theta}{r^2 \sin^2 \theta} \frac{\partial B_\phi}{\partial \phi} \right) \\ \frac{\partial B_\phi}{\partial t} &= \frac{1}{r} \left[\frac{\partial}{\partial r} (r u_\phi B_r - r u_r B_\phi) - \frac{\partial}{\partial \theta} (u_\theta B_\phi - u_\phi B_\theta) \right] - \frac{1}{r \sin \theta} \frac{\partial \eta}{\partial r} \left(\frac{\partial B_r}{\partial \phi} - \frac{\partial (B_\phi r \sin \theta)}{\partial r} \right) \\ &+ \frac{1}{r^2 \sin \theta} \frac{\partial \eta}{\partial \theta} \left(\frac{\partial (B_\phi \sin \theta)}{\partial \theta} - \frac{\partial B_\theta}{\partial \phi} \right) + \eta \left(\nabla^2 B_\phi + \frac{2}{r^2 \sin \theta} \frac{\partial B_r}{\partial \phi} + \frac{2 \cos \theta}{r^2 \sin^2 \theta} \frac{\partial B_\theta}{\partial \phi} - \frac{B_\phi}{r^2 \sin^2 \theta} \right) \end{aligned}$$

Bibliography:

For more on all this stuff, see

- P.M. Morse et H. Feshbach, *Methods of Theoretical Physics*, McGraw-Hill (1953): Chap. 1
- G.K. Batchelor, *An Introduction to Fluid Dynamics*, Cambridge University Press (1967): Appendice 2
- Arfken, G., *Mathematical Methods for Physicists*, 2^e éd., Academic Press (1970): chap. 2.

or, for a concise introduction, Appendix A.2 of the Goedbloed & Poedts tome cited in the bibliography to chapter 1.

Appendix C: Physical and astronomical constants

{pch_app:constants}

C.1 Physical constants

C.2 Astronomical constants

Appendix D: Maxwell's equations and physical units

{pch_app:maxunits}

Electromagnetism is, unfortunately, a subfield of physics where the choice of units does not only influence the numerical values assigned to measurements, but also the mathematical form of the fundamental laws, i.e., Maxwell's equations.

Physical Quantity	Symbol	Value	Units (SI)
Charge of electron	e	1.602×10^{-19}	C
Mass of electron	m_e	9.109×10^{-31}	kg
Mass of proton	m_p	1.673×10^{-27}	kg
Permittivity of vacuum	ϵ_0	8.854×10^{-12}	
Permeability of vacuum	μ_0	$4\pi \times 10^{-7}$	
Speed of light	c	2.998×10^8	m s ⁻¹
Planck constant	h	6.626×10^{-34}	J s
Boltzmann constant	k_B	1.381×10^{-23}	J K ⁻¹
Stefan-Boltzmann constant	σ	5.670×10^{-8}	J K ⁻⁴ m ⁻² s ⁻¹
Gravitational constant	G	6.671×10^{-11}	m ³ kg ⁻¹ s ⁻²

Astronomical Quantity	Symbol	Value	Units (SI)
Earth mass	M_\oplus	5.977×10^{24}	kg
Earth radius	R_\oplus	6.378×10^6	m
Astronomical Unit	AU	1.496×10^{11}	m
Solar mass	M_\odot	1.989×10^{30}	kg
Solar radius	R_\odot	6.960×10^8	m
Solar luminosity	L_\odot	3.83×10^{26}	J s ⁻¹
Parsec	pc	3.086×10^{16}	m
Light-year	ly	9.461×10^{15}	m

D.1 Maxwell's equations

The whole mess in converting SI units to the astrophysically ubiquitous CGS units all harks back to the definition for the unit of charge, as embodied in Coulomb's Law. Under the SI system we write the electrostatic force between two charges q_1 and q_2 located at positions \mathbf{x}_1 and \mathbf{x}_2 as

$$\{\text{eq:SIcoulomb}\} \quad \mathbf{F} = \frac{1}{4\pi\epsilon_0} \frac{q_1 q_2}{r^2} \hat{\mathbf{r}}, \quad [\text{SI}], \quad (304)$$

with electrical charge measured in coulomb, and with $\mathbf{r} \equiv \mathbf{x}_1 - \mathbf{x}_2$ for notational brevity; whereas under the CGS system the constant $1/4\pi\epsilon_0$ is absorbed into the definition of the unit of charge:

$$\{\text{eq:cgscoulomb}\} \quad \mathbf{F} = \frac{q_1 q_2}{r^2} \hat{\mathbf{r}}, \quad [\text{CGS}], \quad (305)$$

with electrical charge now measured in “electrostatic units”, abbreviated “esu” and sometimes also called “statcoulomb”. In electrostatics it is relatively easy to switch from CGS to SI with the simple substitution $\epsilon_0 \rightarrow 1/(4\pi)$. With electrical currents now measured in esu s⁻¹ in the CGS system, and remembering that $c^2 = (\epsilon_0 \mu_0)^{-1}$, the $\mu_0/4\pi$ prefactor in the Biot-Savart

Law now becomes $1/c$:

$$\{\text{eq:cgSB-S}\} \quad \mathbf{B} = \frac{1}{c} \int \frac{d\boldsymbol{\ell} \times \mathbf{r}}{r^2}, \quad [\text{CGS}], \quad [\text{Biot - Savart}] \quad (306)$$

If you then now go through the process of re-constructing Maxwell's equations under these two new forms for the fundamental relations (electric and magnetic forces), you eventually get to

$$\nabla \cdot \mathbf{E} = 4\pi\rho_e, \quad [\text{Gauss' Law}] \quad (307) \quad \{\text{eq:cgsmx1}\}$$

$$\nabla \cdot \mathbf{B} = 0, \quad [\text{Anonymous}] \quad (308) \quad \{\text{eq:cgsmx2}\}$$

$$\nabla \times \mathbf{E} = -\frac{1}{c} \frac{\partial \mathbf{B}}{\partial t}, \quad [\text{Faraday's Law}] \quad (309) \quad \{\text{eq:cgsmx3}\}$$

$$\nabla \times \mathbf{B} = \frac{4\pi}{c} \mathbf{J} + \frac{1}{c} \frac{\partial \mathbf{E}}{\partial t}, \quad [\text{Ampere/Maxwell's Law}] \quad (310) \quad \{\text{eq:cgsmx4}\}$$

In some sense, the CGS system is perhaps more “natural”, as it omits the introduction of new, apparently fundamental physical constants ϵ_0 and μ_0 , to simply stick with the speed of light c , the only price to pay being an extraneous factor 4π in Gauss' Law. The Lorentz force and Poynting vector become, in CGS units:

$$\mathbf{F} = q(\mathbf{E} + \frac{1}{c} \mathbf{u} \times \mathbf{B}) \quad [\text{Lorentz Force}] \quad (311) \quad \{\text{eq:cgsmx5}\}$$

$$\mathbf{S} = \frac{c}{4\pi} (\mathbf{E} \times \mathbf{B}) \quad [\text{Poynting}] \quad (312) \quad \{\text{eq:cgsmx6}\}$$

and the electrostatic and magnetic energies:

$$\mathcal{E}_e = \frac{1}{8\pi} \int \mathbf{E}^2 dV, \quad (313)$$

$$\mathcal{E}_B = \frac{1}{8\pi} \int \mathbf{B}^2 dV. \quad (314)$$

D.2 Conversion of units

The Table that follows gives you the conversion factor (f) required to go **from SI to cgs units**, i.e., $\text{SI Unit} = f \times \text{cgs units}$. Any “3” appearing in a given value for f is a notational shortcut for 2.99792458.

Table 4 Conversion between SI and CGS units

Quantity	SI name	SI symbol	conversion factor f	CGS name	CGS symbol
Length	meter	m	10^2	centimeter	cm
Mass	kilogram	kg	10^3	gram	g
Force	newton	N	10^5	dyne	dyne
Energy	joule	J	10^7	erg	erg
Charge	coulomb	C	3×10^9	electrostatic units	esu
Current	ampere	A	3×10^9	statampere	esu s ⁻¹
Potential	volt	V	1/300	statvolt	statvolt
Electric field	—	V m ⁻¹	$(1/3) \times 10^{-4}$	—	statvolt cm ⁻¹
Magnetic field	tesla	T	10^4	gauss	G
Magnetic flux	weber	Wb	10^8	maxwell	Mx

For a somewhat humorous close to this rather dry Appendix, here are five different ways, actually to be found in various textbooks or research monographs, to express teslas in terms of other fundamental SI units:

$$1 \text{ T} = 1 \frac{\text{Vs}}{\text{m}^2} = 1 \frac{\text{N}}{\text{A m}} = 1 \frac{\text{kg}}{\text{A s}^2} = 1 \frac{\text{Wb}}{\text{m}^2} = 1 \frac{\text{kg}}{\text{C s}} . \quad (315)$$

Bibliography:

The bulk of this Appendix is taken from Appendix C in

Griffith, D.J., *Introduction to Electrodynamics*, 3rd ed., Prentice Hall (1999),

with some additions extracted primarily from:

Huba, J.D., *NRL Plasma Formulary*, Naval Research Laboratory Publication PU/6790-07-500 (2007).