

FLUID DYNAMICS AND DYNAMOS IN ASTROPHYSICS AND GEOPHYSICS

A series edited by Andrew Soward

University of Exeter UK and

Michael Ghil

University of California, Los Angeles, USA

Founding Editor: Paul Roberts

University of California, Los Angeles, USA

Volume 1

Solar Flare Magnetohydrodynamics

Edited by E. R. Priest

Volume 2

Stellar and Planetary Magnetism

Edited by A. M. Soward

Volume 3

Magnetic Fields in Astrophysics

Ya. B. Zeldovitch, A. A. Ruzmaikin and D. D. Sokoloff

Volume 4

Mantle Convection

Plate tectonics and global dynamics

Edited by W. R. Peltier

Volume 5

Differential Rotation and Stellar Convection

Sun and solar-type stars

G. Rüdiger

Volume 6

Turbulence, Current Sheets and Shocks in Cosmic Plasma

Edited by S. I. Vainshtein, A. M. Bykov and I. N. Toptygin

Volume 7

Earth's Deep Interior

The Doornbos memorial volume

Edited by D. J. Crossley

Volume 8

Geophysical and Astrophysical Convection

Edited by P. A. Fox and R. M. Kerr

Volume 9

Advances in Nonlinear Dynamos

Edited by A. Ferriz-Mas and M. Núñez

Magnetohydrodynamics and The Earth's Core

Selected works of Paul Roberts

Edited by A. M. Soward

Volume 11

Earth's Core and Lower Mantle

Edited by C. A. Jones, A. M. Soward and K. Zhang

Volume 12

Fluid Dynamics and Dynamos in Astrophysics and Geophysics

Edited by A. M. Soward, C. A. Jones, D. W. Hughes and N. O. Weiss

FLUID DYNAMICS AND DYNAMOS IN ASTROPHYSICS AND GEOPHYSICS

EDITED BY

**Andrew M. Soward, Christopher A. Jones,
David W. Hughes and Nigel O. Weiss**

Reviews emerging from the Durham Symposium
on Astrophysical Fluid Mechanics
July 29 to August 8, 2002



CRC PRESS

Boca Raton London New York Washington D.C.

Cover: A Solar X-ray Telescope (SXT) image taken on the Yohkoh satellite. The SXT was jointly developed by science teams in the US and Japan.

Library of Congress Cataloging-in-Publication Data

Fluid dynamics and dynamos in astrophysics and geophysics / edited by Andrew M. Soward ... [et al.].

p. cm. — (Fluid mechanics of astrophysics and geophysics)

Includes bibliographical references and index.

ISBN 0-8493-3355-5 (alk. paper)

1. Fluid dynamics. 2. Dynamo theory (Cosmic physics). 3. Magnetohydrodynamics.

I. Soward, A. M. II. Series.

QB466.F58F58 2004

523.01'886—dc22

2004057907

This book contains information obtained from authentic and highly regarded sources. Reprinted material is quoted with permission, and sources are indicated. A wide variety of references are listed. Reasonable efforts have been made to publish reliable data and information, but the author and the publisher cannot assume responsibility for the validity of all materials or for the consequences of their use.

Neither this book nor any part may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, microfilming, and recording, or by any information storage or retrieval system, without prior permission in writing from the publisher.

All rights reserved. Authorization to photocopy items for internal or personal use, or the personal or internal use of specific clients, may be granted by CRC Press, provided that \$.50 per page photocopied is paid directly to Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923 USA. The fee code for users of the Transactional Reporting Service is ISBN 0-8493-3355-5/05/\$0.00+\$1.50. The fee is subject to change without notice. For organizations that have been granted a photocopy license by the CCC, a separate system of payment has been arranged.

The consent of CRC Press does not extend to copying for general distribution, for promotion, for creating new works, or for resale. Specific permission must be obtained in writing from CRC Press for such copying.

Direct all inquiries to CRC Press, 2000 N.W. Corporate Blvd., Boca Raton, Florida 33431.

Trademark Notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation, without intent to infringe.

Visit the CRC Press Web site at www.crcpress.com

© 2005 by CRC Press

No claim to original U.S. Government works

International Standard Book Number 0-8493-3355-5

Library of Congress Card Number 2004057907

Printed in the United States of America 1 2 3 4 5 6 7 8 9 0

Printed on acid-free paper

Contents

Contributors

Preface

Acknowledgments

1 Accretion discs

Gordon I. Ogilvie

1. Introduction
2. Classical theory of accretion discs
3. Angular momentum transport
 - 3.1. The need for an effective viscosity
 - 3.2. The shearing sheet
 - 3.3. Stability of Keplerian discs
 - 3.4. The magnetorotational instability
4. Continuum celestial mechanics
 - 4.1. As a generalization of the classical theory of accretion discs
 - 4.2. As a generalization of classical celestial mechanics
 - 4.3. In relation to continuum mechanics
 - 4.4. Constitutive equations for turbulent accretion discs

2 Discs with MHD turbulence and their response to orbiting planets

John C.B. Papaloizou and Richard P. Nelson

1. Introduction
2. Set up of initial models
 - 2.1. Global models
 - 2.2. Local simulations
 - 2.3. Vertically, horizontally and time averaged stresses, angular momentum transport and energy dissipation

3. Simulation results
 - 3.1. The time averaged radial velocity in the global models
 - 3.2. Interaction with a planet in fixed circular orbit
 - 3.3. Angular momentum transport and gap maintenance
 - 3.4. A local simulation
4. Discussion

3 Mixing at the surface of white dwarf stars

R. Rosner and A. Alexakis

1. Introduction
2. Previous work on mixing
 - 2.1. The early studies
 - 2.2. More recent ideas about surface mixing
3. An alternative: mixing at boundaries by “wind”-driven resonant modes
 - 3.1. The prototypical problem
 - 3.2. The linear stability problem
 - 3.3. The underlying physics and its relation to the astrophysics
 - 3.4. Modeling the mixing limit: the nonlinear regime
 - 3.5. Extending our results
4. Conclusions

4 Pulsar magnetospheres

Leon Mestel

1. Introduction
2. Basic electrodynamics
 - 2.1. An illustrative example
 - 2.2. A live classical model
3. Models with pair production
 - 3.1. The electron-positron plasma
 - 3.2. The relativistic force-free equation
4. Live models
 - 4.1. The outer domain
5. The $S(P)$ relation
6. The inner domain

7. Computational results
8. Alternative models
9. Magnetospheric gaps

5 Magnetic fields in galaxies

Ellen G. Zweibel

1. Introduction
2. Galactic structure
3. Observations of Galactic magnetic fields
4. Galaxy formation and magnetogenesis
5. Building blocks of Galactic magnetic fields
 - 5.1. Differential rotation and turbulence
 - 5.2. Microscopic state of the gas
 - 5.3. Cosmic rays
 - 5.4. Magnetic reconnection
6. Summary and outlook

6 Self-consistent mean field electrodynamics in two and three dimensions

P. H. Diamond, D. W. Hughes and Eun-jin Kim

1. Introduction
2. Turbulent diffusion of magnetic fields
 - 2.1. Overview
 - 2.2. Flux diffusion in 2D—basic model and concepts
 - 2.3. Mean field electrodynamics for $\langle A \rangle$ in 2D
 - 2.4. Turbulent diffusion of flux and field in 3D
 - 2.5. Discussion and conclusion
3. The generation of magnetic fields
 - 3.1. The linear regime
 - 3.2. Small-scale dynamo action
 - 3.3. The nonlinear regime
 - 3.4. The role of boundary conditions
4. Momentum and flux transport in 2D MHD
 - 4.1. Overview
 - 4.2. Mean field theory
 - 4.3. The effect of shear flows on transport
 - 4.4. The effect of shear flows on flux diffusion in 2D MHD

- 4.5. Effect of magnetic fields and shear on momentum transport
- 4.6. Concluding remarks

7 The solar tachocline: Formation, stability and its role in the solar dynamo

S. M. Tobias

1. Introduction
2. Properties of the solar tachocline
3. Why is the solar tachocline so thin?
4. Instabilities of the solar tachocline
 - 4.1. Differential rotation and joint instabilities
 - 4.2. Magnetic buoyancy instabilities
5. The tachocline and the solar dynamo
 - 5.1. The interface dynamo
6. Discussion and conclusion

8 Magnetoconvection

M.R.E. Proctor

1. Introduction
2. The induction equation and the Lorentz force
3. Boussinesq magnetoconvection in a vertical field
 - 3.1. Linear theory
4. Nonlinear Boussinesq magnetoconvection
 - 4.1. Pattern selection near onset
 - 4.2. Fully nonlinear convection
5. Strong magnetic fields
6. Shearing instabilities and streaming flows
7. Convection in extended regions
 - 7.1. Symmetry breaking instabilities and convectons
 - 7.2. Modulational instabilities
8. Effects of stratification
 - 8.1. “Fundamental” models
 - 8.2. Nonlinear pumping
 - 8.3. Oblique fields and traveling waves
 - 8.4. “Realistic” models
9. Discussion

9 Alfvén waves within the Earth's core

D. Jault and G. Légaut

1. Introduction
2. Propagation of Alfvén waves within an entirely fluid spherical body
3. Topographical effects at the CMB
4. Influence of a conducting inner core
5. Concluding remarks

10 Turbulence models and plane layer dynamos

C.A. Jones and P.H. Roberts

1. Introduction
2. Background to the LANS α -models
3. Model description and governing equations
4. The hyperdiffusive and the LANS α -models
5. Subgrid scale boundary conditions
6. Linear theory
7. Numerical simulations
8. Conclusions

11 Planetary and stellar dynamos: challenges for next generation models

Gary A. Glatzmaier

1. Introduction
2. Current 3D geodynamo simulations
3. Laminar approximations of turbulent flows
4. 2D Turbulent magnetoconvection simulations
 - 4.1. The model
 - 4.2. The results
5. Discussion

12 Convection in rotating spherical fluid shells and its dynamo states

F. H. Busse and R. Simitev

1. Introduction
2. Mathematical formulation of the problem

3. Convection in rotating spherical shells
4. Evolution of convection columns at small Prandtl numbers
5. Convection at very small Prandtl numbers
6. Convection driven dynamos at moderate Prandtl numbers
7. Convection driven dynamos in low Prandtl number fluids
8. Torsional oscillations
9. Oscillating dipolar dynamos
10. Discussion and outlook

13 Laboratory experiments on liquid metal dynamos and liquid metal MHD turbulence

Stephan Fauve and Daniel P. Lathrop

1. Introduction
2. Self-generation in Riga and Karlsruhe
 - 2.1. From the Ponomarenko dynamo to the Riga experiment
 - 2.2. From the Roberts flow to the Karlsruhe experiment
 - 2.3. Expected and unexpected results: the nature of the dynamo bifurcation
 - 2.4. Expected and unexpected results: sensitivity of the dynamo threshold to flow perturbations
 - 2.5. Expected and unexpected results: scaling law for the magnetic energy generation above onset
3. Generating a fully three-dimensional turbulent flow in a compact volume
 - 3.1. Maryland 30 cm experiment
 - 3.2. VKS experiment
4. Motivations for rapidly rotating flows
 - 4.1. Astrophysical and geophysical dynamos
 - 4.2. Dynamo bifurcations in rapidly rotating flows
 - 4.3. Scalings of magnetic energy
5. Concluding remarks

Contributors

Alexandros Alexakis

NCAR—Advanced Study Program
Boulder, Colorado, USA
E-mail: alexakis@ucar.edu

Friedrich H. Busse

Physikalisches Institut, Universität Bayreuth
Lehrstuhl Theoretische Physik IV
Bayreuth, Germany
E-mail: Friedrich.Busse@uni-bayreuth.de

Patrick H. Diamond

Department of Physics
University of California, San Diego
La Jolla, California, USA
E-mail: pdiamond@physics.ucsd.edu

Stephan Fauve

Laboratoire de Physique Statistique de l'Ecole Normale Supérieure
CNRS UMR 8550
Paris, France
E-mail: Stephan.Fauve@lps.ens.fr

Gary A. Glatzmaier

Earth Sciences Department
University of California
Santa Cruz, California, USA
E-mail: glatz@es.ucsc.edu

David W. Hughes

Department of Applied Mathematics
University of Leeds
Leeds, UK
E-mail: d.w.hughes@amsta.leeds.ac.uk

Dominique Jault

Laboratoire de Géophysique Interne et Tectonophysique - LGIT
Université Joseph Fourier de Grenoble
Grenoble, France
E-mail: Dominique.Jault@ujf-grenoble.fr

Christopher A. Jones

Department of Mathematical Sciences
University of Exeter
Exeter, UK
E-mail: C.A.Jones@ex.ac.uk

Eun-jin Kim

Department of Physics
University of California, San Diego
La Jolla, California, USA
E-mail: ejk@physics.ucsd.edu

Daniel P. Lathrop

IREAP, IPST, and the Department of Physics
University of Maryland
College Park, Maryland, USA
E-mail: dp1@complex.umd.edu

Gédéon Légaut

Laboratoire de Géophysique Interne et Tectonophysique - LGIT
Université Joseph Fourier de Grenoble
Grenoble, France
E-mail: Gedeon.Legaut@obs.ujf-grenoble.fr

Leon Mestel

Astronomy Centre, SciTech
University of Sussex
Falmer, Brighton, UK
E-mail: lmestel@sussex.ac.uk

Richard P. Nelson

Astronomy Unit, School of Mathematical Sciences
Queen Mary, University of London
London, UK
E-mail: R.P.Nelson@qmul.ac.uk

Gordon I. Ogilvie

Institute of Astronomy
University of Cambridge
Cambridge, UK
E-mail: gogilvie@ast.cam.ac.uk

John C.B. Papaloizou

Astronomy Unit, School of Mathematical Sciences
Queen Mary, University of London
London, UK
E-mail: j.c.b.papaloizou@qmw.ac.uk

Michael R.E. Proctor

Department of Applied Mathematics and Theoretical Physics
University of Cambridge, Centre for Mathematical Sciences
Cambridge, UK
E-mail: M.R.E.Proctor@damtp.cam.ac.uk

Paul H. Roberts

Institute of Geophysics and Planetary Physics
University of California, Los Angeles
Los Angeles, California, USA
E-mail: roberts@math.ucla.edu

Robert Rosner

Departments of Physics and Astronomy & Astrophysics,
and Enrico Fermi Institute
The University of Chicago
Chicago, Illinois, USA
E-mail: r-rosner@uchicago.edu

Radostin Simitev

Physikalisches Institut, Universität Bayreuth
Lehrstuhl Theoretische Physik IV
Bayreuth, Germany
E-mail: radostin.simitev@uni-bayreuth.de

Steven M. Tobias

Department of Applied Mathematics
University of Leeds
Leeds, UK
E-mail: smt@maths.leeds.ac.uk

Ellen G. Zweibel

Departments of Astronomy and Physics,
and Center for Magnetic Self-Organization
University of Wisconsin
Madison, Wisconsin, USA
E-mail: zweibel@astro.wisc.edu



The participants at the LMS Durham Symposium on Astrophysical Fluid Mechanics 29 July to 8 August 2002. (Courtesy of the Department of Mathematical Sciences, The University of Durham, Durham, UK.) Color image can be found at: www.maths.dur.ac.uk/lms/pics/2002AFM.jpg. For a key of participants, please visit: www.maths.dur.ac.uk/lms/pics/2002AFMkey.jpg.

Preface

Astrophysical fluid dynamics is a research area that has developed rapidly in recent years. This is mainly due to the availability of increasingly powerful computer resources, allowing numerical simulations to explore previously uncharted territory, but large improvements in observational data have also had a major impact. The subject spans a wide range, encompassing three distinct yet overlapping communities. Their interests may be loosely identified as (i) accretion discs and high-energy astrophysics; (ii) solar, stellar and galactic magnetic fields; (iii) the geodynamo, planetary magnetic fields and associated experiments. It is not often that all three of these communities come together at the same meeting. Nevertheless, the London Mathematical Society (with the support of EPSRC) provided the editors of this book with such an opportunity in the summer of 2002 by inviting us to be the Scientific Organizers of the 10-day LMS Durham Symposium titled “Astrophysical Fluid Mechanics.” The symposium was well attended by many of the world’s leading experts in the various research areas (see the group photograph), and a lively interchange of ideas ensued. The purpose of this book is to review the current status and provide a balanced coverage of the subject areas mentioned.

Magnetic fields are all-pervasive in astrophysics, giving rise to a host of spectacular phenomena. Through the Lorentz force they influence the dynamical behavior of astrophysical bodies with the possibility of radically affecting their structure and evolution. The equations of magnetohydrodynamics (MHD), which govern the behavior of electrically conducting fluids, are, in general, extremely difficult to solve, and many important issues such as MHD turbulence still present significant theoretical challenges. New observations in objects such as accretion discs and pulsars, however, can suggest new theoretical ideas in MHD, and the theory in turn can influence the development of more realistic models. This means that a better understanding of the fundamental physics of MHD processes progresses alongside the study of these astrophysical objects; many of the chapters in this book illustrate this symbiotic relation between theory and observation.

The dynamo problem, namely, obtaining an understanding of the ability of fluid motions to maintain a magnetic field against its natural tendency to decay, remains one of the most interesting and controversial

topics in astrophysical MHD. In different guises it lies at the very heart of studies of the magnetic fields found in planets, stars, galaxies and accretion discs and so provides a central theme of this book. An important distinguishing feature of these various problems is the size of the magnetic Reynolds number R_m , a measure of the relative importance of advection to diffusion of the magnetic field. It is generally large in the astrophysical context but moderate in the planetary context, this difference accounting for the dynamos in these two cases having quite distinct characteristics. The high R_m limit, in particular, raises some fundamental questions about the strength of large-scale, dynamo-generated fields, an issue not yet fully resolved.

An exciting development in recent years has been an upsurge in the number of relatively low R_m MHD laboratory experiments, which is leading to a deeper understanding of MHD turbulence and the dynamics of rapidly rotating convection. Also, following many decades of unsuccessful attempts, the first working fluid dynamos have been constructed in the last few years. Some of these exciting advances are reviewed here.

Acknowledgments

We are grateful to the London Mathematical Society for inviting the Editors of this book to be the Scientific Organizers of LMS Durham Symposium on Astrophysical Fluid Mechanics, 29 July to 8 August 2002, which was supported by EPSRC. We are particularly grateful to the Department of Mathematical Sciences at the University of Durham for their hospitality and organizational support which ensured the smooth running and success of the symposium.

1 Accretion discs

Gordon I. Ogilvie

Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK

A thin disc of gas in Keplerian orbital motion around a central mass is one of the most characteristic fluid flows in astrophysics. After briefly reviewing the classical theory of accretion discs I discuss a selection of the outstanding questions and recent developments in this field. Particular attention is given to the role of magnetic fields, the possible onset of turbulent motion in the disc, and the modeling of the resulting stresses. I also describe progress toward a general theory of Keplerian discs, in which the orbits are permitted to be eccentric and/or mutually inclined.

1. Introduction

Since the time of Copernicus we have understood that the planets of the solar system orbit the Sun in nearly circular and coplanar paths. Based on this observation, Kant (1755) proposed that the planets formed out of a thin disc that surrounded the Sun earlier in its history. In Kant's theory, as in the currently favored paradigm, particles in the disc followed circular Keplerian orbits around the Sun and aggregated into successively larger bodies that eventually became planets.

Less than a decade ago, such protoplanetary discs were directly imaged for the first time by the Hubble Space Telescope, surrounding young stars in the Orion Nebula (McCaughrean and O'Dell, 1996). The first discovery of a planet orbiting an extrasolar main-sequence star also dates from this time (Mayor and Queloz, 1995), and more than one hundred examples are now known. The formation of a centrifugally supported disc is understood to be an essential part of the process by which a star forms from a slowly rotating cloud of gas that collapses under its own gravitation.

Discs are also found in close binary stars in which one star captures matter lost by its companion through a stellar wind or an overflow of its Roche lobe. If the recipient is sufficiently compact, the captured

matter has too much angular momentum to fall directly on to the star and instead forms an accretion disc around it. Matter is accreted if it loses angular momentum through a torque acting on or within the disc. Systems involving accretion on to a white dwarf (e.g., Warner, 1995) include the classical novae and dwarf novae, which have been known for more than one hundred years but lacked a physical explanation until more recently. Those involving accretion on to a neutron star or black hole (e.g., Lewin *et al.*, 1995) were first revealed by rocket-borne X-ray detectors in the early 1960s.

Accretion of gas through a disc on to a black hole, but having millions or billions of times the mass of the Sun, powers the intense luminosity of active galactic nuclei and quasars (Lynden-Bell, 1969). Spiral galaxies themselves differ from gaseous accretion discs in that the principal, stellar component forms an almost collisionless system that cannot be regarded as a fluid. Furthermore, the time-scale associated with accretion processes in the gaseous component generally exceeds the age of the Universe.

There are many other examples of astrophysical discs. Saturn's rings are extremely thin discs composed of icy "boulders" up to a few meters in size, which undergo very gentle collisions (e.g., Gor'kavyi and Fridman, 1994). Rapidly rotating early-type main-sequence stars, known as Be stars, are surrounded by an equatorial decretion disc that is expelled through the action of torque (Lee *et al.*, 1991). Discs are also formed when compact binary stars, consisting for example of two white dwarfs, or a neutron star and a black hole, spiral in and merge as a result of gravitational radiation (e.g., Benz *et al.*, 1990).

A thin disc of gas in Keplerian orbital motion around a central mass is therefore one of the most characteristic fluid flows in astrophysics. Indeed, the physics of discs is sufficiently general that many theoretical aspects can be investigated without detailed reference to the specific environment in which the disc is found.

In this article I discuss a selection of the outstanding questions and recent developments in this field. I begin by briefly reviewing the classical theory of accretion discs (Section 2). I then discuss the central issue of angular momentum transport in discs (Section 3). Finally, I describe progress toward a general theory of Keplerian discs, in which the orbits are permitted to be eccentric and/or mutually inclined (Section 4).

2. Classical theory of accretion discs

In a thin accretion disc the collective effects of the fluid, such as pressure, magnetic fields, turbulence, self-gravitation or viscosity, are all

weak compared to the dominant balance between the inertia of the flow and the gravitational attraction of the central mass. To a first approximation, therefore, fluid elements attempt to follow the same trajectories as test particles in the gravitational potential. For this reason, the study of astrophysical discs combines ideas from fluid dynamics and celestial mechanics, a synthesis that may be termed “continuum celestial mechanics.”

For the most important example of a point-mass potential in Newtonian theory, the trajectories of test particles are, of course, Keplerian orbits. In the classical theory of accretion discs these orbits are assumed to be circular and coplanar. Not only is this the simplest situation, such orbits also have the least energy for a given value of the vertical component of angular momentum and therefore are often considered a natural outcome of dissipative processes.

The evolution of a standard accretion disc is regulated by the laws of conservation of mass and conservation of angular momentum. It is convenient to extract these in one-dimensional forms from the three-dimensional equation of mass conservation

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{u}) = 0 \quad (1)$$

and equation of motion

$$\frac{\partial}{\partial t}(\rho \mathbf{u}) + \nabla \cdot (\rho \mathbf{u} \mathbf{u}) = -\rho \nabla \Phi + \nabla \cdot \mathbf{W}. \quad (2)$$

Here ρ and \mathbf{u} are the density and velocity of the fluid and Φ is the external gravitational potential, usually dominated by the central mass. The stress tensor \mathbf{W} , which is a symmetric tensor field of second rank, represents all the collective effects of the fluid.

Let (r, ϕ, z) be cylindrical polar coordinates, and let C_r be the cylinder of radius r . Integrating equation (1) over C_r yields

$$2\pi r \frac{\partial \Sigma}{\partial t} - \frac{\partial F}{\partial r} = 2\pi r S, \quad (3)$$

where $\Sigma(r, t)$ is the surface mass density (azimuthally averaged if necessary), $F(r, t)$ is the inward mass flux through C_r and $S(r, t)$ is the net source of mass, if any, per unit area. In the case of an outflow from the disc, $S < 0$.

If we assume that the angular velocity $\Omega = u_\phi/r$ depends only on r , a similar operation applied to r times the azimuthal component of (2) yields

$$2\pi rh \frac{\partial \Sigma}{\partial t} - \frac{\partial}{\partial r}(F h) = \frac{\partial G}{\partial r} + 2\pi r T, \quad (4)$$

where $h(r) = r^2\Omega$ is the specific angular momentum, $G(r, t)$ is the inward angular momentum flux through C_r and $T(r, t)$ is the external torque, if any, per unit area (including angular momentum added along with the mass source S). Note that

$$G = \iint r^2 W_{r\phi} \, d\phi \, dz \quad (5)$$

represents the internal torque of the disc. It is conventional to parameterize G as if it resulted from a kinematic viscosity $\nu(r, t)$, i.e.,

$$G = 2\pi\nu\Sigma r^3 \frac{d\Omega}{dr}. \quad (6)$$

Formally, such a parameterization is always possible, although it is problematic if there is a point in the flow at which $d\Omega/dr = 0$ (as might well occur in a star-disc boundary layer) while $G \neq 0$. It is to be expected on energetic grounds that $\nu \geq 0$ in general.

F can then be eliminated to yield an evolutionary equation for $\Sigma(r, t)$ in the form

$$\frac{\partial \Sigma}{\partial t} + \frac{1}{r} \frac{\partial}{\partial r} \left\{ \left(\frac{dh}{dr} \right)^{-1} \left[\frac{\partial}{\partial r} \left(\nu \Sigma r^3 \frac{d\Omega}{dr} \right) + r(T - S h) \right] \right\} = S. \quad (7)$$

In the important case of Keplerian orbits, for which $\Omega \propto r^{-3/2}$ and $h \propto r^{1/2}$, and in the absence of source or sink terms, this equation reduces to

$$\frac{\partial \Sigma}{\partial t} = \frac{3}{r} \frac{\partial}{\partial r} \left[r^{1/2} \frac{\partial}{\partial r} (r^{1/2} \nu \Sigma) \right], \quad (8)$$

which has the character of a diffusion equation.

To understand the reason for this, suppose that the disc initially takes the form of a narrow ring. The inner part of the ring has a slightly smaller specific angular momentum than the outer part, although it rotates slightly more rapidly. Any process that acts like a viscosity will transport angular momentum from the inner part of the ring to the outer part, thereby increasing the width of the distribution of specific angular momentum. In order to remain in circular Keplerian orbital motion, the fluid must spread out.

Only one assumption was required to derive (7), namely that the angular velocity depends only on r . This can be justified for a standard thin disc in which the characteristic ratio of semi-thickness to radius, H/r , is much less than unity. In such a disc the radial component of the equation of motion requires that the angular velocity equal that of a circular

test-particle orbit, plus a small fractional correction of order $(H/r)^2$. If the azimuthal velocity differs from a simple orbital motion, for example if the disc contains waves or turbulent motion, then any corresponding $r\phi$ -component of the Reynolds stress tensor should be included in $W_{r\phi}$.

A fundamental problem in accretion disc theory has been to identify efficient mechanisms of angular momentum transport that could account for ν (i.e., $W_{r\phi}$), as the molecular viscosity falls short by many orders of magnitude. Noting that the shear stress $W_{r\phi}$ has the same dimensions as the pressure p , Shakura and Sunyaev (1973) introduced the dimensionless viscosity parameter α such that

$$W_{r\phi} = -\alpha p. \quad (9)$$

If we assume that α is a universal constant of order unity, a consideration of the vertical hydrostatic and thermal equilibrium of the disc makes it possible to obtain ν as a function of r and Σ . The vertical structure of a disc is analogous to the radial structure of a star, and depends significantly on the equation of state and the opacity of the gas. For example, if the disc is thin, Keplerian, composed of fully ionized hydrogen that can be described as an ideal gas with negligible radiation pressure, and if the vertical heat flux is carried by radiation under optically thick conditions dominated by Thomson opacity, it can be shown that

$$\nu = C_\nu \alpha^{3/2} \Omega^{-2/3} \Sigma^{2/3}, \quad (10)$$

where C_ν is a combination of the fundamental constants. The alpha viscosity prescription therefore allows the diffusion (8) to be solved and, interestingly, renders it nonlinear.

The basic equations governing viscous accretion discs were formulated by von Weizsäcker (1948) and Lüst (1952), who found steady and time-dependent solutions, respectively, of the diffusion (8) with a simple viscosity prescription. The time-dependent problem was studied more generally by Lynden-Bell and Pringle (1974) and Pringle (1974), while the properties of steady accretion discs were investigated in detail by Lynden-Bell (1969), Novikov and Thorne (1973) and Shakura and Sunyaev (1973). The review articles by Pringle (1981) and Papaloizou and Lin (1995) provide an excellent introduction to the subject.

The classical theory of accretion discs leaves a number of questions unanswered, of which some of the most important are as follows.

1. What provides the effective viscosity of accretion discs? Does it behave according to the alpha prescription, and, if so, what is the value of α ?

2. What happens if the gas in an accretion disc does not follow circular and coplanar orbits, as is usually assumed?
3. What happens if the disc is not thin?

The first question has been a central problem of accretion disc theory since its conception, and Section 3 of this article is devoted to it. The second question, which has received much less attention, is treated in Section 4. The third question relates to the subject of “quasi-spherical”, “non-radiative” or “advection-dominated” accretion flows and is discussed by Blandford (this volume).

3. Angular momentum transport

3.1. The need for an effective viscosity

In principle, an accretion disc can operate without a shear stress $W_{r\phi}$ if angular momentum is removed vertically from the disc by a magnetized outflow. Consider (7) in the case of a Keplerian disc with $v = 0$, but allowing for a removal of mass at a rate $-S > 0$ per unit area and of angular momentum at a rate $-T = -qS h$ per unit area. For a purely hydrodynamic outflow in which the specific angular momentum of fluid elements is conserved, $q = 1$. However, by enforcing an approximate corotation of the outflow as far as the Alfvén surface, a magnetic field allows a more efficient removal of angular momentum, with $q \gg 1$. In the case of a steady, axisymmetric outflow in ideal magnetohydrodynamics, $q = [r_A(r)/r]^2$, where $r_A(r)$ is the cylindrical radius of the Alfvén point on the magnetic field line having its foot-point on the disc at radius r (Mestel, 1968). Furthermore, the magnetic field can provide great assistance in accelerating the outflow from the surface of the disc (Blandford and Payne, 1982).

Equation (7) then reduces to

$$\frac{\partial \Sigma}{\partial t} + \frac{1}{r} \frac{\partial}{\partial r} [2(q-1)r^2 S] = S. \quad (11)$$

If mass is supplied steadily to the disc at its outer radius r_{out} , a stationary solution is possible in which

$$-S \propto r^{-2+\delta}, \quad \delta = \frac{1}{2(q-1)}, \quad (12)$$

assuming for simplicity that q is independent of r . In this solution the inward mass flux $F \propto r^\delta$. The accretion rate at the inner radius r_{in} is smaller than the mass supply rate by a factor $(r_{\text{in}}/r_{\text{out}})^\delta$, which differs little from unity if $\delta \ll 1$.

This mechanism is an attractive one, not least because it has the potential to explain both accretion and outflow, which are almost invariably associated in observed systems. However, most models of magnetized outflows rely on the existence of a large-scale poloidal magnetic field threading the disc. It is not clear that such a field can be assembled and maintained against dissipation or turbulent expulsion (Lubow *et al.*, 1994a). In addition, there are concerns over the stability of outflow-driven accretion (Lubow *et al.*, 1994b). Consequently, it is still very uncertain whether a magnetized outflow can provide an effective removal of angular momentum over many decades of radius.

Outward angular momentum transport in the plane of the disc is associated with a negative shear stress $W_{r\phi}$. Such a stress can be provided, in principle, by fluid motions (the Reynolds stress), magnetic fields (the Maxwell stress) or non-axisymmetric gravitational fields (Lynden-Bell and Kalnajs, 1972). In each case the stress could be associated either with large-scale spiral structures (shocks, magnetic arms or density waves) or with small-scale features (waves or turbulence). Many of these possibilities are discussed in the review article by Papaloizou and Lin (1995). Large-scale spiral structures typically require specific formation mechanisms and are difficult to sustain over many orbital periods in a differentially rotating disc. Small-scale features, such as turbulent motion resulting from a local instability, are generally considered more promising as possible mechanisms of angular momentum transport.

3.2. *The shearing sheet*

The shearing sheet is a very useful local model of a differentially rotating disc. Originally conceived with application to the stellar dynamics of spiral galaxies (Goldreich and Lynden-Bell, 1965), it has rightly acquired fundamental importance in both analytical and numerical studies of accretion discs. By formally separating the rotation and shear of the disc, straightening the streamlines and removing the horizontal boundaries to infinity, it creates the simplest realistic environment in which to carry out local stability analyses or studies of turbulence in discs. I recount some of the properties of this important model here.

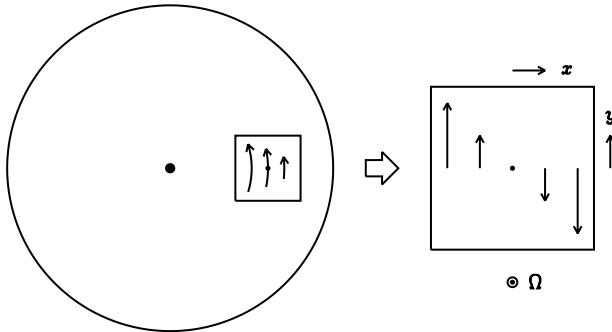


Figure 1. The shearing sheet. The differential rotation of the disc is represented locally as a uniform rotation plus a linear shear flow.

In a differentially rotating disc with angular velocity $\Omega(r)$, the quantity

$$A(r) = -\frac{r}{2} \frac{d\Omega}{dr} \quad (13)$$

measures the shear rate. Consider a reference point (Fig. 1), situated in the mid-plane of the disc and orbiting the central mass in a circular orbit of radius r_0 and angular velocity $\Omega_0 = \Omega(r_0)$. It is used as the origin of a local, rotating Cartesian coordinate system (x, y, z) , with unit vectors $(\mathbf{e}_x, \mathbf{e}_y, \mathbf{e}_z)$ pointing in the radial, azimuthal and vertical directions, respectively. The flow is represented locally as a uniform rotation $\Omega_0 \mathbf{e}_z$ plus a linear shear flow

$$\mathbf{u}_0 = -2A_0 x \mathbf{e}_y, \quad (14)$$

where $A_0 = A(r_0)$ is known as Oort's first constant in galactic dynamics.

The shearing sheet is considered to be horizontally unbounded, corresponding to the physical situation when we view a thin disc on a scale comparable to its thickness. The vertical dimension can be treated in different ways. It can be neglected altogether, in the sense that the strictly two-dimensional dynamical equations are employed. If the fluid is treated as incompressible and of uniform density, three-dimensional equations can be retained without the complications of vertical stratification. A similar simplification is possible for a compressible fluid if the vertical component of the gravitational acceleration is neglected. Alternatively, a full treatment of the vertical structure is possible.

In this connection it is relevant to point out that, although very many authors have represented the dynamics of a thin accretion disc by

a naive application of the two-dimensional fluid dynamical equations, there are almost no circumstances in which such a description can be formally justified (Gor'kavyi and Fridman, 1994).

The shearing sheet has a much higher degree of symmetry than the original disc. First, every point in the xy -plane is equivalent, modulo a Galilean boost. This means that the system is horizontally homogeneous and, if it becomes turbulent through an instability, the turbulence can be expected to be statistically horizontally homogeneous. Second, the system is invariant under a rotation by π about the z -axis. This means that the system cannot distinguish whether the central mass lies in the $-x$ direction or the $+x$ direction.

The equation of motion in the rotating frame may be written

$$\left(\frac{\partial}{\partial t} + \mathbf{u} \cdot \nabla \right) \mathbf{u} + 2\Omega_0 \mathbf{e}_z \times \mathbf{u} = -\nabla \Phi^{\text{cg}} + \frac{1}{\rho} \nabla \cdot \mathbf{W}, \quad (15)$$

where $\Phi^{\text{cg}} = \Phi - r^2\Omega_0^2/2$ is the centrifugal–gravitational potential. In a centrifugally supported disc, $\partial\Phi/\partial r = r\Omega^2$ and so

$$\frac{\partial \Phi^{\text{cg}}}{\partial r} = r(\Omega^2 - \Omega_0^2) = -4\Omega_0 A_0 x + O(x^2). \quad (16)$$

In the shearing sheet, the shear flow is represented in a linear approximation and therefore the $O(x^2)$ remainder is neglected. Also $\partial\Phi/\partial z$ is approximated by $\Omega_{z0}^2 z$, where $\Omega_z(r)$ is the vertical oscillation frequency given by $\Omega_z^2 = \Phi_{,zz}(r, 0)$. By writing the total velocity as $\mathbf{u} = -2A_0 x \mathbf{e}_y + \mathbf{v}$, where \mathbf{v} is the velocity perturbation, we then obtain the equation of motion

$$\left(\frac{\partial}{\partial t} - 2Ax \frac{\partial}{\partial y} + \mathbf{v} \cdot \nabla \right) \mathbf{v} + 2\Omega \mathbf{e}_z \times \mathbf{v} = -\Omega_z^2 z \mathbf{e}_z + \frac{1}{\rho} \nabla \cdot \mathbf{W}, \quad (17)$$

where the subscript zero has been omitted. The Coriolis force acting on the background shear flow has cancelled with the $\partial\Phi^{\text{cg}}/\partial x$ term in (15).

The basic state of the shearing sheet is $\mathbf{v} = \mathbf{0}$, with hydrostatic equilibrium in the vertical direction. Owing to the horizontal homogeneity of the sheet, any stress associated with the basic state, such as a viscous stress acting on the background shear flow, is independent of x and y and therefore does not disturb the horizontal components of the equation of motion. In particular, such a stress does not cause an accretion flow to develop. This is connected with the fact mentioned above, that the shearing sheet cannot distinguish in which direction the central mass lies.

In numerical studies it is impractical to deal with a system of unbounded spatial extent. The most natural way to truncate the system is to solve the equations in a finite horizontal domain, $0 < x < L_x$, $0 < y < L_y$, and to apply the boundary conditions

$$f(L_x, y, z, t) = f(0, y', z, t), \quad f(x, L_y, z, t) = f(x, 0, z, t), \quad (18)$$

where f represents any component of \mathbf{v} or any other physical quantity, and $y' = (y + 2AL_x t) \bmod L_y$. This is the “shearing box” (Hawley *et al.*, 1995). As for the shearing sheet, the vertical dimension can be treated in various ways with periodic or other boundary conditions.

The meaning of the horizontal boundary conditions is that all quantities f are periodic in coordinates that shear with the background flow, a technique introduced by Rogallo in studies of turbulent shear flow (see, e.g., Pumir, 1996) and used by Wisdom and Tremaine (1988) in the context of planetary rings. For analytical studies they have the important property that the linearized equations admit solutions in the form of the sheared plane waves,

$$f = \operatorname{Re} \left\{ \tilde{f}(z, t) \exp [ik_x(t)x + ik_y y] \right\}, \quad (19)$$

first employed by Lord Kelvin (Thomson, 1887). Here \tilde{f} is a complex wave amplitude and \mathbf{k} is a time-dependent horizontal wavevector that evolves according to $k_x = k_x(0) + 2Ak_y t$ as the background shear flow tilts the plane wavefronts. Such solutions are also admitted in the unbounded shearing sheet, where \mathbf{k} is otherwise unrestricted. In the shearing box \mathbf{k} is quantized such that $k_x(0) = 2n_x\pi/L_x$ and $k_y = 2n_y\pi/L_y$, for integers n_x and n_y ; the allowed wavevectors therefore occupy a shearing lattice in Fourier space. Kelvin’s waves also form a natural basis for treatments of the nonlinear equations and, in particular, for studies of turbulence.

3.3. Stability of Keplerian discs

Circular test-particle orbits around a point mass are stable in Newtonian dynamics. Moving in an axisymmetric gravitational potential $\Phi(r, z)$, a particle with conserved specific angular momentum h experiences an effective potential

$$\Phi^{\text{eff}}(r, z) = \Phi(r, z) + \frac{h^2}{2r^2}, \quad (20)$$

such that its equation of motion is

$$\frac{d^2r}{dt^2} = -\frac{\partial \Phi^{\text{eff}}}{\partial r}, \quad \frac{d^2z}{dt^2} = -\frac{\partial \Phi^{\text{eff}}}{\partial z}. \quad (21)$$

If Φ is symmetric about $z = 0$, a circular orbit can occur in the mid-plane if $\partial \Phi^{\text{eff}}/\partial r(r, 0) = 0$, which implies $h^2 = r^3 \partial \Phi/\partial r(r, 0)$. Infinitesimal perturbations about this orbit satisfy

$$\frac{d^2\delta r}{dt^2} = -\kappa^2 \delta r, \quad \frac{d^2\delta z}{dt^2} = -\Omega_z^2 \delta z, \quad (22)$$

where

$$\kappa^2 = \frac{\partial^2 \Phi^{\text{eff}}}{\partial r^2}(r, 0) = \frac{\partial^2 \Phi}{\partial r^2}(r, 0) + \frac{3}{r} \frac{\partial \Phi}{\partial r}(r, 0), \quad (23)$$

$$\Omega_z^2 = \frac{\partial^2 \Phi^{\text{eff}}}{\partial z^2}(r, 0) = \frac{\partial^2 \Phi}{\partial z^2}(r, 0) \quad (24)$$

define the epicyclic oscillation frequency $\kappa(r)$ and the vertical oscillation frequency $\Omega_z(r)$. In fact κ^2 can be related to the angular velocity $\Omega(r)$ of circular orbits by

$$\kappa^2 = \frac{1}{r^3} \frac{d}{dr}(r^4 \Omega^2). \quad (25)$$

In the case of a point-mass potential, $\kappa = \Omega_z = \Omega$ and the perturbed orbit forms a closed figure, which is a slightly eccentric and/or slightly inclined Keplerian orbit.

The question of the hydrodynamic stability of a gaseous Keplerian disc is of fundamental importance, but is much more difficult to resolve definitively and must still be regarded as an open question. As noted in Section 2, the basic motion of fluid elements in a thin disc is essentially ballistic and only weakly influenced by pressure, viscosity or other collective effects. Although the stability properties of the test-particle orbits are not the sole determinant of the hydrodynamic stability of a gaseous disc, they naturally have an important bearing on the problem. Indeed, κ^2 given by (25) is precisely the stability discriminant in Lord Rayleigh's analysis of the stability of a cylindrical shear flow $\mathbf{u} = r\Omega(r)\mathbf{e}_\phi$ of an inviscid, incompressible fluid with respect to axisymmetric perturbations (Rayleigh, 1917). Although Rayleigh's result cannot be extended to the case of non-axisymmetric perturbations, some authors consider that Keplerian discs are stable because $\kappa^2 > 0$.

An alternative point of view is that, because the motion in a disc is a shear flow of extremely large Reynolds number (typically exceeding 10^{10}), it ought to be vigorously turbulent. If this were correct, the resulting turbulent transport of angular momentum might well account for the effective viscosity of accretion discs.

The association of a large Reynolds number (Re) with turbulence is based on the experience of laboratory shear flows such as those driven by a pressure gradient (Hagen–Poiseuille flow in a circular pipe or plane Poiseuille flow between parallel plates) or those driven by a shearing motion of the boundaries (Taylor–Couette flow between differentially rotating cylinders or plane Couette flow between parallel plates). In most cases the flow is observed to be turbulent at sufficiently large Re . Taylor–Couette flow is an exception because, if the ratios of the radii and of the angular velocities of the two cylinders are freely adjustable, there are regions of parameter space in which turbulence has not been observed at any value of Re . Broadly speaking, these regions correspond to situations in which $\kappa^2 > 0$ throughout the flow. However, these regions are relatively unexplored in experiments.

It was proven by Serrin (1959) that all flows of an incompressible fluid are monotonically and nonlinearly stable for sufficiently small Re . Although this obviously implies that turbulent motion *requires* a large Reynolds number, there is no corresponding theorem to the effect that all flows are unstable for sufficiently large Re . Perhaps the closest equivalents are the results of Lifschitz and Hameiri (1991) and Friedlander and Vishik (1991) on the stability of steady flows of an incompressible, inviscid fluid, which go some way to proving that “almost all” steady flows are linearly unstable at $\text{Re} = \infty$. In particular, any flow containing a non-degenerate stagnation point ($\mathbf{u} = \mathbf{0}$, $\nabla \mathbf{u} \neq \mathbf{0}$) is unstable. Intriguingly, these results do not apply to differentially rotating discs in which the angular velocity is strictly of one sign.

In fact, the traditional approach to hydrodynamic stability analysis, as expounded, for example, by Drazin and Reid (1981), fails spectacularly to account for the transition to turbulence in parallel shear flows such as plane Couette flow. The standard technique is to identify a steady flow, obtain the linearized perturbation equations, and seek solutions in the form of exponentially growing normal modes. Squire’s theorem implies that only two-dimensional disturbances (invariant in a direction orthogonal to the plane of the flow) need be considered. Rayleigh’s theorem implies that no growing modes exist at $\text{Re} = \infty$ if there is no inflection point in the velocity profile. For plane Couette flow, the linearized equations have no solutions in the form of exponentially

growing modes for any value of Re , yet the flow is observed to become turbulent for values of Re greater than a few hundred or so.

The reasons why this approach is misleading are well documented (e.g., Schmid and Henningson, 2001). The linear operator governing perturbations of a parallel shear flow at large Re is non-normal and highly ill-conditioned, with the result that the normal modes are a treacherous basis for a discussion of stability. In fact, at $Re = \infty$, all parallel shear flows are linearly unstable to algebraically growing disturbances, which are not two-dimensional in the sense of Squire's theorem (Ellingsen and Palm, 1975). For large but finite Re , infinitesimal disturbances of a similar form experience a large but ultimately transient amplification (Butler and Farrell, 1992). The transition to turbulence in plane Couette flow very probably occurs through a nonlinear instability in which nonlinear feedback couples with the algebraic linear amplification mechanism to provide sustained growth.

A clearer perspective is obtained by considering the shearing sheet as a family of models characterized by the inverse Rossby number Ω/A . When $\Omega/A = 0$ we have a non-rotating shear flow equivalent to an unbounded plane Couette flow (PCF); when $\Omega/A = 4/3$ we have a Keplerian shear flow (KSF), a local model of the flow in a standard accretion disc. The Rayleigh stability discriminant in the shearing sheet, $\kappa^2 = 4\Omega(\Omega - A)$, is plotted as a function of Ω/A in Fig. 2. It is clear from this figure that, while KSF amply satisfies Rayleigh's stability criterion, PCF lies at a point of marginal stability according to the same criterion. It is for precisely this reason that the transition to turbulence in PCF occurs through a nonlinear instability. The basin of attraction of the laminar flow diminishes to zero as Re increases (Thomson, 1887), with the result that PCF is stable to infinitesimal disturbances except at $Re = \infty$. The algebraic growth that occurs at $Re = \infty$ is a

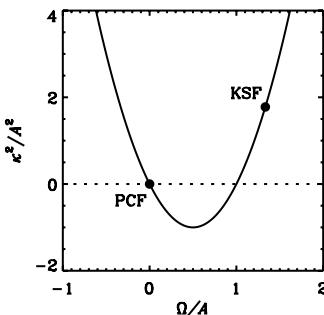


Figure 2. Normalized Rayleigh stability discriminant versus inverse Rossby number, for the shearing sheet. Non-rotating plane Couette flow (PCF) lies at marginal linear stability (dotted line), while Keplerian shear flow (KSF) amply satisfies Rayleigh's stability criterion.

consequence of the degenerate zero frequency eigenvalue corresponding to the condition $\kappa^2 = 0$.

Nagata (1990) has considered Taylor–Couette flow in the limit of a narrow gap, obtaining a model equivalent to the shearing sheet but bounded in the x -direction by rigid walls. When the Rossby number is fixed such that $0 < \Omega/A < 1$, and the Reynolds number is increased from zero to a certain critical value, the flow becomes linearly unstable and the solution branch known as Taylor vortex flow bifurcates from the basic state. Nagata followed this solution branch to $\Omega/A = 0$ where it is disconnected from the basic state but, in a sense, bifurcates from it at $Re = \infty$. As this solution branch can itself be unstable it provides a route to a subcritical transition to turbulence in PCF. However, there is no evidence that the solution branch continues significantly into the Rayleigh-stable interval $\Omega/A < 0$.

These lines of reasoning would suggest that Keplerian discs are not subject to the nonlinear hydrodynamic instability that occurs in laboratory shear flows. Balbus *et al.* (1996) and Balbus and Hawley (1998) have argued vigorously in favor of this viewpoint. While they easily reproduced the nonlinear hydrodynamic instability of PCF in numerical simulations in a shearing box, any putative equivalent instability of KSF failed to appear under equivalent conditions.

Nevertheless, other authors maintain a contrary viewpoint. Richard and Zahn (1999) draw attention to early experiments on Taylor–Couette flow in which turbulence and significant angular momentum transport were found to occur at sufficiently large Re , even in situations for which apparently $\kappa^2 > 0$ throughout the flow. The interpretation of these results is complicated by the fact that experimental Taylor–Couette flow, even in a laminar regime, differs from the theoretical model owing to end effects. Nevertheless, it appears that sustained turbulence is not absolutely prohibited when $\kappa^2 > 0$. One possibility is that the nonlinear instability of PCF persists to values of Ω/A slightly greater than 1, or less than 0, as suggested by the numerical results of Hawley *et al.* (1999). Alternatively, global effects may be important.

Whatever the eventual resolution of this question, the result will be an interesting one. If KSF is unstable, it raises the possibility that all accretion discs are turbulent even in cases when magnetic fields (see Section 3.4) are ineffective. If there is a well defined interval of Ω/A , wider than $[0, 1]$, in which instability is possible, this will also be of fundamental interest. If KSF is stable with respect to arbitrary perturbations, this will be a remarkable and unique result in fluid dynamics. The only way to demonstrate such a result would be to prove a theorem, because experimental or numerical approaches are limited in the

Reynolds numbers that can be achieved, and are subject to additional uncertainties. Either the shearing sheet or the shearing box is the correct setting for such a demonstration, because global models of discs are typically subject to weak non-axisymmetric instabilities that depend on the boundary conditions (e.g., Goldreich and Narayan, 1985). It would be a major achievement of mathematical fluid dynamics if the question of hydrodynamic stability could be definitively resolved in this way.

3.4. The magnetorotational instability

In discs that are sufficiently ionized that the magnetic field is dynamically coupled to the gas, the stability properties of the basic flow are dramatically altered. A magnetized fluid is more inventive, probably because it is not constrained to conserve vorticity, and is able to tap the shear energy even in differentially rotating flows that are stable according to Rayleigh's criterion. Balbus and Hawley (1991) first drew attention to this possibility for astrophysical discs, leading to a flurry of analytical and numerical calculations. The findings of many of these investigations, as well as the historical background, are described in the review article by Balbus and Hawley (1998).

Briefly stated, a Keplerian disc is linearly and locally unstable in the presence of a weak magnetic field of arbitrary configuration. The magnetorotational instability (MRI) is dynamical and extremely efficient, with a maximal growth rate equal to the Oort constant A . This means that the energy of the disturbance grows by a factor of up to $\exp(3\pi) \approx 12392$ per orbit. In the nonlinear regime, the instability saturates in a state of sustained, three-dimensional magnetohydrodynamic turbulence with a shear stress $W_{r\phi} < 0$ contributed by the Reynolds and Maxwell stresses associated with the fluctuating velocity and magnetic field.

The MRI is therefore very promising as an explanation of the effective viscosity of accretion discs. In most astrophysical environments the provision of a seed magnetic field to initiate the instability presents no difficulty. The MRI is suppressed only if the gas is insufficiently ionized, or if the magnetic field exceeds a certain strength, typically corresponding approximately to equipartition with the internal energy of the gas. The former restriction means that the MRI may fail in large regions of protoplanetary discs, and possibly in the quiescent states of accreting binary stars. In these systems, therefore, the issue of hydrodynamic stability and the possibility of alternative means of angular momentum transport remain of considerable interest.

Numerical studies of magnetorotational turbulence have been carried out in the shearing box and also, more recently, in global configurations. The simplest initial configuration involves a fluid of uniform density in a shearing box, which is possible either if the fluid is incompressible or if the vertical gravitational acceleration is neglected (Hawley *et al.*, 1995). As explained in Section 3.2, the turbulence that develops in such a system can be expected to be statistically steady and homogeneous, although anisotropic. Certain characteristics of the turbulent state are reproduced in all calculations, but the physics of the nonlinear regime is relatively poorly understood. A number of open questions remain.

Perhaps surprisingly, it is not yet possible to quote a definitive value of α resulting from the MRI. The magnitude of the shear stress in the turbulent state depends on the process by which the instability saturates, and this is not yet understood physically. A well established result is that the stress obtained in the shearing box is larger if a weak uniform magnetic field, especially a vertical one, is imposed on the box. The boundary conditions of the shearing box exactly conserve the total vertical magnetic flux, which therefore acts as a continuous source of linear instability. What is not well understood is how the stress depends on the horizontal size of the box, on the Reynolds number or on the magnetic Reynolds number (R_m). An attractive hypothesis is that the mean stress per unit volume should be asymptotically independent of all these quantities in the limit that they are all very large, as this corresponds to the physical situation in a typical disc. This has yet to be verified because numerical simulations in these limiting regimes are very costly. In fact, most calculations to date rely on numerical dissipation rather than explicit viscosity and resistivity, and therefore suffer the additional problem that the stress depends in an unknown way on the numerical algorithm and resolution. I would argue that there is a need for more resolved calculations with explicit dissipative mechanisms, in order to obtain robust quantitative results, even if they are at unrealistically small Reynolds numbers.

A case of special interest and difficulty is that in which there is no imposed magnetic flux to act as a continuous source of linear instability. The magnetorotational turbulence must then truly regenerate the magnetic field that gives rise to the instability, in a unique type of nonlinear dynamo process. It is known from numerical studies that a much larger magnetic Reynolds number is required to sustain turbulence in this case (Fleming *et al.*, 2000), and it is likely that no calculation to date has resolved the full stress available in this situation at very large R_m .

The important question of whether the magnetorotational turbulence can generate a magnetic field with a scale much larger than the thickness of the disc is largely unexplored to date. The numerical challenge is to devise a system that is sufficiently global that this question can be meaningfully addressed, but in which it is also computationally feasible to follow the evolution for the long time-scale that is probably required for such a dynamo process.

Many interesting numerical studies have been carried out recently to examine global effects or more complicated local physics. Of particular importance are the outcome of the MRI in accretion flows crossing the marginally stable circular orbit close to a black hole (e.g., Hawley and Krolik, 2001), the MRI in stratified discs covering many scale-heights (Miller and Stone, 2000), the MRI in weakly ionized discs, where Ohmic resistivity, ambipolar diffusion and the Hall effect can all be important (Sano and Stone, 2002), and the MRI in the inner parts of luminous discs around black holes and neutron stars, where radiation pressure exceeds the gas pressure (Turner *et al.*, 2002).

While numerical simulations are an essential and enlightening tool for investigating the nonlinear behavior of the MRI, there is something of a danger that accretion disc theory will be stalled by the prospect that every self-consistent calculation will require a three-dimensional numerical simulation of magnetohydrodynamic turbulence. Realistically, such simulations can only be conducted by expert practitioners and their results need to be distilled into a form that can be applied more widely. Furthermore, global numerical studies of magnetorotational turbulence in realistically thin discs are well beyond the currently available resources. There is a clear need for models that can bridge the gap between analytical studies and local numerical simulations, a subject to which I return in Section 4.4 below.

4. Continuum celestial mechanics

4.1. As a generalization of the classical theory of accretion discs

In Section 2 it was noted that, to a first approximation, fluid elements in a thin disc attempt to follow the same trajectories as test particles in the gravitational potential. In the case of a point-mass potential in Newtonian theory, the general solution allows for orbits that are eccentric and mutually inclined. The class of “Keplerian discs” is therefore much larger than is usually considered. Indeed, a substantial generalization

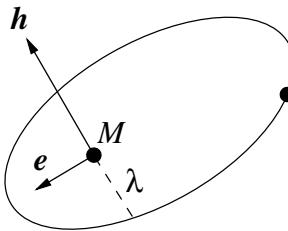


Figure 3. Elements of a Keplerian orbit.

of the classical theory of accretion discs is obtained by allowing for these possibilities.

A Keplerian orbit has six parameters, or “elements”. One of these, the time of periastron passage, is not relevant if the orbit is occupied by a continuous ring of matter. The remaining five orbital elements can be conveniently encoded in two vector quantities: the specific angular momentum vector \mathbf{h} , and the eccentricity vector \mathbf{e} (Fig. 3). The magnitude of \mathbf{h} determines the size of the orbit, specifically the semi-latus rectum λ of the ellipse, according to the relation $h^2 = GM\lambda$. The unit vector $\mathbf{l} = \mathbf{h}/h$ specifies the plane of the orbit, being orthogonal to it. The vector \mathbf{e} is defined to point toward the periastron, while its magnitude $e = |\mathbf{e}|$ is the eccentricity of the orbit.¹

The general solution for the shape of a Keplerian disc is obtained by allowing $\mathbf{h}(\lambda)$ and $\mathbf{e}(\lambda)$ to be arbitrary, continuous functions of λ , subject to a number of constraints. From the definitions of \mathbf{h} and \mathbf{e} it follows that $h = (GM\lambda)^{1/2}$ and $\mathbf{h} \cdot \mathbf{e} = 0$. There are therefore only four degrees of freedom for each orbit: two select the direction of the unit vector \mathbf{l} , a third selects the direction of \mathbf{e} within the plane orthogonal to \mathbf{l} , and a fourth selects the eccentricity e . Further constraints apply if the orbits are to be closed and non-intersecting: if \mathbf{e}' represents $d\mathbf{e}/d\ln\lambda$, the inequalities

$$[\mathbf{e} \cdot (\mathbf{e} - \mathbf{e}')]^2 + [\mathbf{l} \cdot (\mathbf{e} \times \mathbf{e}')]^2 < |\mathbf{e}|^2 < 1 \quad (26)$$

ensure that the orbits are closed and do not intersect their immediate neighbors.

For example, if $\mathbf{e} = \mathbf{0}$ but \mathbf{l} varies continuously as a function of λ , we have a warped disc composed of circular orbits (Fig. 4a). There is no restriction, in principle, on how rapidly the unit vector \mathbf{l} may vary with λ , which is equivalent to the spherical radius in this case. If, however,

¹ \mathbf{e} is sometimes called the Runge–Lenz vector, but see Lynden-Bell (2000) for a historical note.

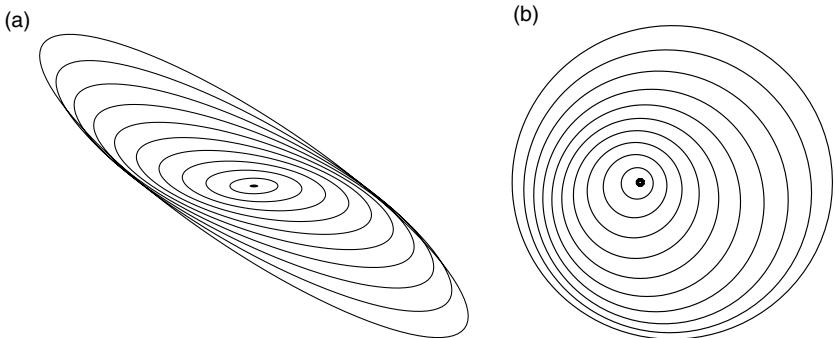


Figure 4. Warped and eccentric Keplerian discs.

\mathbf{l} is constant while \mathbf{e} varies continuously as a function of λ , we have a planar, eccentric disc (Fig. 4b). In this case the restrictions on the variation of \mathbf{e} amount to $|\mathbf{e} - \mathbf{e}'| < 1$ and $|\mathbf{e}| < 1$.

The general solution involves a combination of warping and eccentricity, which can be difficult to visualize. It is a remarkable fact that the class of discs composed of closed and non-intersecting orbits is so diverse and, although the very names “warped” and “eccentric” emphasize the “distorted” nature of these solutions, we should remember that they are equally valid members of the class of Keplerian discs.

4.2. As a generalization of classical celestial mechanics

The problem of continuum celestial mechanics is to determine how such discs evolve in time. Recall that the orbital solution follows from balancing the inertia of the flow and the gravitational attraction of the central mass. The collective effects of the fluid, by providing small additional contributions to the equation of motion, lead to a slow evolution of the disc. The time-scale of this evolution is generally of the order of $(H/r)^{-2}$ times the orbital time-scale, although in some circumstances it may be significantly longer or shorter. In the case of a standard disc, this evolution corresponds to the redistribution of the surface density $\Sigma(r, t)$ by the stress component $W_{r\phi}$, according to the diffusion (8). For warped and/or eccentric discs, however, the shape of the disc will also evolve slowly. The general problem therefore amounts to determining a set of partial differential equations governing the evolution of $\Sigma(\lambda, t)$, $\mathbf{l}(\lambda, t)$ and $\mathbf{e}(\lambda, t)$.

This problem can be understood as a generalization of one of the central problems of celestial mechanics, which is to determine the slow

evolution of the orbital elements of a body under the action of a perturbing force. Lagrange's solution of this problem using the method of variation of arbitrary constants (Lagrange, 1788), and later formulations by Gauss, Delaunay and others, effectively provide ordinary differential equations that relate the evolution of $\mathbf{h}(t)$ and $\mathbf{e}(t)$ to the perturbing force acting on the body. The evolution of $\Sigma(\lambda, t)$, $\mathbf{l}(\lambda, t)$ and $\mathbf{e}(\lambda, t)$ in a continuous disc is closely analogous. Even in the absence of an external force, an *internal* perturbing force is always provided by the collective effects of the disc, as embodied in the stress tensor \mathbf{W} .

Therefore continuum celestial mechanics generalizes the classical theory of accretion discs by allowing for eccentric and/or mutually inclined orbits, and at the same time generalizes classical celestial mechanics to the case of a continuous disc rather than a single orbiting body.

4.3. In relation to continuum mechanics

Continuum mechanics is concerned with the bulk motion and other macroscopic properties of continuous media. The conservation laws of mass, momentum and any other relevant quantities dictate the form of the equations governing the medium, in particular the equation of mass conservation,

$$(\partial_t + \mathcal{L}_q)\rho + \nabla_a(\rho u^a) = 0, \quad (27)$$

and the equation of motion,

$$(\partial_t + \mathcal{L}_q)(\rho u^a) + \nabla_b(\rho u^a u^b) = -\rho \nabla^a \Phi + \nabla_b W^{ab}. \quad (28)$$

For later convenience, these equations are written in a general, time-dependent coordinate system $\{q_a\}$. The operators \mathcal{L} and ∇ are the Lie derivative and the covariant derivative, respectively.

One of the central problems of continuum mechanics is to find the constitutive equations that characterize a particular substance by providing additional closure relations, which are needed because the conservation laws by themselves do not form a complete set of equations. In particular, the stress tensor W^{ab} that appears in the equation of motion must be related to the motion itself. Familiar examples of constitutive equations include those for elastic solids and viscous fluids, in which the stress is proportional to the strain or to the rate of strain, respectively.

It is possible to develop a dynamical theory of distorted discs without initially assuming a detailed constitutive equation. I defer a discussion

of constitutive equations until Section 4.4, and sketch below the method for eccentric discs (Ogilvie, 2001), although the same principles apply more generally.

The first step is to define a convenient coordinate system, one related to the geometry of the orbits. Orbital coordinates (λ, ϕ, z) use the semi-latus rectum λ as a label for the orbits and as a quasi-radial coordinate, while ϕ and z measure the azimuthal angle around the orbit and the distance above the mid-plane. Although these coordinates are time-dependent and non-orthogonal, they have the great advantage that the orbital motion can be expressed very simply.

The second step makes use of the existence of a small parameter $\epsilon \ll 1$ in the problem, such that the angular semi-thickness of the disc is $H/r = O(\epsilon)$. Although informal thin-disc approximations are widely used in studies of astrophysical discs, the systematic application of asymptotic expansions is practically essential in the treatment of distorted discs. The physical simplification associated with the fact that $\epsilon \ll 1$ is a separation of scales: the vertical structure of the disc decouples from the large-scale radial inhomogeneities, while the characteristic time-scale of the evolution of the shape of the disc is much longer than the orbital time-scale. This separation is achieved formally by the introduction of scaled variables $\zeta = z/\epsilon$ and $T = \epsilon^2 t$, if a system of units is adopted in which λ and Ω are of order unity.

The eccentricity vector is more conveniently expressed as a complex eccentricity variable $E(\lambda, T)$ formed from the x - and y -components of \mathbf{e} in the combination $E = e_x + ie_y$. The magnitude of E is then the eccentricity of the orbit, while the phase of E describes the precession of the orbit. Physical quantities, for example the (contravariant) velocity components, are given appropriate asymptotic expansions of the form

$$\begin{aligned} u^\lambda &= \epsilon^2 u_2^\lambda(\lambda, \phi, \zeta, T) + O(\epsilon^4), \\ u^\phi &= \Omega(\lambda, \phi, T) + \epsilon^2 u_2^\phi(\lambda, \phi, \zeta, T) + O(\epsilon^4), \\ u^z &= \epsilon u_1^z(\lambda, \phi, \zeta, T) + \epsilon^3 u_3^z(\lambda, \phi, \zeta, T) + O(\epsilon^5). \end{aligned} \quad (29)$$

The dominant, orbital motion $u_0^\phi = \Omega$ is very simply expressed in the orbital coordinate system. Among the small corrections to this motion, the term $\epsilon^2 u_2^\lambda$ includes the slow accretion flow that carries mass across the orbits. The existence of a significant vertical velocity ϵu_1^z in a planar, eccentric disc may be surprising, but it plays an important role and cannot be neglected.

When these expansions, together with those for ρ , Φ and W^{ab} , are substituted into the basic equations (27) and (28), a series of

mathematical problems is obtained at successive orders in ϵ . After some effort it is possible to extract the desired evolutionary equations for the disc. Two of these express the conservation of mass and angular momentum in a one-dimensional form identical to (3) and (4) except for certain geometrical factors relating to the eccentric shape of the disc. These can be combined to provide a diffusion equation for the surface density that differs little from (8). Therefore only small changes occur to the physics of accretion in an eccentric disc. However, a new equation is also obtained, which governs the evolution of the shape of the disc, and has the form

$$h\Sigma \left(\frac{\partial E}{\partial t} + v \frac{\partial E}{\partial \lambda} - \frac{vE}{\lambda} \right) = \frac{\partial}{\partial \lambda} \left\langle \left(2r^2 \mathcal{W}^{\lambda\phi} - i\lambda \frac{\partial r}{\partial \lambda} \mathcal{W}^{\lambda\lambda} \right) e^{i\phi} \right\rangle + \left\langle \frac{r^2}{\lambda} \mathcal{W}^{\lambda\phi} \left(e^{i\phi} - E + \lambda \frac{\partial E}{\partial \lambda} \right) - ir^2 \mathcal{W}^{\phi\phi} e^{i\phi} \right\rangle, \quad (30)$$

where $v(\lambda, t)$ is a mean accretion velocity related to the net mass flux across the orbits.

This equation relies on minimal assumptions and is therefore very general. It involves orbital averages of certain combinations of components of the vertically integrated stress tensor $\mathcal{W}^{ab} = \int W^{ab} dz$. At this stage we therefore face the closure problem of continuum mechanics: the stress must somehow be related to the motion of the fluid. This problem also appeared in the classical theory, but in a much simpler form. Only one component of the stress tensor plays a role in the evolution of a standard disc, and it is reasonable to parameterize this using a single dimensionless parameter α . In the case of distorted discs we must deal with several components of the stress tensor, and it is therefore important to find a constitutive equation that respects the tensorial nature of the stress.

The case of a warped disc, in which the orbits are circular but mutually inclined, is perhaps easier to understand. The evolutionary equations derived from the asymptotic analysis (Ogilvie, 1999) can be viewed as straightforward vectorial generalizations of (3) and (4),

$$2\pi r \frac{\partial \Sigma}{\partial t} - \frac{\partial F}{\partial r} = 2\pi r S, \quad (31)$$

$$2\pi rh \frac{\partial}{\partial t} (\Sigma \mathbf{l}) - \frac{\partial}{\partial r} (F h \mathbf{l}) = \frac{\partial \mathbf{G}}{\partial r} + 2\pi r \mathbf{T}. \quad (32)$$

Elimination of F again yields an evolutionary equation for $\Sigma(r, t)$ closely related to (7). However, a new equation is also obtained for $\partial \mathbf{l}/\partial t$, which

governs the evolution of the shape of the disc. Equations of this general form were first written down by Papaloizou and Pringle (1983), correcting earlier approaches initiated by Bardeen and Petterson (1975), and were investigated further by Pringle (1992).

Eccentric discs have received much less attention than warped discs, although there is at least as much observational and circumstantial evidence for them. Part of the difficulty lies in the fact that the eccentricity is not a conserved quantity. More precisely, while the eccentricity of an orbiting particle can be deduced from a knowledge of its mass, energy and angular momentum, the longitude of periastron, and therefore the direction of the eccentricity vector, cannot. Equation (30) is a non-trivial complex, or vectorial, equation that must be deduced from a detailed analysis of the equation of motion, and its form cannot be written down purely on the basis of conservation principles.

4.4. Constitutive equations for turbulent accretion discs

We have seen that it is possible to develop a dynamical theory of distorted discs that is fully nonlinear, in the sense that the orbital eccentricities and inclinations need not be small, and is asymptotically exact in the limit of a thin disc. The missing ingredient in the theory is a constitutive equation for the stress in the disc. The velocity field of the fluid is determined (mainly) by the instantaneous shape of the disc, i.e., by $\mathbf{l}(\lambda, t)$ and $\mathbf{e}(\lambda, t)$. If the stress in the disc can be related to its motion, a closed system of equations for $\Sigma(\lambda, t)$, $\mathbf{l}(\lambda, t)$ and $\mathbf{e}(\lambda, t)$ can be obtained.

The general problem of formulating appropriately invariant constitutive equations for continuous media is described in an important paper by Oldroyd (1950). In a local constitutive equation the stress at any point in the medium depends only on the deformation history of that element. Is it reasonable to expect such a relation to exist, in principle, for an accretion disc? Or must the framework of continuum celestial mechanics be discarded because turbulent discs can be studied only using direct numerical simulations?

If the stress in a disc is associated with large-scale phenomena such as magnetized outflows or spiral density waves, there is little reason to suppose a local constitutive equation to exist. However, if the stress is associated with turbulence that results from a local instability such as the MRI, there is good reason for such a supposition. Shearing-box studies indicate, and in fact are founded on the premise, that the physics of the MRI, including its nonlinear saturation, is determined within

a volume comparable to H^3 . It is also clear from these studies that the turbulence has a finite “memory”, lasting perhaps for a few orbits. This suggests the mean turbulent stress at any point in the disc should depend only on the deformation history of the fluid within that limited range of space and time. In the context of continuum celestial mechanics, where the distortion of the disc has a length-scale large compared to H and a time-scale long compared to Ω^{-1} , this means that a local constitutive equation with a limited memory is appropriate.

The simplest reasonable model of the turbulent stress is a viscous model, equivalent to the Navier–Stokes equation, in which the stress tensor is instantaneously and linearly related to the rate-of-strain tensor. Physically, this corresponds to the traditional notion that turbulence provides an isotropic “eddy viscosity”. It has been widely used in studies of accretion discs, especially in connection with the alpha parameterization of Shakura and Sunyaev (1973). When applied to the problem of continuum celestial mechanics, the viscous model allows an exact nonlinear closure of the equations (Ogilvie, 1999, 2000, 2001). For warped discs the model makes quite reasonable predictions: while pressure causes a warping disturbance to propagate through the disc in a wavelike manner, viscosity leads to a diffusion and dissipation of the warp. The fully nonlinear evolutionary equations have been applied successfully to the radiative warping of accretion discs in X-ray binaries (Ogilvie and Dubus, 2001).

For eccentric discs, however, the viscous model encounters a serious problem. While pressure allows eccentricity to propagate through the disc as a dispersive wave, viscosity generally leads to a *negative* diffusivity for eccentricity. This would imply that circular discs are unstable to small-scale eccentric perturbations, and appears to contradict an argument referred to in Section 2, that dissipative processes should result in circular and coplanar orbits. In fact there is no contradiction, because energy can be made available through accretion, but I have shown that the negative diffusivity is an artifact of the assumption of an instantaneous relation between the stress and the rate of strain (Ogilvie, 2001). The instability is generally eliminated when a viscoelastic model of the turbulent stress is used, which takes account of the non-zero relaxation time of the turbulence.

Ogilvie and Proctor (2003) explore in some detail the close relation between viscoelasticity and magnetohydrodynamics. Viscoelastic models such as Oldroyd’s liquid B are relevant to dilute solutions of high-molecular weight polymer molecules. Like the magnetic field lines in an astrophysical plasma, the polymer molecules are advected and distorted by the fluid flow, and react on it through their tension. The physical and

mathematical analogy is so close that a direct analogue of the MRI can be identified in viscoelastic Couette flow.

The wider prospect of finding constitutive equations for magnetorotational turbulence is related to the search for Reynolds-stress closure models for hydrodynamic turbulence or turbulent convection, for example in engineering applications (e.g., Speziale, 1991). The application of these methods to accretion discs was initiated by Kato and Yoshizawa (1993), who also modeled the Maxwell stress. More recently I introduced a simple but properly invariant model that aims to address the nonlinear saturation of the MRI and is able to reproduce many of the known properties of magnetorotational turbulence (Ogilvie, 2003). Although this is only a beginning, it is to be hoped that this type of approach can help to bridge the gap between analytical studies and numerical simulations, leading to a better understanding of astrophysical discs.

Acknowledgments

I am grateful to the organizers of the LMS Durham Symposium, and acknowledge the support of the Royal Society through a University Research Fellowship.

References

- Balbus, S. A. and Hawley, J. F., "A powerful local shear instability in weakly magnetized disks. I. Linear analysis," *Astrophys. J.* **376**, 214–233 (1991).
- Balbus, S. A. and Hawley, J. F., "Instability, turbulence, and enhanced transport in accretion disks," *Rev. Mod. Phys.* **70**, 1–53 (1998).
- Balbus, S. A., Hawley, J. F. and Stone, J. M., "Nonlinear stability, hydrodynamical turbulence, and transport in disks," *Astrophys. J.* **467**, 76–86 (1996).
- Bardeen, J. M. and Petterson, J. A., "The Lense–Thirring effect and accretion disks around Kerr black holes," *Astrophys. J.* **195**, L65–L67 (1975).
- Benz, W., Cameron, A. G. W., Press, W. H. and Bowers, R. L., "Dynamic mass exchange in doubly degenerate binaries. I. 0.9 and $1.2 M_{\odot}$ stars," *Astrophys. J.* **348**, 647–667 (1990).
- Blandford, R. D. and Payne, D. G., "Hydromagnetic flows from accretion discs and the production of radio jets," *Mon. Not. R. Astron. Soc.* **199**, 883–903 (1982).
- Butler, K. M. and Farrell, B. F., "Three-dimensional optimal perturbations in viscous shear flow," *Phys. Fluids A* **4**, 1637–1650 (1992).
- Drazin, P. G. and Reid, W. H., *Hydrodynamic Stability*, Cambridge University Press, Cambridge (1981).

- Ellingsen, T. and Palm, E., "Stability of linear flow," *Phys. Fluids* **18**, 487–488 (1975).
- Fleming, T. P., Stone, J. M. and Hawley, J. F., "The effect of resistivity on the nonlinear stage of the magnetorotational instability in accretion disks," *Astrophys. J.* **530**, 464–477 (2000).
- Friedlander, S. and Vishik, M. M., "Instability criteria for the flow of an inviscid incompressible fluid," *Phys. Rev. Lett.* **66**, 2204–2206 (1991).
- Goldreich, P. and Lynden-Bell, D., "II. Spiral arms as sheared gravitational instabilities," *Mon. Not. R. Astron. Soc.* **130**, 125–158 (1965).
- Goldreich, P. and Narayan, R., "Non-axisymmetric instability in thin discs," *Mon. Not. R. Astron. Soc.* **213**, 7P–10P (1985).
- Gor'kavyi, N. N. and Fridman, A. M., *Fizika Planetnykh Kolets*, Nauka, Moscow (1994). English translation: Fridman, A. M. and Gor'kavyi, N. N., *Physics of Planetary Rings* (trans. D. ter Haar), Springer, Berlin (1999).
- Hawley, J. F. and Krolik, J. H., "Global MHD simulation of the inner accretion disk in a pseudo-Newtonian potential," *Astrophys. J.* **548**, 348–367 (2001).
- Hawley, J. F., Balbus, S. A. and Winters, W. F., "Local hydrodynamic stability of accretion disks," *Astrophys. J.* **518**, 394–404 (1999).
- Hawley, J. F., Gammie, C. F. and Balbus, S. A., "Local three-dimensional magnetohydrodynamic simulations of accretion disks," *Astrophys. J.* **440**, 742–763 (1995).
- Kant, I., *Allgemeine Naturgeschichte und Theorie des Himmels*, Petersen, Königsberg and Leipzig (1755). English translation in: Hastie, W., *Kant's Cosmogony*, Maclehose, Glasgow (1900).
- Kato, S. and Yoshizawa, A., "A model of hydromagnetic turbulence in accretion disks," *Publ. Astron. Soc. Japan* **45**, 103–112 (1993).
- Lagrange, J.-L., *Mécanique Analytique*, Desaint, Paris (1788). English translation: Lagrange, J.-L., *Analytical Mechanics* (trans./ed. A. Boissonade and V. N. Vagliente), Kluwer, Dordrecht (1997).
- Lee, U., Saio, H. and Osaki, Y., "Viscous excretion discs around Be stars," *Mon. Not. R. Astron. Soc.* **250**, 432–437 (1991).
- Lewin, W. H. G., van Paradijs, J. and van den Heuvel, E. P. J., *X-Ray Binaries*, Cambridge University Press, Cambridge (1995).
- Lifschitz, A. and Hameiri, E., "Local stability conditions in fluid dynamics," *Phys. Fluids A* **3**, 2644–2651 (1991).
- Lubow, S. H., Papaloizou, J. C. B. and Pringle, J. E., "Magnetic field dragging in accretion discs," *Mon. Not. R. Astron. Soc.* **267**, 235–240 (1994a).
- Lubow, S. H., Papaloizou, J. C. B. and Pringle, J. E., "On the stability of magnetic wind-driven accretion discs," *Mon. Not. R. Astron. Soc.* **268**, 1010–1014 (1994b).
- Lüst, R., "Die Entwicklung einer um einen Zentralkörper rotierenden Gasmasse. I. Lösungen der hydrodynamischen Gleichungen mit turbulenter Reibung," *Z. Naturforsch.* **7a**, 87–98 (1952).
- Lynden-Bell, D., "Galactic nuclei as collapsed old quasars," *Nature* **223**, 690–694 (1969).

- Lynden-Bell, D., "The Newton wonder in mechanics," *Observatory* **120**, 131–136 (2000).
- Lynden-Bell, D. and Kalnajs, A. J., "On the generating mechanism of spiral structure," *Mon. Not. R. Astron. Soc.* **157**, 1–30 (1972).
- Lynden-Bell, D. and Pringle, J. E., "The evolution of viscous discs and the origin of the nebular variables," *Mon. Not. R. Astron. Soc.* **168**, 603–637 (1974).
- McCaughrean, M. J. and O'Dell, C. R., "Direct imaging of circumstellar disks in the Orion Nebula," *Astron. J.* **111**, 1977–1986 (1996).
- Mayor, M. and Queloz, D., "A Jupiter-mass companion to a solar-type star," *Nature* **378**, 355–359 (1995).
- Mestel, L., "Magnetic braking by a stellar wind — I," *Mon. Not. R. Astron. Soc.* **138**, 359–391 (1968).
- Miller, K. A. and Stone, J. M., "The formation and structure of a strongly magnetized corona above a weakly magnetized accretion disk," *Astrophys. J.* **534**, 398–419 (2000).
- Nagata, M., "Three-dimensional finite-amplitude solutions in plane Couette flow: bifurcation from infinity," *J. Fluid Mech.* **217**, 519–527 (1990).
- Novikov, I. D. and Thorne, K. S., "Astrophysics of black holes", in: *Black Holes* (Eds. C. DeWitt and B. S. DeWitt), Gordon and Breach, New York (1973).
- Ogilvie, G. I., "The non-linear fluid dynamics of a warped accretion disc," *Mon. Not. R. Astron. Soc.* **304**, 557–578 (1999).
- Ogilvie, G. I., "An alpha theory of time-dependent warped accretion discs," *Mon. Not. R. Astron. Soc.* **317**, 607–622 (2000).
- Ogilvie, G. I., "Non-linear fluid dynamics of eccentric discs," *Mon. Not. R. Astron. Soc.* **325**, 231–248 (2001).
- Ogilvie, G. I., "On the dynamics of magnetorotational turbulent stresses," *Mon. Not. R. Astron. Soc.* **340**, 969–982 (2003).
- Ogilvie, G. I. and Dubus, G., "Precessing warped accretion discs in X-ray binaries," *Mon. Not. R. Astron. Soc.* **320**, 485–503 (2001).
- Ogilvie, G. I. and Proctor, M. R. E., "On the relation between viscoelastic and magnetohydrodynamic flows and their instabilities," *J. Fluid Mech.* **476**, 389–409 (2003).
- Oldroyd, J. G., "On the formulation of rheological equations of state," *Proc. R. Soc. Lond. A* **200**, 523–541 (1950).
- Papaloizou, J. C. B. and Lin, D. N. C., "Theory of accretion disks I: angular momentum transport processes," *Annu. Rev. Astron. Astrophys.* **33**, 505–540 (1995).
- Papaloizou, J. C. B. and Pringle, J. E., "The time-dependence of non-planar accretion discs," *Mon. Not. R. Astron. Soc.* **202**, 1181–1194 (1983).
- Pringle, J. E., "Accretion and binary X-ray sources," Ph. D. dissertation, University of Cambridge (1974).
- Pringle, J. E., "Accretion discs in astrophysics," *Annu. Rev. Astron. Astrophys.* **19**, 137–162 (1981).
- Pringle, J. E., "A simple approach to the evolution of twisted accretion discs," *Mon. Not. R. Astron. Soc.* **258**, 811–818 (1992).

- Pumir, A., "Turbulence in homogeneous shear flows," *Phys. Fluids* **8**, 3112–3127 (1996).
- Lord Rayleigh, "On the dynamics of revolving fluids," *Proc. R. Soc. Lond. A* **93**, 148–154 (1917).
- Richard, D. and Zahn, J.-P., "Turbulence in differentially rotating flows: what can be learned from the Couette–Taylor experiment," *Astron. Astrophys.* **347**, 734–738 (1999).
- Sano, T. and Stone, J. M., "The effect of the Hall term on the nonlinear evolution of the magnetorotational instability. II. Saturation level and critical magnetic Reynolds number," *Astrophys. J.* **577**, 534–553 (2002).
- Schmid, P. J. and Henningson, D. S., *Stability and Transition in Shear Flows*, Springer, New York (2001).
- Serrin, J., "On the stability of viscous fluid motions," *Arch. Rat. Mech. Anal.* **3**, 1–13 (1959).
- Shakura, N. I. and Sunyaev, R. A., "Black holes in binary systems. Observational appearance," *Astron. Astrophys.* **24**, 337–355 (1973).
- Speziale, C. G., "Analytical methods for the development of Reynolds-stress closures in turbulence," *Annu. Rev. Fluid Mech.* **23**, 107–157 (1991).
- Thomson, W. (Lord Kelvin), "Stability of fluid motion — rectilineal motion of viscous fluid between two parallel planes," *Phil. Mag.* **24**(5), 188–196 (1887).
- Turner, N. J., Stone, J. M. and Sano, T., "Local axisymmetric simulations of magnetorotational instability in radiation-dominated accretion disks," *Astrophys. J.* **566**, 148–163 (2002).
- Warner, B., *Cataclysmic Variable Stars*, Cambridge University Press, Cambridge (1995).
- von Weizsäcker, C. F., "Die Rotation kosmischer Gasmassen," *Z. Naturforsch.* **3a**, 524–539 (1948).
- Wisdom, J. and Tremaine, S., "Local simulations of planetary rings," *Astron. J.* **95**, 925–940 (1988).

2 Discs with MHD turbulence and their response to orbiting planets

John C.B. Papaloizou and Richard P. Nelson

*Astronomy Unit, Queen Mary, University of London,
Mile End Road, London E1 4NS, UK*

We consider both local and global simulations of models of accretion discs with MHD turbulence driven by the magnetorotational instability (MRI). The simulations have conserved zero net vertical and toroidal magnetic flux. Time averaging for a period of at least 7 – 8 orbital periods at the outer boundary of the active domain of global models was required to reduce fluctuations sufficiently to reveal behavior even approximately corresponding to predictions from classical viscous disc theory which only has potential use for discs evolving on very long time scales.

We also describe simulations of turbulent disc models interacting with a perturbing protoplanet with sufficient mass to significantly affect background properties. Phenomena such as spiral wave excitation and gap clearing occur in both global and local models consistently with ideas from earlier theories.

1. Introduction

The recent and ongoing discovery of extrasolar giant planets has stimulated renewed interest in the theory of planet formation (e.g., Mayor and Queloz, 1995; Marcy *et al.*, 2000; Vogt *et al.*, 2002). In particular the discovery of giant planets close to their central stars has led to the idea that they migrated inward due to gravitational interaction with the gaseous protoplanetary disc out of which they formed.

In the usual picture of disc–protoplanet interactions, the presence of a protoplanet orbiting in a gaseous accretion disc leads to the excitation of spiral density waves at Lindblad resonances, which propagate from the locations where they are launched (e.g., Goldreich and Tremaine, 1979). These spiral waves carry with them an associated

angular momentum flux that is deposited in the disc where the waves are damped, either through shocks or viscous effects.

Early work on the interaction between a giant protoplanet of mass m_p in circular orbit at radius R about a central star of mass M_* and a laminar protoplanetary disc of semi-thickness H which transports angular momentum through the action of an anomalous viscosity, indicates that an annular gap is expected to form in the vicinity of the protoplanet orbit provided that two criteria are satisfied (e.g., Papaloizou and Lin, 1984; Lin and Papaloizou, 1993). The first criterion (the so-called viscous criterion) requires that

$$\frac{m_p}{M_*} > 40\alpha \left(\frac{H}{R}\right)^2, \quad (1)$$

which arises from the requirement that tidal torques induced by the protoplanet exceed viscous torques in the disc. Here we parameterize the anomalous disc viscosity using the well known “ α ” parameterization of Shakura and Sunyaev (1973). That also parameterizes the rate of flow of angular momentum through the disc that occurs without perturbation by a protoplanet.

The second criterion (the so-called thermal criterion) requires that

$$\frac{m_p}{M_*} > 3 \left(\frac{H}{R}\right)^3 \quad (2)$$

and comes from the requirement that the spiral waves introduced into the disc flow by the planet be nonlinear so that spiral shocks form (Ward, 1997). It is also the requirement that the radius of the Hill sphere, interior to which the gravity of the protoplanet dominates, exceeds the disc semi-thickness.

A variety of numerical studies of the interaction between a protoplanet and a laminar but viscous disc (Lin and Papaloizou, 1986, 1993; Bryden *et al.*, 1999; Nelson *et al.*, 2000; D’Angelo *et al.*, 2002) indicate that a protoplanet in the Jovian mass range will open a gap and that the torques exerted through the disc protoplanet interaction can produce inward orbital migration.

However, such simulations are normally carried out in two-dimensions and the effect of the turbulence producing the anomalous viscosity has yet to be taken into account. The origin of this turbulence was uncertain until Balbus and Hawley (1991) provided an explanation for its origin through the operation of the magnetorotational instability (MRI). Simulations of protoplanets interacting with discs

self-consistently maintaining MHD turbulence should therefore be undertaken in order to investigate the disc protoplanet interaction and its consequences for orbital migration.

Improved computational resources now make it feasible to consider three-dimensional simulations of turbulent discs interacting with protoplanets.

Here we describe some results of some of the first attempts at such simulations. We focus on the properties of both global and local turbulent disc models prior to the introduction of a perturbing protoplanet (see Steinacker and Papaloizou, 2002). We go on to consider the effects of introducing a protoplanet with sufficient mass to significantly affect the properties of the turbulence and the disc structure in the neighborhood of its orbit (see Papaloizou and Nelson, 2003; and Nelson and Papaloizou, 2003 for more details).

To ease computational requirements we adopt cylindrical disc models with no vertical stratification. For this first study we assume that the disc is adequately ionized throughout so that ideal MHD applies and consider models with zero net magnetic flux. We study the onset and maintenance of the turbulence in such models which seems to define a state attained independently of reasonable initial conditions. We relate the behavior of the disc to the classical “ α ” viscous disc theory of Shakura and Sunyaev (1973). This is because in addition to providing the most common and simple conceptual framework for disc modeling it still remains the main contact with observations (Balbus and Papaloizou, 1999). We here focus on time averages of disc quantities such as the $r\phi$ stress and radial inflow velocity and study the extent to which stable behavior of these quantities relates to classical disc theory. In general significant periods of time averaging are required to provide even an approximate correspondence and so the disc does not behave like a laminar one with anomalous viscosity on dynamical time scales.

We then go on to illustrate the response of turbulent disc models to forcing by a perturbing protoplanet focusing on spiral wave generation and gap formation.

2. Set up of initial models

2.1. Global models

The governing equations for MHD written in a frame rotating with uniform angular velocity $\Omega_p \hat{\mathbf{k}}$ with $\hat{\mathbf{k}}$ being the unit vector in the vertical

direction are

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \rho \mathbf{v} = 0, \quad (3)$$

$$\rho \left(\frac{\partial \mathbf{v}}{\partial t} + \mathbf{v} \cdot \nabla \mathbf{v} + 2\Omega_p \hat{\mathbf{k}} \times \mathbf{v} \right) = -\nabla P - \rho \nabla \Phi + \frac{1}{4\pi} (\nabla \times \mathbf{B}) \times \mathbf{B}, \quad (4)$$

$$\frac{\partial \mathbf{B}}{\partial t} = \nabla \times (\mathbf{v} \times \mathbf{B}), \quad (5)$$

where \mathbf{v} , P , ρ , \mathbf{B} and Φ denote the fluid velocity, pressure, density, magnetic field and potential, respectively. Φ contains contributions due to gravity and the centrifugal potential $-(1/2)\Omega_p^2 r^2$. Adopting cylindrical coordinates (z, r, ϕ) , with origin at the central star, we use a locally isothermal equation of state

$$P = \rho c(r)^2, \quad (6)$$

where $c(r)$ denotes the sound speed which is a fixed function of r .

Note that this formulation means that the models investigated have no vertical dependence of the gravitational potential and thus no vertical stratification and may thus be described as cylindrical discs (e.g., Armitage, 1998; Hawley, 2001). The gravitational potential is taken to be independent of z , so that

$$\Phi = -\frac{GM_*}{r} - \frac{Gm_p}{\sqrt{r^2 + D^2 - 2rD \cos(\phi - \phi_p) + b^2}} - \frac{1}{2}\Omega_p^2 r^2. \quad (7)$$

Here M_* is the central mass, m_p is the mass of the perturbing planet located at $(0, D, \phi_p)$, if included, G is the gravitational constant, b is the softening parameter for global models taken to be $R_H/2$, with $R_H = D[m_p/(3M_*)]^{1/3}$ being the radius of the Hill sphere and the final term represents the centrifugal potential where applicable. Thus the cylindrical disc models do not include a full treatment of the disc vertical structure. Models of this type are employed due to the high computational overhead that would be required to resolve fully the disc vertical structure of a stratified model.

Here we consider two global models labeled 1 and 2 which contain zero net vertical and toroidal magnetic flux. These are of course both conserved quantities for the simulations presented here (global simulations of planets in discs with a non-zero toroidal flux have been recently considered by Winters *et al.*, 2003a). Our calculations were commenced in the inertial frame ($\Omega_p = 0$).

The radial domain (R_1, R_2) of each model consists of an interior Keplerian domain (R_{K1}, R_{K2}) in which the gas initially is in a state of

Keplerian rotation with angular velocity $\Omega \propto r^{-3/2}$ which, when a magnetic field is then imposed, becomes unstable to the MRI. This unstable Keplerian domain is then buffered between two stable regions which prevent magnetically active material from coming into contact with the boundaries and so enables the accurate conservation of magnetic flux during the simulations.

The inner stable region can be regarded as modeling the boundary layer region between the central star and disc. For model 1 the initial Ω was taken as constant in this domain while for model 2, the initial $\Omega \propto r^2$. In the outer stable region Ω was constant for both global models. Of course in all cases, constants and scaling parameters were chosen so as to give a continuous initial rotation profile.

In the Keplerian domain the initial density was such that $\rho \propto 1/r$ for both global models. In their Keplerian and outer stable domains (see below) we adopted $c^2(r) \propto 1/r$ with constant of proportionality such that $c/(r\Omega) = 0.1$ giving a putative disc aspect ratio of 0.1. In the inner stable domain we took $c^2(r) \propto r^{-5/4}$. As in Steinacker and Papaloizou (2002) the angular velocity scaling in the Keplerian domain together with the initial density distribution were chosen such that hydrostatic equilibrium was satisfied in the radial direction.

Periodic boundary conditions were adopted in the vertical and azimuthal directions and the boundary conditions adopted in the radial direction were as in Steinacker and Papaloizou (2002). Model parameters are summarized in [Table 1](#).

Dimensionless units of length and time are adopted such that $R_1 = 1$ and $GM = 1$.

The initial magnetic field was given by

$$\mathbf{B} = B_0 \sin \left(2n_R \pi \frac{r - r_m}{r_{m1} - r_m} \right) \mathbf{e}_i \quad (8)$$

with the integer n_R being the number of 2π cycles. The index $i = 1, 2$ indicates either the vertical or the toroidal field component with the corresponding unit vector \mathbf{e}_i , respectively. For toroidal fields B_0 is constant while for vertical fields $B_0 \propto 1/r$. In order to show that the final turbulent state was independent of the initial magnetic field, provided that it was not too large in magnitude and varied on a sufficiently small scale, model 1 was initiated with a toroidal field and model 2 with a vertical field. The initial magnetic field was applied in an annulus within the Keplerian domain with inner and outer bounding radii r_m and r_{m1} , respectively. The normalization of B_0 was chosen such that the initial magnetic energy in the Keplerian domain expressed in units of the volume integrated pressure there was 0.03 for model 1 and 0.002

Table 1. Model parameters*

Model	z_1	z_2	R_1	R_2	ϕ_2	R_{K1}	R_{K2}	n_z	n_R	n_ϕ	r_m	r_{m1}	n_R
1	-0.2	0.2	1.0	6.1	$\pi/3$	1.25	5.0	40	380	100	2.35	4.35	2
2	-0.3	0.3	1.0	8.8	$\pi/3$	1.2	7.2	60	370	100	3.5	6.5	3

*The first column gives the model label, the second and third give the vertical domain of extent L_z , the fourth and fifth give the radial domain while the sixth column gives the maximum extent of the azimuthal domain ϕ_2 . The Keplerian domain is specified in columns seven and eight and the numbers of equally spaced grid points in the vertical, radial and azimuthal directions are given in columns nine, ten and eleven. The final three columns give the boundaries of the domain in which the initial magnetic field was applied and the number of 2π cycles in the functional form.

for model 2. The smaller value is used in the model 2 case because of the inherently more disruptive behavior of initial vertical fields as compared to toroidal fields of the same magnitude. In addition, for these models, a small amplitude sinusoidal radial velocity perturbation was imposed in (r_m, r_{m1}) .

As simulations progress, the MRI develops eventually leading to MHD turbulence which diffuses throughout the Keplerian domain. A quasi-steady state constituting a dynamo is set up that has been found to last for up to 50 orbital periods measured at the outer boundary of the Keplerian domain. This state has been found to be independent of a wide range of initial conditions provided the zero net magnetic flux condition is fulfilled.

Model 2 was initiated in the inertial frame with an azimuthal domain of extent $\pi/3$, and was subsequently used to study disc–planet interaction. Initially it was run up to time $t = 3825.5$ at which it exhibited a fully turbulent state. For the purpose of studying the interaction between a turbulent protostellar disc and an embedded protoplanet, using the periodic boundary condition, the azimuthal domain for this model was then extended to 2π by joining six $\pi/3$ sectors together. A transformation to a frame rotating with an angular velocity Ω_p corresponding to the fixed circular Keplerian motion of an inserted planet of 5 Jupiter masses was carried out and the evolution continued.

2.2. Local simulations

Although the calculation of the interaction of an orbiting planet strictly requires a global simulation using the full 2π in azimuth, it is of interest to see whether the main features of the disc–planet interaction

can be captured in a local simulation. The higher resolution one obtains in a local simulation would then permit greater in-depth studies. The local patch in a shearing disc corresponds to the well known shearing box (Goldreich and Lynden-Bell, 1965). This is a small Cartesian box centered on an arbitrary point at which the Keplerian angular velocity is Ω_K . Local Cartesian coordinates (x, y, z) with associated unit vectors $(\mathbf{i}, \mathbf{j}, \mathbf{k})$ are adopted. The direction \mathbf{i} points from the central object to the origin, \mathbf{j} is along the unperturbed rotational shear flow and \mathbf{k} points in the vertical direction. The shearing box setup has been used many times for conducting MHD simulations in a shearing medium corresponding to a Keplerian accretion disc (e.g., Hawley *et al.*, 1995, 1996; Brandenburg *et al.*, 1996; Flemming *et al.*, 2000).

The equation of motion is

$$\frac{\partial \mathbf{v}}{\partial t} + \mathbf{v} \cdot \nabla \mathbf{v} + 2\Omega_K \hat{\mathbf{k}} \times \mathbf{v} - 3\Omega_K^2 x \mathbf{i} = -\frac{\nabla P}{\rho} - \nabla \Phi_p + \frac{(\nabla \times \mathbf{B}) \times \mathbf{B}}{4\pi\rho}. \quad (9)$$

The continuity and induction equations (3) and (5) can be used straightforwardly in the local box centered Cartesian coordinate system. The term $\propto x$ in (9) is the net first order gravitational plus centrifugal acceleration. As for the global disc simulations, the z dependence of the gravitational potential and therefore vertical stratification have been neglected. Thus the simulations are of unstratified boxes of the type considered by Hawley *et al.* (1995). We adopt a strictly isothermal equation of state $P = \rho c^2$ and the standard boundary conditions corresponding to periodicity in shearing coordinates. The sound speed c defines a natural length scale $H = c/\Omega_K$. This would correspond to the vertical scale height were the box stratified. We adopt a box of extent $8H$ in the x direction, $4\pi H$ in the y direction and H in the z direction. This is extended by factors of 8 and 2 in the x and y directions, respectively, when compared to the standard boxes used for MHD simulations of this type. This is because a box of the dimensions we use is required to represent the response of the disc to the forcing of a protoplanet adequately.

Nonetheless because of the periodicity conditions the calculation in a shearing box is affected by protoplanets in neighboring boxes. Such effects might be manifest at a low level when considering low mass protoplanets which provide propagating wave responses in the linear regime that suffer only small damping. However, they are less apparent when the protoplanet mass is large enough to make the response nonlinear, when gap formation occurs and turbulence is present. Then it is difficult for significant disturbance to propagate to and from neighboring boxes.

The planet, when present, is located in the center of the box exerting a softened gravitational potential given by

$$\Phi_p = -\frac{Gm_p}{\sqrt{x^2 + y^2 + b^2}}, \quad (10)$$

where the softening parameter $b = 0.3H$. But at distances exceeding $3H$ from the protoplanet, Φ_p is taken as constant so as to match continuously the form at smaller distances in order to satisfy the box periodicity conditions. For box simulations the dimensionless units of length and time are H and Ω_K^{-1} , respectively. The simulation reported here was performed on a (261, 200, 35) grid. It was initiated with a weak vertical magnetic field of wavelength H in x and a random radial velocity with magnitude not exceeding ten percent of the sound speed applied at each grid point.

2.3. Vertically, horizontally and time averaged stresses, angular momentum transport and energy dissipation

In order to describe average properties of the turbulent models, we use quantities that are both vertically and azimuthally averaged over the (ϕ, z) domain (e.g., Hawley, 2000) together with an additional time average. The vertical, horizontal and time average of Q is defined through

$$\bar{Q}(r, t) = \frac{\int_{t-\Delta}^{t+\Delta} \int_S \rho Q dz d\phi d\tau}{\int_{t-\Delta}^{t+\Delta} \int_S \rho dz d\phi d\tau}. \quad (11)$$

Here the spatial averaging is over the vertical and azimuthal domain S while the time average is taken over an interval of the running time variable τ of width 2Δ centered on some time t . Thus the averaged state variables so defined remain functions of both the radius, r , and the time, t , used to center the average. In addition we work in the inertial frame ($\Omega_p = 0$).

We can derive equations governing the above defined averages of state variables from the basic equations (4), (5) by appropriate averaging.

In this way the averaged continuity (3) yields

$$\frac{\partial \Sigma}{\partial t} + \frac{1}{r} \frac{\partial}{\partial r} (r \Sigma \bar{v}_r) = 0, \quad (12)$$

where the disc surface density is given by

$$\Sigma = \frac{1}{A} \int_{t-\Delta}^{t+\Delta} \int_S \rho dz d\phi d\tau, \quad (13)$$

where $A = 2\phi_2\Delta$. (Note that the averaging of the basic equations is performed without first multiplying by the density.)

In order to quantify angular momentum transport we find it convenient to define vertically and azimuthally averaged Maxwell and Reynolds stresses through

$$T_M(r, t) = A\Sigma \overline{\left(\frac{B_r(z, r, \phi, \tau) B_\phi(z, r, \phi, \tau)}{4\pi\rho} \right)} \quad (14)$$

and

$$T_{Re}(r, t) = A\Sigma \overline{\delta v_r(z, r, \phi, \tau) \delta v_\phi(z, r, \phi, \tau)}, \quad (15)$$

respectively. The velocity fluctuations δv_r and δv_ϕ are defined through

$$\delta v_r(z, r, \phi, \tau) = v_r(z, r, \phi, \tau) - \bar{v}_r(r, t), \quad (16)$$

$$\delta v_\phi(z, r, \phi, \tau) = v_\phi(z, r, \phi, \tau) - \bar{v}_\phi(r, t). \quad (17)$$

The Shakura and Sunyaev (1973) α stress parameter appropriate to the total stress is given by

$$\alpha(r, t) = \frac{T_{Re} - T_M}{A\Sigma P/\rho}. \quad (18)$$

Averaging of the azimuthal component of the equation of motion (4) gives

$$\frac{\partial}{\partial t} (\Sigma \bar{j}) + \frac{1}{r} \left[\frac{\partial}{\partial r} (r \Sigma \bar{v}_r \bar{j}) + \frac{\partial}{\partial r} (\Sigma r^2 \alpha \bar{P}/\rho) \right] = -\frac{1}{A} \int \rho \frac{\partial \Phi}{\partial \phi} d\phi dz d\tau \quad (19)$$

(see Balbus and Papaloizou, 1999). Here $j = rv_\phi$ is the specific angular momentum. Note that the last term, which vanishes in the case of an axisymmetric potential as in an isolated Keplerian disc, corresponds to the time averaged torque per unit area produced by a perturbing object such as an orbiting planet.

Using (12), (19) may also be written as

$$\Sigma r \left(\frac{\partial \bar{j}}{\partial t} + \bar{v}_r \frac{\partial \bar{j}}{\partial r} \right) + \frac{\partial}{\partial r} \left(\Sigma r^2 \alpha \bar{P}/\rho \right) = -\frac{r}{A} \int \rho \frac{\partial \Phi}{\partial \phi} d\phi dz d\tau. \quad (20)$$

From (4) we may derive the rate of energy dissipation and doing PdV work per unit volume (see Steinacker and Papaloizou, 2002) in the conservation law form without averaging

$$\frac{\partial}{\partial t}(\rho\epsilon) + \nabla \cdot \mathbf{F} - \rho \frac{\partial \Phi}{\partial t} = -Q_T, \quad (21)$$

where the energy per unit mass and flux are, respectively,

$$\epsilon = \frac{1}{2}|\mathbf{v}|^2 + \Phi + \frac{\mathbf{B}^2}{8\pi\rho}$$

and

$$\mathbf{F} = \rho\mathbf{v} \left(\frac{1}{2}|\mathbf{v}|^2 + \Phi + \frac{|\mathbf{B}|^2}{4\pi\rho} + \frac{P}{\rho} \right) - \frac{(\mathbf{v} \cdot \mathbf{B})\mathbf{B}}{4\pi}.$$

As our aim is to derive averaged equations that are appropriate to near Keplerian discs and which can also be related to the classical viscous or ' α ' disk theory (see, e.g., Pringle, 1981) we define an energy per unit mass $\epsilon_k(r)$ and specific angular momentum $j_k(r) = rv_k(r)$. These can be used to define an angular velocity

$$\Omega_k = \left(\frac{d\epsilon_k(r)}{dr} \right) / \left(\frac{dj_k(r)}{dr} \right),$$

and may, but do not have to, be chosen to correspond to a free particle in circular Keplerian orbit. Note that the quantities $\epsilon_k(r)$, $j_k(r)$, $\Omega_k(r)$ are independent of time.

Averaging (21) and (19) after multiplying by Ω_k , we obtain

$$\begin{aligned} \frac{\partial}{\partial t}(\Sigma\bar{\mathcal{E}}) + \frac{1}{r} \frac{\partial}{\partial r}(r\bar{\mathcal{F}}) + \Sigma [\bar{v}_r(\bar{v}_\phi - v_k) + \alpha(r)\bar{P}/\rho] r \frac{d\Omega_k}{dr} = \\ -\frac{1}{A} \int Q_T dz d\phi d\tau + \frac{1}{A} \int \rho \left(\frac{\partial \Phi}{\partial t} + \Omega_k \frac{\partial \Phi}{\partial \phi} \right) dz d\phi d\tau. \end{aligned} \quad (22)$$

Here we have

$$\mathcal{E} = \epsilon - \epsilon_k - \Omega_k(\bar{j} - j_k)$$

and

$$\begin{aligned} \mathcal{F} = \Sigma v_r \left(\epsilon - \epsilon_k - \Omega_k(\bar{j} - j_k) + \frac{P}{\rho} + \frac{|\mathbf{B}|^2}{8\pi\rho} - r\Omega_k \hat{\phi} \cdot (\mathbf{v} - \bar{\mathbf{v}}) \right) \\ - \Sigma (\mathbf{v} - r\Omega_k \hat{\phi}) \cdot \frac{\mathbf{B}B_r}{4\pi\rho}. \end{aligned}$$

Equations (20) and (22) which govern averaged state variables are exact. In order to derive expressions from them which can be regarded as having equivalence to classical viscous disc theory we have to assume that if the time average is carried out over a sufficiently long interval 2Δ , we can neglect the time variation of the mean specific angular momentum \bar{J} . Then, as in the normal viscous disc theory with no perturbing protoplanet, we expect to have an expression for the mean radial velocity of the form

$$\Sigma r \bar{v}_r \frac{\partial \bar{J}}{\partial r} = - \frac{\partial}{\partial r} \left(\Sigma r^2 \bar{P}/\rho \right). \quad (23)$$

But note that simulations show that the typical velocity fluctuation is much larger than the mean \bar{v}_r (see Steinacker and Papaloizou, 2002; Papaloizou and Nelson, 2003, and below). Therefore a long period of time averaging is expected to be necessary in order to eliminate the effects of velocity fluctuations.

Similarly it can be argued (adopting Keplerian values for ϵ_k, j_k) that in a thin disk that is in a time average near steady state, so that $\bar{v}_r \sim \alpha \bar{P}/\rho / (r \Omega_k)$ as implied by (23), the term proportional to α on the left-hand side of (22) is second order in c/v_ϕ , while the others are at least third order (see also Balbus and Papaloizou, 1999).

Then with no perturbing planet present we have

$$\alpha(r) |\bar{P}/\rho| r \frac{d\Omega_k}{dr} = - \frac{1}{A} \int Q_T dz d\phi d\tau,$$

which is to be compared with the relation of classical viscous disc theory (e.g., Pringle, 1981)

$$\nu \left(r \frac{d\Omega_k}{dr} \right)^2 = \frac{1}{A} \int Q_T dz d\phi d\tau$$

with ν being the kinematic viscosity. Again, to obtain this we require that time averaging for long enough to reduce the effect of velocity fluctuations to a level where the underlying mean flow is revealed is possible. This may be problematic unless the underlying disc evolves on a very long timescale such that a proper separation between that timescale and the timescale required for averaging is possible. That in turn may mean that one should only consider classical viscous disc theory to be appropriate for discs with mean state variables evolving on the long viscous diffusion timescale r^2/ν .

3. Simulation results

We now describe the numerical simulations. Model 1 and model 2, prior to the introduction of a perturbing planet, attain a quasi-steady state MHD turbulence that is characteristic of simulations starting with magnetic fields with zero net magnetic flux that are not too strong. This is also the case for the local simulation carried out in a shearing box.

In order to characterize this behavior, we consider the time dependent evolution of the total magnetic energy in the Keplerian domain expressed in units of the volume integrated pressure, $1/\langle\beta\rangle = \int_V (|\mathbf{B}|^2/8\pi) dV / \int_V P dV$. In the above the averaging indicated by the angular brackets is over the spatial domain at a fixed time. Clearly this quantity can be defined for both the global and the local models.

The ratio of the total magnetic energy to volume integrated pressure $1/\langle\beta\rangle = \int |\mathbf{B}|^2/(8\pi)dV / \int P dV$ is plotted as a function of time for model 1 in Fig. 1. This model was initiated with a purely toroidal magnetic field and was run for up to 2300 time units corresponding to 366 orbits at $r = 1$ and 33 orbits at $r = 5$. The initial value of $1/\langle\beta\rangle$ was 0.03. After the onset of the MRI, some loss of the rather high initial magnetic energy due to reconnection occurs, and a relaxed turbulent

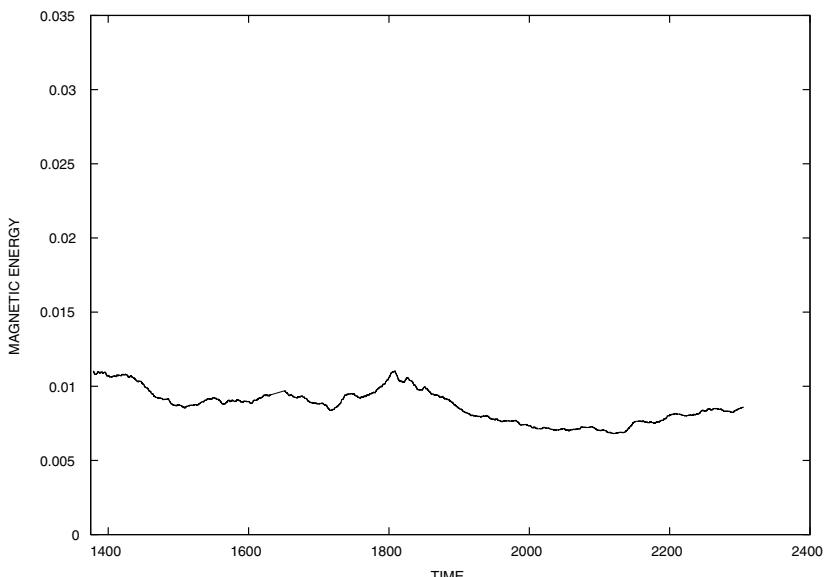


Figure 1. The magnetic energy in the Keplerian domain expressed in units of the volume integrated pressure is plotted as a function of time for model 1 for $t > 1378.6$.

state is attained after about ten orbits at the outer boundary of the Keplerian domain. The statistical properties of this do not depend on the initial conditions for models with zero net flux (see Steinacker and Papaloizou, 2002, and below).

By comparison, model 2 was initiated with a vertical magnetic field. Like model 1 it had a value of $c/(r\Omega) = 0.1$ throughout the Keplerian domain but that was larger in this case with the inner edge located at $r = 1.2$ and the outer edge at $r = 7.2$. As for model 1, this model was begun with an azimuthal extent of $\pi/3$.

The temporal evolution of the magnetic energy in units of the volume integrated pressure is shown in Fig. 2. Because model 2 was initiated with a vertical field, the initial evolution differs from that displayed by model 1 on account of the appearance of the well known channel solutions (see, e.g., Steinacker and Papaloizou, 2002). This causes an early rise in $1/\langle \beta \rangle$ followed by a fall and approach to a saturated turbulent state. The value of $1/\langle \beta \rangle \sim 0.01$ in the saturated turbulent state is similar to what is found in model 1. This supports the idea that the final state for fields with zero net flux is, as is expected, independent of initial conditions. The final state that is simulated numerically has been seen to maintain a dynamo for up to 50 orbits at the outer boundary of the Keplerian domain and most likely persists as long as the disc itself does or is not affected by interaction with the outermost boundaries.

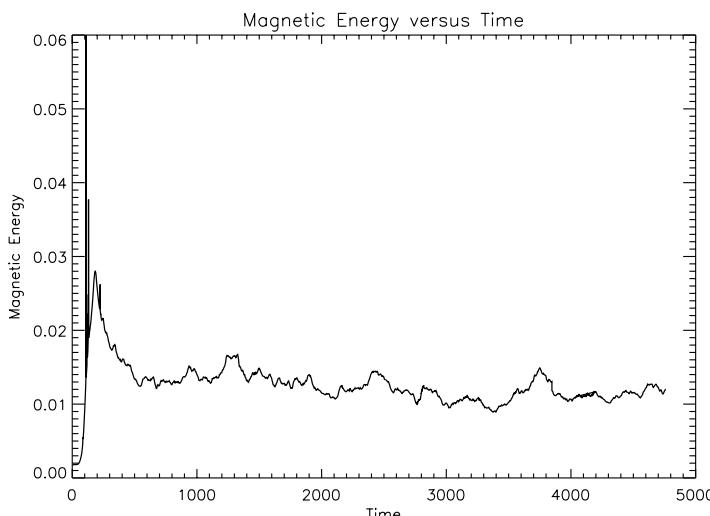


Figure 2. The magnetic energy in the Keplerian domain expressed in units of the volume integrated pressure is plotted as a function of time for model 2.

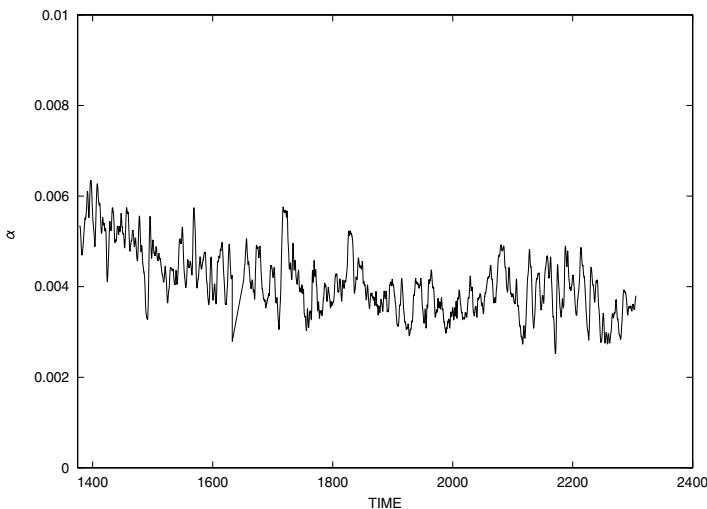


Figure 3. The stress parameter α , volume averaged over the Keplerian domain is plotted as a function of time for model 1.

The stress parameter α , defined in (18), volume averaged over the Keplerian domain is plotted as a function of time for model 1 in Fig. 3. After turbulence has become established, values are typically ~ 0.004 .

We now discuss the variation of the stress parameter α in space and time for the global models and discuss their relationship to the expectations of classical viscous disc theory (e.g., Pringle, 1981). This can only be done in a meaningful manner through the introduction of time averaging. Time averages of the stress parameter α are plotted as a function of dimensionless radius for model 1 in Fig. 4. We consider five different time averages. In each case the time averaging starts at $t = 1378.6$, ending at times $t = 1391.6$, $t = 1430.0$, $t = 1549.7$, $t = 1646.4$ and $t = 1841.6$, in each of the five cases, respectively.

Although a snapshot of α may reveal quite large variations (see also, e.g., Steinacker and Papaloizou, 2002) after quite a short time, even though noise is apparent, the beginning of a stable picture starts to emerge. This is apparent after time averaging over an interval as short as 12 units, which represents 2 orbits at the inner boundary and only 0.2 orbits at the outer boundary. Although from Fig. 4 there is still some erratic behavior visible for an averaging of 50 time units, for averaging periods exceeding 70 units, or one orbital period at the outer boundary of the active domain, the time averaged stress parameter appears to be

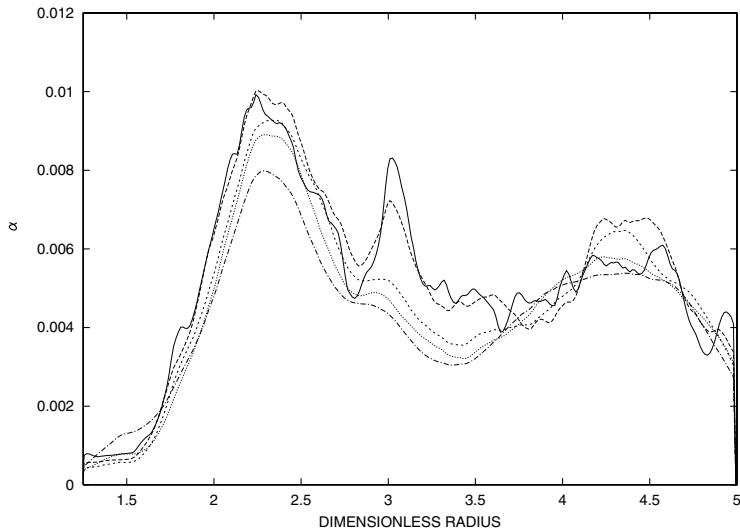


Figure 4. Five different time averages of the stress parameter α are plotted as a function of dimensionless radius for model 1. In each case the time averaging starts at $t = 1378.6$ ending at $t = 1391.6$ solid line, $t = 1430.0$ dashed line, $t = 1549.7$ short dashed line, $t = 1646.4$ dotted line and $t = 1841.6$ dot dashed line in each of the five cases, respectively.

a reasonably smooth function of r . The variation of the time averaged α in the active domain is between 0.008 and 0.003. These values are typical of those seen in local shearing box simulations starting with magnetic fields with zero net flux (Hawley *et al.*, 1996; Brandenburg *et al.*, 1996; Brandenburg, 1998; Fleming *et al.*, 2000). Smaller values are obtained near the inner boundary. However, at this stage, magnetic field has not yet diffused to the inner boundary and a long term turbulent steady state has not been reached there. This is because the initial magnetic field was applied in an annulus away from the boundary domains as described above.

For comparison, the volume averaged stress parameter, α , is plotted as a function of time for model 2 in Fig. 5. Once turbulence is established, this shows a similar behavior to model 1, saturating at a value of $\sim 5 \times 10^{-3}$.

The radial variation of the time averaged stress parameter α is shown in Fig. 6, with the solid line denoting the total α , and the dashed line denoting the magnetic contribution. The process of time averaging was started at $t = 3825.5$ and ended at $t = 4244.1$, corresponding to

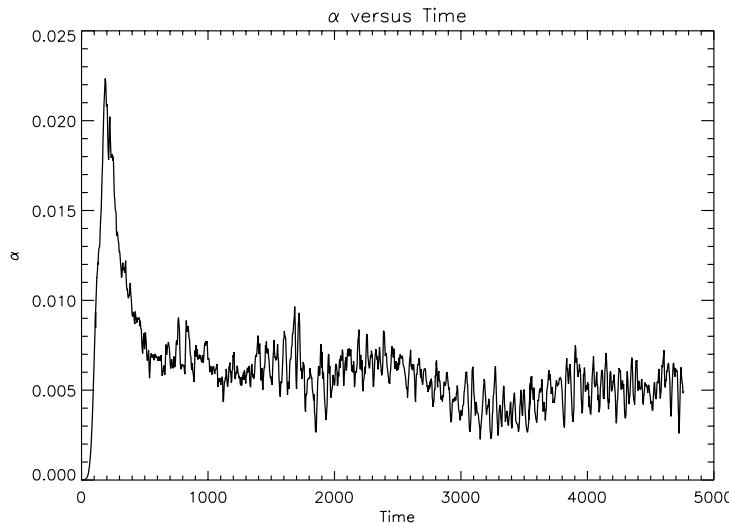


Figure 5. The stress parameter α volume averaged over the Keplerian domain is plotted as a function of time for model 2.

an interval of 66.7 orbits at $r = 1$ and 3.5 orbits at $r = 7.2$. The form is qualitatively similar to that found in model 1 as displayed in Fig. 4. This also supports the notion that the saturated turbulent state does not depend on initial conditions.

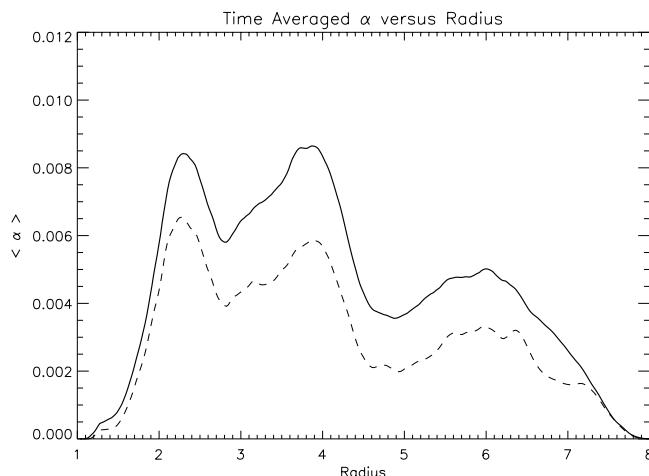


Figure 6. The time averaged stress parameter α as a function of radius for the model 2.

In general, although noise is always present, the time averaged stress parameter α , and because of the relatively small fluctuations in time of Σ , the time averaged stress itself attains a stable pattern after only a short period of time averaging.

The existence of the non-zero average stresses means that angular momentum is being transported outward through the disc. As in standard accretion disc theory we expect that this should be associated with some mass accretion toward the center. To illustrate this we plot a snapshot of $\Sigma/L_z = 1/(AL_z) \int \rho d\phi dz$ as a function of radius for model 1 at $t = 1841.6$ in Fig. 7 (where L_z is the vertical extent of the computational domain). The temporal fluctuations in this quantity are found to be small. As it represents the density averaged over the azimuthal and vertical domains it indicates some mass accretion in the inner regions. For comparison, the azimuthally averaged midplane density as a function of radius is plotted in Fig. 8 for model 2 at $t = 3825.5$. The profile is qualitatively similar to that of the surface density that was found for model 1.

The density has a significant variation of up to a factor of three with azimuth and z . To illustrate this a grayscale plot of the density at the vertical midplane of the disc is shown in Fig. 9 for model 2 corresponding to this time. Trailing waves generated by the turbulence can clearly be seen in this figure.

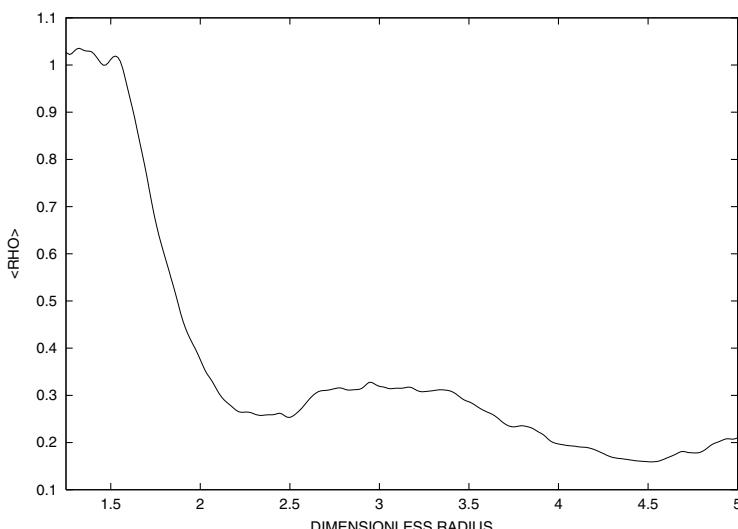


Figure 7. A snapshot of Σ/L_z as a function of radius for model 1 at $t = 1841.6$.

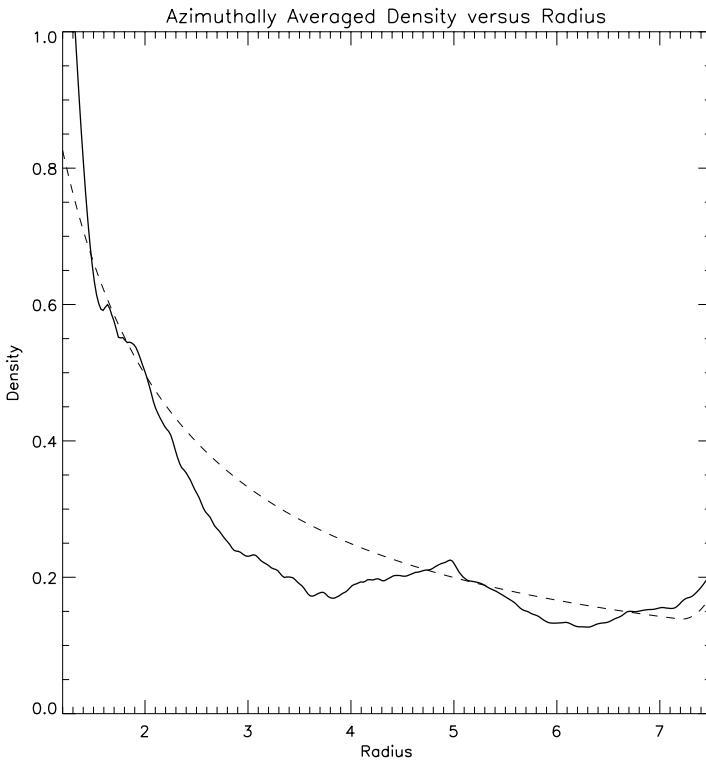


Figure 8. The radial variation of the azimuthally averaged density at the vertical mid-plane for model 2 at time $t = 3825.5$. In addition the dashed line shows the initial values at $t = 0$.

3.1. The time averaged radial velocity in the global models

A stable pattern for the time averaged radial velocity takes significantly longer to emerge. This is because from viscous disc theory we expect a characteristic value $\sim 1.5\alpha(H/r)^2(r\Omega) \sim 8 \times 10^{-5}r^{-1/2}$. However, a snapshot of the radial velocity averaged over the vertical and azimuthal domain is typically between one and two order of magnitudes higher (Hawley, 2000; Steinacker and Papaloizou, 2002). This means that the temporal fluctuations in the radial velocity are much larger in magnitude than the mean value. Accordingly averaging for a large number of characteristic times Ω^{-1} is required for these fluctuations to be averaged out to the extent that the underlying mean radial velocity is revealed.

A snapshot of the product of the vertically and azimuthally averaged values of v_r , with Σ/L_z , is plotted as a function of dimensionless radius

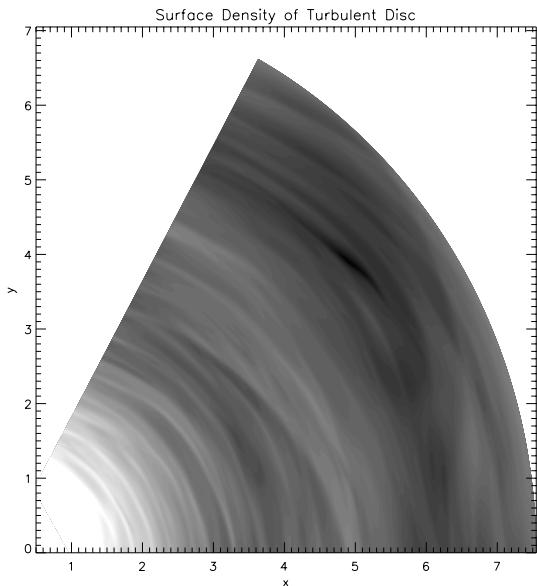


Figure 9. This plot shows the variations in density at the vertical midplane for disc model 2.

for model 1 at $t = 1841.6$ in Fig. 10. This quantity is related to the instantaneous radial mass flux. Characteristic values of the averaged v_r are seen to be much larger than those $\sim 5 \times 10^{-5}$ expected from classical viscous disc theory. Such values can only be recovered after a long period of time averaging.

To illustrate this, we plot the product of the time averages of the vertically and azimuthally averaged values of v_r with the time averaged Σ/L_z , as a function of dimensionless radius in Fig. 11 for model 1. As previously, the averages end at times $t = 1430.0$, $t = 1549.7$, $t = 1646.4$ and $t = 1841.6$. The time averaging starts at $t = 1378.6$. Even the longest two averages over 463 and 268 time units, although indicating magnitudes comparable to those expected from viscous disc theory, deviate significantly. The shortest average over 52 time units is very different in character. From this we conclude that periods of time up to 7 orbital periods at the outer boundary of the active domain are required to obtain radial velocities that can be compared with viscous disc theory, and then the comparison can be moderately successful.

We illustrate this by plotting the product of the longest time average of the vertically and azimuthally averaged values of v_r with the time

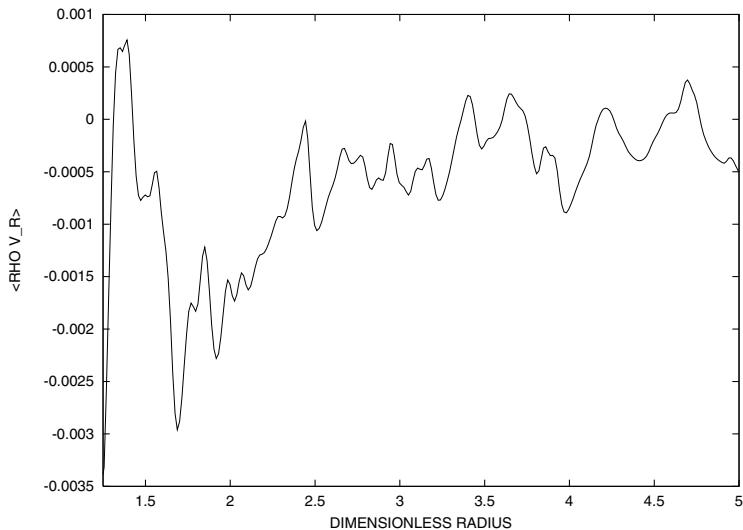


Figure 10. A snapshot of the product of the vertically and azimuthally averaged values of v_r with Σ/L_z , as a function of dimensionless radius for model 1 at $t = 1841.6$.

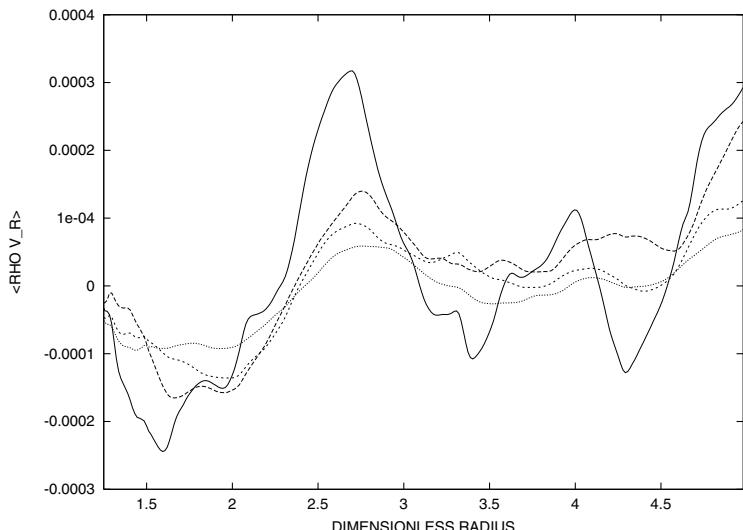


Figure 11. The product of the time average of the vertically and azimuthally averaged v_r with the time average of Σ/L_z , is plotted as a function of dimensionless radius for five different cases for model 1. The time averaging starts from $t = 1378.6$ ending at times $t = 1430.0$ (solid line), $t = 1549.7$ (dashed line), $t = 1646.4$ (short dashed line) and $t = 1841.6$ (dotted line) in the five cases, respectively.

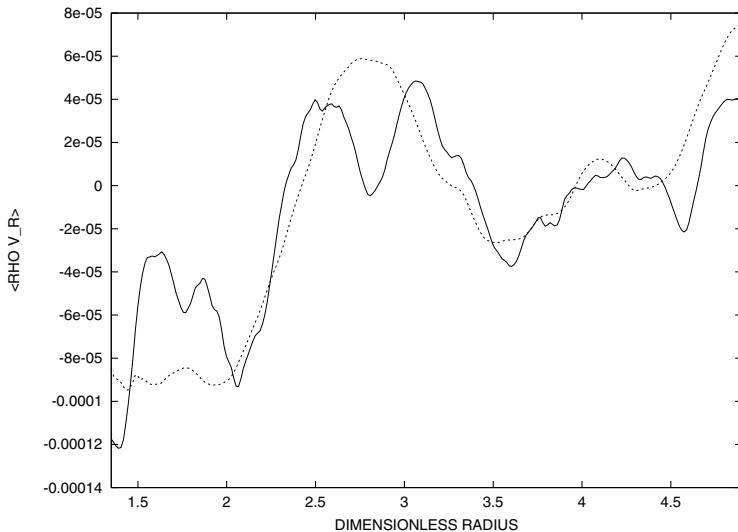


Figure 12. The product of the time average of the vertically and azimuthally averaged values of v_r with the time averaged Σ/L_z , is plotted as a function of dimensionless radius for model 1 in (dotted line). The time average starts at $t = 1378.6$ and ends at $t = 1841.6$. The value obtained from the time averaged stress using equation (23) is also plotted (solid line).

averaged Σ/L_z , as a function of dimensionless radius for model 1 in Fig. 12. The value of this quantity that we obtained from (23), using the time averaged stress $\overline{\alpha P/\rho}$ obtained in the simulation and the Keplerian value for \bar{j} , found using numerical differentiation, is also plotted (solid line). This latter quantity varies somewhat erratically because of the numerical differentiation, but nonetheless the general agreement is reasonable.

3.2. Interaction with a planet in fixed circular orbit

At a time of $t = 3825.5$ the periodicity in azimuth, as imposed by the boundary conditions of the disc model 2, was used to construct an extended model that covered the entire azimuthal extent of 2π . This was done by patching six identical copies of the $\pi/3$ model together. In addition, the frame of reference was transformed to a rotating frame with $\Omega_p = 0.30645$. This corresponds to the angular velocity of material in circular Keplerian orbit at $r = 2.2$ where the planet was inserted in circular orbit. This thus appears to be stationary in the rotating frame

used for the simulation. We found that the behavior of the models with no planet was essentially unaffected by the choice of reference frame or increasing the size of azimuthal extent of a model once that exceeded $\pi/3$. Although the inherent instability of chaotic nonlinear systems causes small deviations due to round-off error to eventually result in significant deviations, the essential statistical properties of the turbulence are unaffected (Winters *et al.*, 2003b).

The time dependent evolution of the turbulent disc extending over the full azimuth was continued together with a planet with m_p equal to 5 Jupiter masses.

However, when such a high mass planet is suddenly and discontinuously embedded in the disc, large scale transient disturbances are set up which, if they decay, take a long time to do so. As a result, we found that a steady gap could not be established for runs of less than about 100 planet orbits.

In order to work around this difficulty, a gap in the density profile, centered on the planet's radial location, was set up by hand before inserting it. The disc was then allowed to relax for a short while until the radial density profile was stabilized and the planet then inserted. The form of the azimuthally averaged density profile in the central disc midplane at the time the planet was inserted is illustrated in Fig. 13.

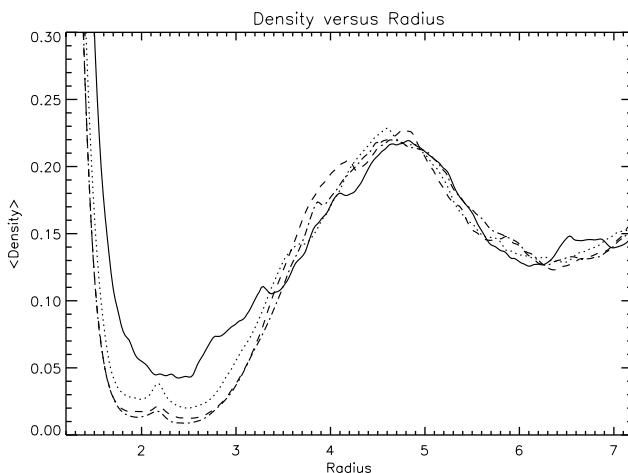


Figure 13. The azimuthally averaged density at the vertical midplane of the disc is plotted as a function of radius at $t = 0$ (solid line). The remaining curves correspond to times of $t = 675.6$ (dotted line), $t = 1351.2$ (dashed line) and $t = 2047.28$ (dot-dashed line) which correspond to 0, 33, 66 and 100 protoplanet orbits, respectively. Note that the gap deepens and widens considerably.

To emphasize that the condition of the simulation represented a significant departure from the original model 2, the running time t was reset to zero at this point.

The simulation of the disc interacting with the planet was continued for 2040 time units corresponding to ~ 324 orbits at $r = 1$, and about 100 orbits of the planet. The gap was seen to considerably deepen and accretion into the central parts of the disc occurred producing something like an inner cavity as has been noted in previous work (e.g., Nelson *et al.*, 2000). We note that this is unable to become very extensive in the present work due to our adoption of an inner boundary layer which prevents mass from flowing through the inner boundary. The behavior of the azimuthally averaged value of $\rho(r)$ at the disc midplane throughout the run is indicated in Fig. 13.

Throughout the run MHD turbulence is sustained in regions of the disc not strongly affected by the planet. But the planet has a large enough mass to strongly perturb the disc locally generating outward and inward propagating density waves.

The magnetic energy in the Keplerian domain expressed in units of the volume integrated pressure is given as a function of time in Fig. 14. A value of ~ 0.01 is maintained throughout. This is similar to what is found for models without the planet, and indicates that the perturbing presence of the protoplanet does not have a strong effect on the dynamo

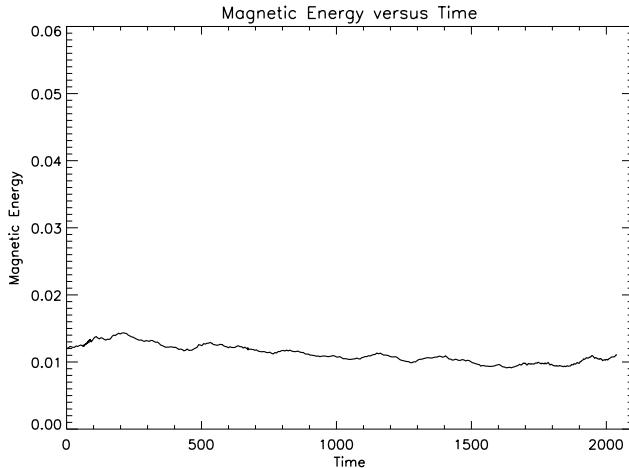


Figure 14. Magnetic energy in the Keplerian domain expressed in units of the volume integrated pressure as a function of time after the protoplanet is introduced. Note that the value of ~ 0.01 is very similar to that obtained in the absence of the protoplanet.

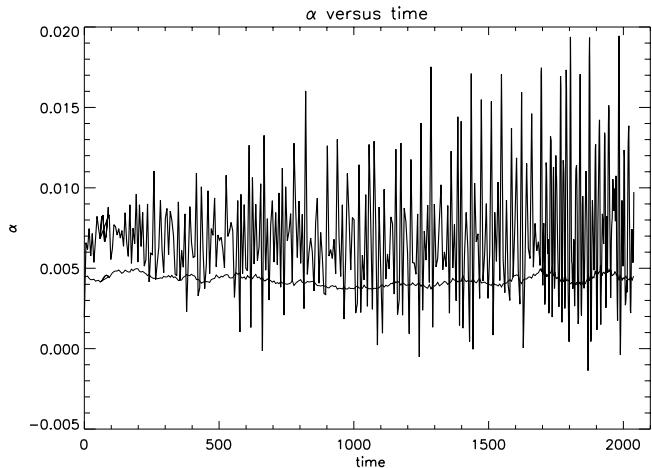


Figure 15. The stress parameter α volume averaged over the Keplerian domain is plotted as a function of time. The upper curve is derived from the total stress while the lower curve is derived from the magnetic contribution.

operating *globally* within the disc. However, there is evidence, that is also supported by local simulations, that the protoplanet affects the turbulence locally. In particular, the magnetic field is compressed and ordered where the material passes through the spiral shocks in the gap region.

The stress parameter α volume averaged over the Keplerian domain is plotted as a function of time in Fig. 15. The magnetic contribution ~ 0.004 is similar to that occurring without the planet. However, the typical total stress parameter is much larger, attaining values up to ~ 0.02 . These high values are due to the large contribution of the Reynolds' stresses associated with the spiral waves launched by the protoplanet and this effect dominate close to it. This phenomenon is also found in local simulations (see below).

3.3. Angular momentum transport and gap maintainance

We now explore the behavior of time averages of the stress parameter α and its variation with radius. This leads one naturally to a consideration of angular momentum transport within the disc and maintenance of the gap through transport associated with waves launched through the disc planet interaction working against the tendency of material to be transported through the action of magnetic stresses to fill it.

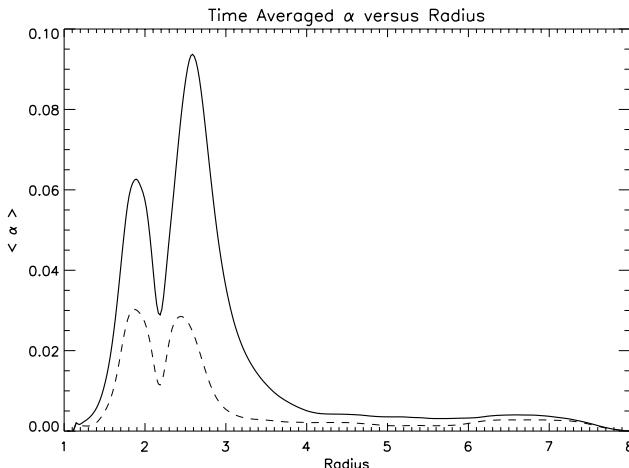


Figure 16. A time average of the vertically and azimuthally averaged value of the stress parameter α as a function of dimensionless radius. The solid curve corresponds to the total stress whereas the dashed curve gives the contribution of the magnetic stresses. The average is between $t = 1650.0$ and $t = 2040.0$.

This is just the balance between angular momentum transport induced by tidal interaction and angular momentum transport resulting from disc viscosity discussed in earlier theories of gap formation in viscous discs (e.g., Lin and Papaloizou, 1993).

A time average of the stress parameter α taken between times $t = 1650.0$ and $t = 2040.0$ is plotted as a function of dimensionless radius in Fig. 16. Both the total stress parameter α and the magnetic contribution to it become large in the vicinity of the planet $r < 3$ in the gap, but there is a dip close to the location of the planet. Such a dip is also seen in the local simulations and is probably associated with material that forms an atmosphere bound to the planet.

Although $\langle \alpha \rangle$ may rise in the gap, the actual magnetic stress $T_m = B_r B_\phi / 4\pi$ decreases there once it has formed. This is as indicated by Fig. 17, because the pressure P in the gap decreases by an even larger amount (recall that $\alpha = -T_m/P$). Beyond about two planet orbital radii values similar to those occurring in the disc without a planet are attained. This is consistent with the notion that the planet does not significantly affect the turbulence outside the gap region.

The angular momentum transport in the disc is described by (19). That contains the non-adverted part of the angular momentum flux which is proportional to $r^2 \times$ stress or $\Sigma r^2 \alpha \bar{P}/\rho$. This quantity is actually proportional to the total angular momentum flow rate through a

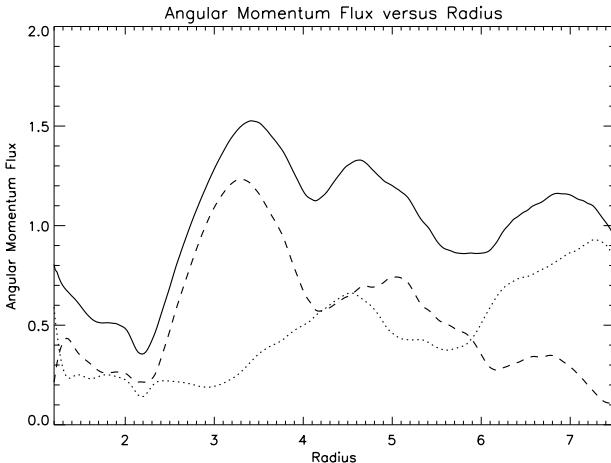


Figure 17. Time averages of the vertically and azimuthally averaged non-adverted part of the angular momentum flux in arbitrary units plotted as a function of dimensionless radius. The solid line gives the total stress while the dashed line gives the Reynolds' stress contribution and the dotted line the magnetic stress contribution. The average is taken between time $t = 1650.0$ and $t = 2040.0$.

circle of radius r . It is important because it operates when there is no mean radial flow in the disc as might be expected under quasi-steady conditions. We plot a time average of the non-adverted angular momentum flux in Fig. 17. The time average is taken over the time interval $1650 < t < 2040$. A similar form is seen if a different but long enough interval is used.

The form of the above plot may be understood in the context of the older gap formation theory as applied to laminar discs with anomalously large Navier–Stokes viscosity (e.g., Bryden *et al.*, 1999). In that theory, a gap is formed when that angular momentum flux due to turbulent stresses is smaller than that arising from waves launched by the planet.

Consider (19) under the assumption of a disc in a quasi-steady state with $\bar{v}_r = 0$. Suppose there is a completely empty gap in which the planet is located. Equation (19) predicts under these conditions that the angular momentum flux tends to a constant value at large distances from the planet. The flux is zero within the gap where there is no material, and is generated by the action of the planet through its perturbing potential (tidal torque term) in the regions of the gap where the density increases as the background disc is approached.

The behavior of the fluxes illustrated in Fig. 17 is generally in line with this view. This is indicated by the fact that the total flux dips to a

minimum at the location of the planet orbit but rises to an asymptotic value at large distances from it. The contribution from the Reynolds' stress rises near the gap edge representing the launching of spiral waves by the protoplanet in that region and beyond. The flux due to Reynolds' stresses reaches a maximum close to the outer 1:2 Lindblad resonance location. Beyond that no further launching of spiral waves is expected. In fact the contribution from the Reynolds' stresses decreases from there outward as the waves are dissipated. The contribution of the magnetic stresses rises from a small but non-zero value near the planet to become dominant at larger distances where the behavior becomes that of a turbulent disc undisturbed by a planet. It is interesting to speculate that the persistence of the magnetic stresses through the gap might enable the transmission of angular momentum from one side of the gap to the other without needing the contribution of a planetary tidal torque. This has the potential to enhance the gap formation process when compared to a non-magnetic laminar disc.

The total angular momentum budget revealed by Fig. 17 indicates a net flow away from the planet that is directed outward. This has to be supplied by the interaction with the planet which in turn results in an angular momentum loss from its orbit. This corresponds to an inward (type II) migration as can be verified by direct calculation of the torques acting on the planet itself. In line with the migration rate expected from interaction with a disc with the corresponding effective viscosity, the inward migration time is defined by $|J_{\text{planet}}/\dot{J}| \simeq 2 \times 10^4$ planetary orbits (Lin and Papaloizou, 1986, 1993; Nelson *et al.*, 2000).

As for the unperturbed turbulent disc, the radial velocity shows large temporal fluctuations about a small mean value. However, when the planet is present the mean values tend to be smaller which would be expected if planetary tides are tending to balance viscous transport. Time averages of the product of the vertically and azimuthally averaged values of Σ/L_z and v_r , in arbitrary units, are plotted as a function of dimensionless radius in Fig. 18. The averages are taken between $t = 1650.0$ and $t = 1920.0$ and between $t = 1650.0$ and $t = 2040.0$. Because of the small mean values and inadequate length of averaging time, the two averages show significant deviation. Nonetheless the net mass flow is apparently small. This is notably so interior to the planet where there is always a zero value at some radius near to that of the orbit. That corresponds to no flow from outside the planet for the duration of the averaging interval. Thus while there is material within the gap, there is no evidence of a net flow through the gap in this simulation. Note that results from laminar disc theory indicate that in some circumstances significant mass flow through gaps may occur, as has been recently

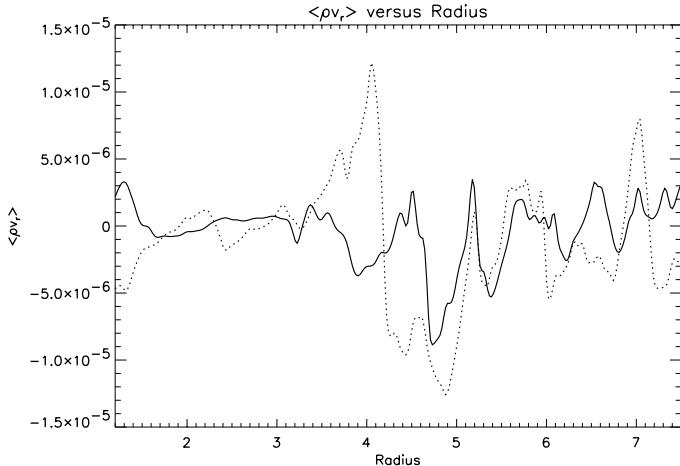


Figure 18. Time averages of the product of the vertically and azimuthally averaged values of Σ/L_z and v_r in arbitrary units are plotted as a function of dimensionless radius. For the solid curve the average is taken between time $t = 1650.0$ and $t = 2040.0$ while for the dotted curve it is taken between $t = 1650.0$ and $t = 1920.0$.

emphasized by Lubow *et al.* (1999). The simulation we have reported here suggests that different behavior may be exhibited in discs with fully developed MHD turbulence.

The density distribution in the gap region reproduces features obtained from laminar disc theory. A contour plot of the midplane density is given in Fig. 19. The wakes induced by the planet and the excitation of spiral waves are clearly visible. But because of the turbulence, the

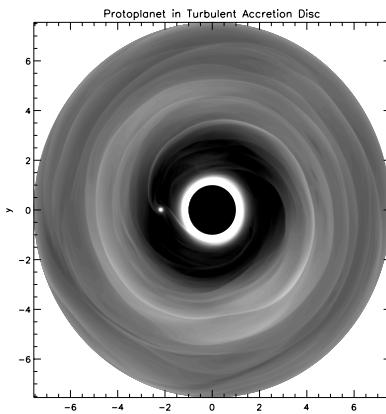


Figure 19. A typical contour plot for the midplane disc density at time $t = 1680.0$. The planet, gap, wakes and spiral waves it causes in the disc are apparent. The turbulence blurs this response relative to that obtained for laminar discs with appropriate anomalous Navier-Stokes viscosity.

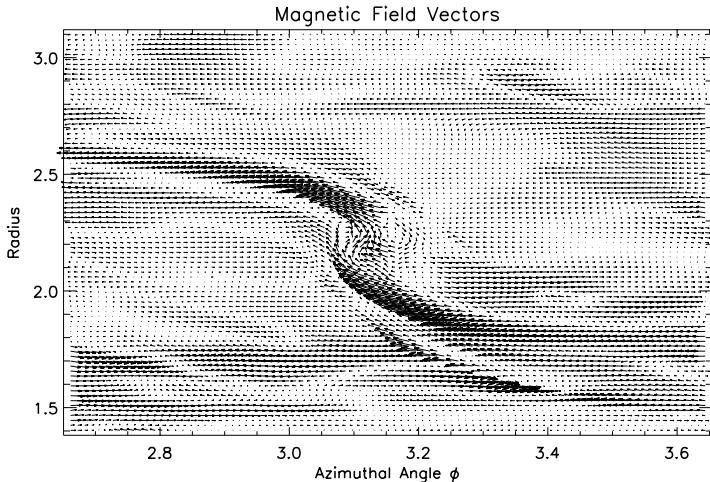


Figure 20. This figure shows the magnetic field vectors (B_r, B_ϕ) in the disc midplane plotted in the r - ϕ plane at time $t = 2040$. The field is compressed and ordered in the vicinity of the wakes induced by the perturbation of the protoplanet.

waves appear blurred when compared to those that can be produced in the laminar disc case (see, e.g., Nelson *et al.*, 2000).

In the region of the wakes the disc and turbulence are strongly perturbed. Magnetic field vectors in the disc midplane in the neighborhood of the planet are illustrated in Fig. 20, along with corresponding density plots for the equivalent region in Fig. 21. An inspection of the magnetic field vectors indicates that these tend to line up along the location of the wakes but in a somewhat broadened region slightly behind the shocks. An ordered structure appears to be imposed on the flow and magnetic field by the protoplanet. The magnetic stress is largely communicated in these regions.

3.4. A local simulation

We now describe some aspects of a local simulation carried out in a shearing box. The parameters are given in Section 2. The simulation began with no perturbing planet and starting with a small vertical magnetic field was run until time 353. This model attained a turbulent state with many similarities to that found in the global models. Values of $\langle \beta \rangle$ and $\langle \alpha \rangle$ were comparable. At time 353 a perturbing planet with $m_p/M_* = 2H^3/R^3$ was introduced and a second simulation

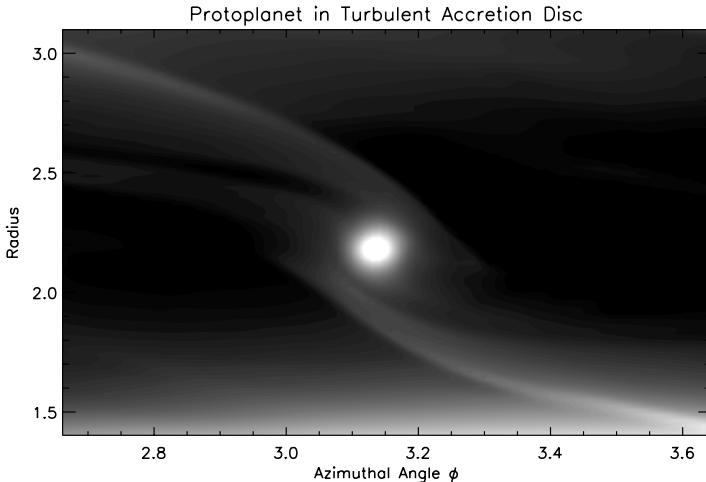


Figure 21. This is a density contour plot for the area illustrated in [fig. 20](#).

continued from then. The original first simulation was continued without the perturbing planet for purposes of comparison. Various aspects of the results are illustrated in [Fig. 22](#) (see color insert following page 234).

As expected (e.g., Korycansky and Papaloizou, 1996 and Section 1 above) the perturbing mass is large enough to open a gap and as in the global models the Reynolds' stress shows a large increase due to wave excitation. The lower panels of Fig. 22 indicate that the density structure near the planet is similar to that found in the global models and that the magnetic energy is similarly found to be concentrated in the region of the wakes.

4. Discussion

We have considered both local and global simulations of models of accretion discs undergoing turbulence driven by the magnetorotational instability (MRI). We have also considered the response of such discs to protoplanets in fixed circular orbits. Only simulations with zero net vertical and toroidal magnetic flux were considered. The local and global simulations are in essential agreement as far as the time averages of the typical magnetic energy to pressure ratio and the stress parameter α are concerned.

We have also attempted to relate the properties of the turbulent models to expectation from classical viscous disc theory (Shakura and

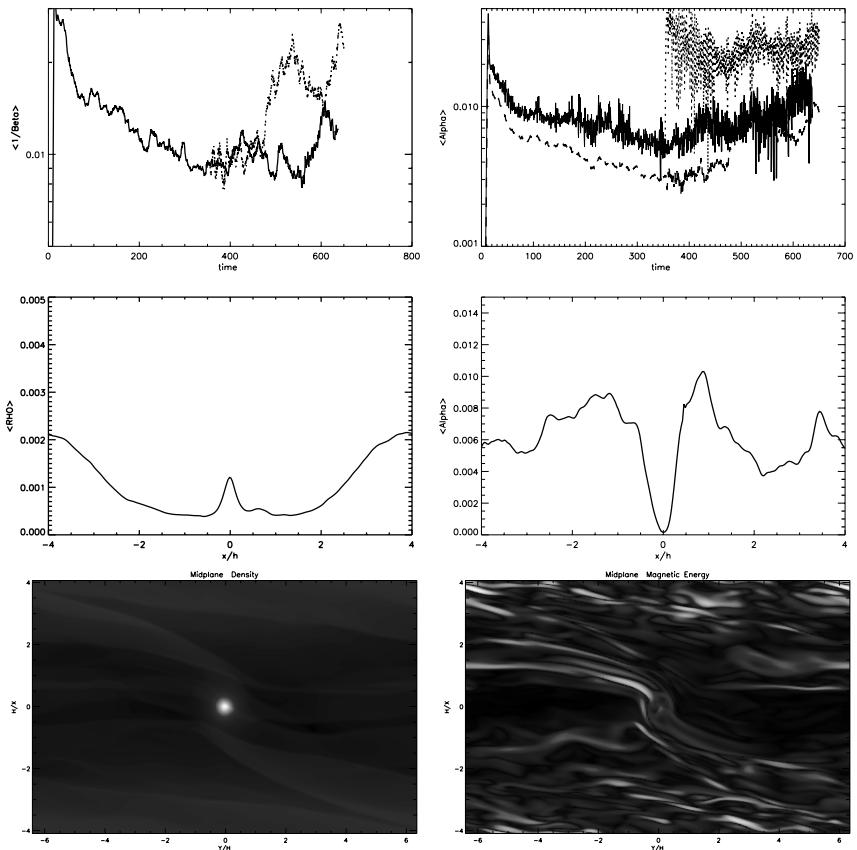


Figure 22. This figure illustrates the results of local simulations carried out in a shearing box. (See color insert.) The first simulation had no perturbing planet while in the second a perturbing planet with $m_p/M_* = 2H^3/R^3$ was introduced at time 353 and continued from then. The upper left panel shows $1/\langle \beta \rangle$ against time for the first simulation (full line) and the second beyond $t = 353$ (dashed line). The upper right panel shows $\langle \alpha \rangle$ against time for the first simulation (full line). The uppermost dotted curve corresponds to the continuation with the planet present and is dominated by the Reynolds' stress contribution. The contribution of the magnetic stress is given by the lowest curve with times beyond 353 corresponding to the second simulation. The left middle panel shows the vertically and y averaged density near the end of the second simulation indicating the gap. The right middle panel shows the vertically and azimuthally averaged total stress as a function of x near the end of the second simulation. The lower left panel is a contour plot of the midplane density near the end of the second simulation and the lower right panel is a contour plot of the midplane magnetic energy near the end of the second simulation.

Sunyaev, 1973). This is an important issue, because this still provides the conceptual framework for many accretion disc studies (Balbus and Papaloizou, 1999).

Models were found to attain a turbulent state with volume averaged stress parameter $\alpha \sim 5 \times 10^{-3}$ and mean $\beta^{-1} \sim 0.01$.

The vertically and azimuthally averaged stress parameter showed large radial fluctuations. Time averaging for a period exceeding 3 orbital periods was found to significantly reduce them.

The vertically and azimuthally averaged radial velocity showed large radial and temporal fluctuations of up to two orders of magnitude larger than the inflow velocity expected from classical viscous disc theory. Time averaging for a period of at least 7–8 orbital periods at the outer boundary of the Keplerian domain was required to reveal values of a magnitude comparable to the expected viscous inflow velocity. It is likely that very long averaging times are needed to eliminate residual fluctuations in the mean radial velocity and that such an averaging operation may only be possible for a very quiet and thin disc that evolves on a very long timescale.

Here we described some results of some of the first attempts at simulations of turbulent disc models interacting with a perturbing protoplanet with sufficient mass to significantly affect the properties of the turbulence and the disc structure in its neighborhood. Phenomena such as compression of magnetic field in the region of the planetary wakes potentially leading to coupling between disc and planet were indicated.

It is important to note that such studies are in their infancy and many physical processes of potential importance remain to be included. These include such phenomena as vertical stratification, effects of heating, cooling, radiative transfer and imperfect conductivity. It is hoped that progress will be made on incorporating some of these in the near future.

Acknowledgments

The computations reported here were performed using the UK Astrophysical Fluids Facility (UKAFF) and the GRAND consortium supercomputing facility.

References

- Armitage, P. J., “Turbulence and angular momentum transport in global accretion disk simulation,” *Astrophys. J.* **501**, L189–L192 (1998).

- Balbus, S. A. and Hawley, J. F., "A powerful local shear instability in weakly magnetized disks. I - Linear analysis. II - Nonlinear evolution," *Astrophys. J.* **376**, 214–233 (1991).
- Balbus, S. A. and Papaloizou, J. C. B., "On the dynamical foundations of alpha disks," *Astrophys. J.* **521**, 650–658 (1999).
- Brandenburg, A., "Disc turbulence and viscosity," in *Theory of Black Hole Accretion Discs*, (Eds. M.A. Abramowicz, G. Bjornsson and J.E. Pringle), pp. 61–90, Cambridge University Press (1998).
- Brandenburg, A., Nordlund, Å., Stein, R. F. and Torkelsson, U., "The disk accretion rate for dynamo-generated turbulence," *Astrophys. J.* **458**, L45–L48 (1996).
- Bryden, G., Chen, X., Lin, D. N. C., Nelson, R. P. and Papaloizou, J. C. B., "Tidally induced gap formation in protostellar disks: Gap clearing and suppression of protoplanetary growth," *Astrophys. J.* **514**, 344–367 (1999).
- D'Angelo, G., Henning, T. and Kley, W., "Nested-grid calculations of disk-planet interaction," *Astron. Astrophys.* **385**, 647–670 (2002).
- Fleming, T. P., Stone, J. M. and Hawley, J. F., "The effect of resistivity on the nonlinear stage of the magnetorotational instability in accretion disks," *Astrophys. J.* **530**, 464–477 (2000).
- Goldreich, P. and Lynden-Bell, D., "Spiral arms as sheared instabilities," *Mon. Not. R. Astron. Soc.* **130**, 125–158 (1965).
- Goldreich, P. and Tremaine, S. D., "The excitation of density waves at the Lindblad and corotation resonances by an external potential," *Astrophys. J.* **233**, 857–871 (1979).
- Hawley, J. F., "Global magnetohydrodynamical simulations of accretion tori," *Astrophys. J.* **528**, 462–479 (2000).
- Hawley, J. F., "Global magnetohydrodynamic simulations of cylindrical keplerian disks," *Astrophys. J.* **554**, 534–547 (2001).
- Hawley, J. F., Gammie, C. F. and Balbus, S. A., "Local three-dimensional magnetohydrodynamic simulations of accretion disks," *Astrophys. J.* **440**, 742–763 (1995).
- Hawley, J. F., Gammie, C. F. and Balbus, S. A., Local three-dimensional simulations of an accretion disk hydromagnetic dynamo, *Astrophys. J.* **464**, 690–703 (1996).
- Kley, W., "Mass flow and accretion through gaps in accretion discs," *Mon. Not. R. Astron. Soc.* **303**, 696–710 (1999).
- Korycansky, D. G. and Papaloizou, J. C. B., "A method for calculations of non linear shear flow: Application to formation of giant planets in the solar nebula," *Astrophys. J. Suppl.* **105**, 181–190 (1996).
- Lin, D. N. C. and Papaloizou, J. C. B., "On the tidal interaction between protoplanets and the protoplanetary disk. III-Orbital migration of protoplanets," *Astrophys. J.* **309**, 846–857 (1986).
- Lin, D. N. C. and Papaloizou, J. C. B., *On the Tidal Interaction Between Protostellar Disks and Companions, Protostars and Planets III*, pp. 749–835, University of Arizona Press (1993).

- Lubow, S. H., Seibert, M. and Artymowicz, P., "Disk accretion onto high-mass planets," *Astrophys. J.* **526**, 1001–1012 (1999).
- Marcy, G. W., Cochran, W. D. and Mayor, M., *Extrasolar Planets around Main-Sequence Stars, Protostars and Planets IV*, (Eds. V. Mannings, A.P. Boss, and S.S. Russell), pp. 1285–1312, University of Arizona Press, Tucson (2000).
- Mayor, M. and Queloz, D., "A Jupiter-mass companion to a solar type star," *Nature* **378**, 355–359 (1995).
- Nelson, R. P. and Papaloizou, J. C. B., "The interaction of a giant planet with a disc with MHD turbulence-II. The interaction of the planet with the disc," *Mon. Not. R. Astron. Soc.* **339**, 993–1005 (2003).
- Nelson, R. P. and Papaloizou, J. C. B., Masset, F. and Kley, W., "The migration and growth of protoplanets in protostellar discs," *Mon. Not. R. Astron. Soc.*, **318**, 18–36 (2000).
- Papaloizou, J. C. B. and Lin, D. N. C., "On the tidal interaction between protoplanets and the primordial solar nebula. I-Linear calculation of the role of angular momentum exchange," *Astrophys. J.* **285**, 818–834 (1984).
- Papaloizou, J. C. B. and Nelson, R. P., "The interaction of a giant planet with a disc with MHD turbulence-I. The initial turbulent disc models," *Mon. Not. R. Astron. Soc.* **339**, 983–992 (2003).
- Pringle, J. E., "Accretion discs in astrophysics," *Annu. Rev. Astron. Astrophys.*, **19**, 137–162 (1981).
- Shakura, N. I. and Sunyaev, R. A., "Black holes in binary systems. Observational appearance," *Astron. Astrophys.* **24**, 337–353 (1973).
- Steinacker, A. and Papaloizou, J. C. B., "Three-dimensional magnetohydrodynamic simulations of an accretion disk with star-disk boundary layer," *Astrophys. J.*, **571**, 413–428 (2002).
- Vogt, S. S., Butler, R. P., Marcy, G. W., Fischer, D. A., Pourbaix, D., Apps, K. and Laughlin, G., "Ten low-mass companions from the Keck Precision Velocity Survey," *Astrophys. J.* **568**, 352–362 (2002).
- Ward, W., "Protoplanet migration by nebula tides," *Icarus*, **126**, 261–281 (1997).
- Winters, W. F., Balbus, S. A. and Hawley, J. F., "Gap formation by planets in turbulent protostellar disks," *Astrophys. J.* **589**, 543–545 (2003a).
- Winters, W. F., Balbus, S. A. and Hawley, J. F., "Chaos in turbulence driven by the magnetorotational instability," *Mon. Not. R. Astron. Soc.* **340**, 519–524 (2003b).

3 Mixing at the surface of white dwarf stars

R. Rosner and A. Alexakis

*Departments of Physics and Astronomy & Astrophysics, and
Enrico Fermi Institute, The University of Chicago, 5640 S. Ellis
Avenue, Chicago, IL 60637, USA*

We have recently reconsidered the role played by shear instabilities in stratified fluids in compositional mixing at the surface of accreting white dwarf stars. The main result is that surface gravity waves localized near the interface between the C/O white dwarf and the overlying accreted envelope can be driven by a resonant instability. These waves may ultimately break; and when they do, they can effectively mix carbon and oxygen into the overlying accreted envelope. This process may explain the observed enhancement of the white dwarf envelope's C/O abundance to levels required by extant models for nova outbursts.

1. Introduction

Nova explosions – the observed nuclear incineration of accreted (mostly hydrogen) matter on the surface of white dwarf stars, and subsequent ejection of much of the “ashes” resulting from this explosion – have posed a number of long-standing puzzles regarding both the energetics governing these stellar transients and the composition of the ejecta. The puzzling aspect of the ejecta composition is related to the fact that the observed ejecta abundances are substantially enriched by CNO nuclei (cf., Truran, 1985; Gehrz *et al.*, 1998 and references therein). This is difficult to understand because these nuclei cannot be the result of nuclear burning in the accreted hydrogen envelope prior to ejection – theoretical calculations of nuclear processing during nova outbursts carried out as early as the late 1970s and early 1980s (e.g., Starrfield *et al.*, 1978; Fujimoto, 1982) clearly showed that nuclear processing during the hydrogen burning leading to the nova flash could not account for the observed CNO abundances, which can reach 30% by mass.

This early work also clearly identified the only rational source of the contaminants, namely the white dwarf star itself: White dwarfs are largely composed of carbon and oxygen; and if one were to posit some sort of “dredge-up” process that led to mixing of the stellar material with the envelope (either prior to the nova outburst, or perhaps during it), then the solution would be in hand. It turns out that if such mixing were to occur, then it would have the additional attractive feature of resolving another puzzle, namely the problem that – because the energy production rate per unit mass during runaway depends directly on the metallicity (e.g., Wallace and Woosley, 1981) – nuclear runaway in a pure hydrogen/helium envelope cannot provide enough energy to eject sufficient matter, with sufficient velocity, to match the observations (Starrfield *et al.*, 1978; Fujimoto, 1982).

2. Previous work on mixing

The above mentioned nova studies clearly identified the need for a specific mixing mechanism at the white dwarf surface; but little effective progress was made in the intervening two decades. We will now briefly discuss the early work on mixing; and suggest why so little progress was made.

2.1. The early studies

Coincidentally with the nova modeling of the late 1970s, a number of researchers at that time focused on the problem of mixing in stellar interiors driven by differential rotation, possibly driven by accretion at the stellar surface (e.g., Goldreich and Schubert 1967; Sung, 1974; Kippenhahn and Thomas, 1978). These studies established the relevant linear stability properties, based on using the Richardson number¹

$$Ri = gAl_o/U_o^2$$

(which measures the competition between the stabilizing effect of buoyancy and the destabilizing effect of the shear flow) as the control parameter. These studies are closely related to Kelvin-Helmholtz

¹ g is the local gravitational acceleration, l_o is a local characteristic length scale, and U_o is the shear flow amplitude; A is the Atwood number across the material interface separating the white dwarf surface and the H/He envelope [$\equiv (\rho_{\text{star}} - \rho_{\text{envelope}})/(\rho_{\text{star}} + \rho_{\text{envelope}})$].

instability for horizontal shear flows in the presence of (vertical) gravitational stratification. As in this latter case (see Chandrasekhar, 1962), a Richardson-like stability criterion can be established (cf., Goldreich and Schubert, 1967); and Kippenhahn and Thomas argued that the accreted matter would largely remain in a relatively narrow accretion belt.

However, MacDonald (1983) shortly thereafter demonstrated that shear instability would lead to rapid dispersal of the accreted matter over the entire white dwarf surface, but also suggested that the radial mixing time was long (set by the thermal timescale of the envelope). Furthermore, MacDonald argued that this dispersal of accreted matter might be related to the mixing of C/O into the envelope material, but did not pursue this idea in detail. It is likely that MacDonald and others realized that quantitative results for such mixing could not be obtained analytically; and as numerical simulations were insufficiently powerful at that time to get around this difficulty, little further work was done in this direction – not because the mechanism could not account for the sought-for mixing, but simply because it was too difficult to carry out quantitative calculations.²

Remarkably enough, it appears that the relative neglect of the shear mixing process was (mis)understood as a negative comment about the efficacy of shear mixing, and as a result, shear mixing was not regarded as a likely candidate for the required mixing process by the nova modeling community.³

2.2. More recent ideas about surface mixing

This situation did not materially change until the late 1990s, when sufficiently powerful simulation tools – both the numerical codes and the computers to run them – finally became available. Perhaps because of the just-mentioned negative perceptions regarding shear mixing, the new efforts to study mixing on the basis of numerical simulations focused on entirely new ideas about how surface mixing might occur.

²Indeed, it is only very recently that such shear mixing has been studied quantitatively in the nonlinear regime; that is, Brüggen and Hillebrandt (2001a,b) have begun a study of nonlinear behavior of shear-driven mixing in the stellar context by means of simulations. However, the study by Brüggen and Hillebrandt (2001a) appears to be at variance with the linear stability results; and a full resolution of this discrepancy is not yet in hand.

³In the mid-1980s, an alternative solution based on elemental diffusion at the interface was proposed by Kovetz and Prialnik (1985); however, for a variety of reasons not germane to the present discussion, this idea has not received much support in the nova community.

Thus, by the mid-90s, several authors had suggested that the convection in the accretion envelope (which was thought to start \sim 1,000 years before runaway) might lead to mixing; the specific mixing mechanism suggested involved convective undershoot and convective penetration from the H/He envelope into the C/O stellar surface layers (Shankar *et al.*, 1992). One possibility, explored by Shankar and Arnett (1994), was that local temperature fluctuations associated with convection might lead to localized “plumes” that penetrated into the C/O layers and led to mixing; others (Glasner and Livne, 1995; Glasner *et al.*, 1997; Kercek *et al.*, 1998) have studied the details of the convective undershoot and penetration zone in order to estimate quantitatively whether these penetrating flows could erode C/O material from the stellar surface and mix it into the H/He envelope. Unfortunately, Kercek *et al.* (1998, 1999) argued on the basis of a series of numerical studies that varied both the dimensionality of the flows (i.e., 2- vs. 3-D) and the spatial resolution of the computational grid that convective undershoot was most likely not an effective mixing process. Indeed, they showed that (more realistic) three-dimensional flows mixed less well than two-dimensional flows; and that less mixing occurred as grid resolution was increased.

Before abandoning this class of models entirely, it is however useful to consider the physical reasons why the existing numerical results lead to such pessimistic consequences. We recall that Rosner *et al.* (2001) could explain the results found by Kercek *et al.* (1998) quite simply: Because of current limitations on computational power, it remains true that the effective Reynolds number of the convecting flows is not high enough to lead to fully developed turbulence. Thus, the flows observed by Kercek *et al.*, while highly chaotic, are nevertheless still in the laminar regime, even at the highest grid resolution attempted by them. This is especially true in the boundary layer separating the stellar (C/O) surface and the accreted (H/He, convecting) envelope: this boundary layer is laminar, and thus its thickness depends on the viscosity (with decreasing viscosity associated with decreasing boundary layer thickness). Since the fluid viscosity associated with their numerical code is entirely dominated by numerical viscosity, and since that viscosity decreases as the spatial resolution increases, so does the boundary layer thickness. But in the laminar case, mixing depends entirely on the efficacy of convective motions in sweeping C/O material out from this boundary layer – and since this layer thins as the grid resolution increases, the amount of material subject to mixing in the boundary layer also decreases.

Note that the above argument depends entirely on the assertion that the boundary layer is laminar: if it were dominantly turbulent instead, its thickness would no longer be controlled by the viscosity (it becomes independent of the viscosity), and therefore the argument would fail. This suggests two possible ways of resolving our problem: First, it may be that in the actual physical situation, the convective motions are in fact turbulent, so that the boundary layer itself becomes dominantly turbulent, and its thickness becomes independent of the viscosity (and grid resolution); alternatively, some as yet not considered mechanism leads to a turbulent boundary layer, with the same consequence. In the following, we shall show that there is in fact a natural process that does lead to a turbulently broadened boundary layer.⁴

3. An alternative: mixing at boundaries by “wind”-driven resonant modes

For the reasons discussed above, we recently reexamined the physics of shear flow instabilities in a gravitationally stratified medium (Alexakis *et al.*, 2002, 2004; Alexakis *et al.*, 2004a,b). In the following, we extract the critical aspects of our recent results that apply to the problem at hand.

3.1. The prototypical problem

For the sake of clarity, it is most convenient to consider the simplest prototypical model, namely a shear layer created by having a low-density “wind” blow across the surface of a higher-density fluid; thus, we shall think of the accreted H/He envelope of the white dwarf as an atmosphere, which remains in relative motion with respect to the white dwarf surface (either because it retains some of the specific angular momentum it carried when accreted, or because of the convective motions driven in the envelope by the energy released from slow nuclear “smoldering”).

⁴We note here that it remains likely that the convective motions are – in the realistic astrophysical case – turbulent. At present, it is unfortunately not possible to compute definitively and quantitatively the consequent broadening of the boundary layer resulting from this turbulence.

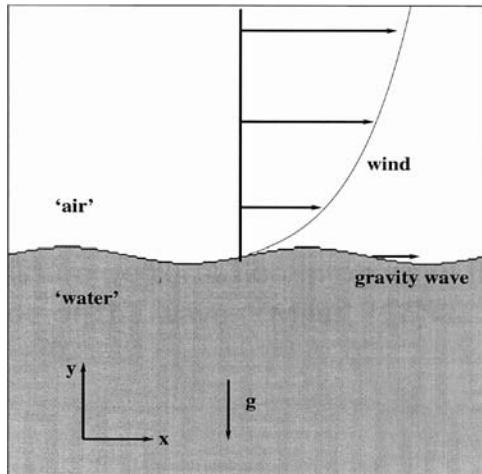


Figure 1. A sketch of the two-dimensional configuration of “air” and “water” (viz., H/He envelope and C/O stellar surface) considered in our simplest model.

For illustration, we consider a two-dimensional system, so that x and y are the horizontal and vertical directions, respectively. Our model system consists of two fluids of differing densities (ρ_1, ρ_2 constants) lying on top of each other, with $\rho_1 < \rho_2$, the heavier fluid being on the bottom. In the absence of any other complicating features, this configuration is of course stable. These two layers are separated by an interface, given by $y = h(x; t)$, which initially is taken to be flat ($y = h(x; 0) = 0$). The upper layer ρ_1 is moving with velocity $U(y)$ in the x direction parallel to the initially flat interface, while the lower layer ρ_2 remains still; this configuration is sketched out in Fig. 1.

This picture suggests some simple functional forms for the vertical velocity profile in the “atmosphere,” viz., we have considered prototypical velocity profiles of the form

$$U(y) = U_o + U_1 \ln(ay + 1) \quad (3.1)$$

for $y \geq 0$ ($U(y) = 0$ below the interface), or

$$U(y) = U_o + U_1 \tanh(ay) \quad (3.2)$$

for $y \geq 0$ ($U(y) = 0$ below the interface); U_o is the velocity jump (if any) at the interface between the H/He envelope and the C/O stellar surface, y is the vertical coordinate ($y = 0$ marks the mean envelope-star surface interface position), and a^{-1} is the characteristic vertical scale length of

the shear flow in the envelope. These two profiles are “generic” in that they sensibly represent the likely variation of the horizontal velocity of the upper (envelope) fluid near the interface. Thus, the logarithmic profile (3.1) is commonly used to describe the (turbulent) wind velocity profile in the boundary layer separating the terrestrial atmosphere from the ground or water (e.g., ocean) surface (e.g., Miles, 1957); the hyperbolic tangent profile (3.2) is similar, but has the not inconsiderable advantage of having a bounded shear velocity far from the interface. Now, the linear theory for stratified shear flows of the type encountered in ocean wave generation was thought to be relatively well understood (e.g., Chandrasekhar, 1962), but it was also well established that this linear theory predicted a cutoff in wave generation for modest wind speeds which was not observed (e.g., Munk, 1947). This difficulty was in fact the basis of the argument by Miles (1957; see also Phillips, 1957) that Kelvin-Helmholtz instability cannot be the dominant mechanism underlying the generation of terrestrial surface water waves by winds. Since we nevertheless know from direct experience that winds are able to amplify surface (gravity) waves to finite amplitude, this can only be understood if there is another mechanism at work – and indeed Miles (1957) identified such a mechanism in this landmark paper. The mechanism in question is now known as a “critical-layer instability,” and has seen very broad applications in areas well outside the present context, especially in plasma physics (see, e.g., Stix, 1992); considerable work has been done in this area, with the papers by Miles (1960) and Howard (1961) being of particular note,⁵ as well as Miles (1959 a, b, 1962, 1967, 1993), and Miles and Ierley (1998).

3.2. The linear stability problem

Critical layer instability can occur when surface gravity waves satisfying a linear dispersion relation travel at the same velocity as the wind at some height above the interface. The stability properties of such stratified shear flows has of course been investigated by others (see especially Miles, 1957; Howard, 1961), albeit under the far more limited physical circumstances related to the wind-ocean wave problem. We have now studied this problem in full generality, allowing for a variety of effects (including broad ranges in the values of the Atwood number/gravity and in compressibility), in order to establish the role of these instabilities under more general (astrophysical) circumstances. For greater details regarding this more general treatment, we refer the reader to Alexakis

⁵Lighthill (1962) provides a particularly insightful account of the physical basis of the instability in the wind-driven wave context.

et al. (2002, 2004) and Alexakis *et al.* (2004a,b), from which the present discussion is extracted.

Our aim in the present discussion is instead to focus on the simplest problem – incompressible, unstratified, no surface tension – that illustrates the main points and the basic ideas. Thus, we start with a wind (shear flow) that is assumed to flow only in the layer of light fluid ρ_1 and is zero in the heavy fluid ρ_2 . Within each layer, the governing equations are the continuity equation

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{u}) = 0, \quad (3.3)$$

and the two-dimensional Euler equation

$$\rho \left(\frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \cdot \nabla \mathbf{u} \right) = -\nabla P + \rho \mathbf{g}. \quad (3.4)$$

The system is then closed by adding an equation of state, which we express in dynamical terms,

$$\frac{\partial P}{\partial t} + \mathbf{u} \cdot \nabla P = \frac{DP}{Dt} = \frac{\gamma P}{\rho} \frac{D\rho}{Dt}; \quad (3.5)$$

γ is the polytropic exponent. The background density and pressure are assumed to be in hydrostatic equilibrium, $\partial P_o / \partial y = -\rho_o g$, with g the gravitational acceleration. The basic state is then defined by a shear flow $U(y)$ in the upper layer (e.g., flows of the type given in (3.1) and (3.2)), and hydrostatic pressure P_o and density profiles ρ_o . We then perturb around this basic state,

$$\mathbf{u} = U(y)\hat{\mathbf{x}} + \mathbf{u}', \quad \rho = \rho_o(y) + \rho', \quad P = P_o(y) + p', \quad (3.6)$$

and study the growth of the perturbations (primed variables). Using (3.5), and expanding the perturbations in normal modes $\exp[ik(x - ct)]$, we obtain the linearized equations in the perturbation quantities (where the primes have been dropped for convenience):

$$ik(U - c)u + w \frac{du}{dy} = -i \frac{k}{\rho_o} p, \quad (3.7a)$$

$$ik(U - c)w = -\frac{1}{\rho_o} \frac{\partial p}{\partial y} + \frac{g}{\rho_o} \rho, \quad (3.7b)$$

$$ik(U - c)\rho = -\rho_o \left(ik u + \frac{\partial w}{\partial y} \right) - w \frac{d\rho_o}{dy}, \quad (3.7c)$$

$$ik(U - c)p - wgp_o = c_\rho^2 \left[ik(U - c)\rho + w \frac{d\rho_o}{dy} \right]; \quad (3.7d)$$

here w is the vertical component of the perturbation velocity, and $c_s = \sqrt{\gamma P_0 / \rho_0}$ is the sound speed for the background state. These equations form an eigenvalue problem for the complex number c . As we are focusing on the (simpler) incompressible problem, we now take the limit

$$c_s \rightarrow \infty.$$

If we further define the stream function Ψ such that $u = \partial\Psi/\partial y$ and $w = -\partial\Psi/\partial x$, then the two-dimensional Euler equation will read

$$\frac{\partial}{\partial t}(\nabla^2\Psi) - \frac{\partial\Psi}{\partial x}\nabla^2\left(\frac{\partial\Psi}{\partial y}\right) + \frac{\partial\Psi}{\partial y}\nabla^2\left(\frac{\partial\Psi}{\partial x}\right) = 0, \quad y \neq h. \quad (3.8)$$

The total stream function $\Psi = \Psi_0 + \psi$ consists of a background stream function $\Psi_0 = \int_0^y U(z)dz$ and a perturbation $\psi = \phi(y) \exp[ik(x - ct)]$. The linear equation for ϕ then finally gives the well-studied Rayleigh equation,

$$\frac{d^2\phi}{dy^2} - \left(k^2 + \frac{d^2U/dy^2}{U - c}\right)\phi = 0. \quad (3.9)$$

The corresponding boundary conditions at the interface for the continuity of the normal component of the velocity and pressure are given by

$$(U - c)\tilde{h} - \phi^\pm = 0, \quad (3.10)$$

$$\Delta \left\{ \rho_i \left[(U - c) \frac{d\phi}{dy} - \frac{dU}{dy} \phi \right] \right\} + g\tilde{h}(\rho_1 - \rho_2) = 0, \quad (3.11)$$

where Δ indicates the difference across the interface, and \tilde{h} is the amplitude of the perturbed interface, $h = \tilde{h} \exp[ik(x - ct)]$. Our problem has thus simplified greatly with the incompressibility assumption; the fluid is now described by the Rayleigh equation (3.9) within each layer; and as indicated earlier, we shall ignore surface tension in all of the following in order to keep our discussion as simple as possible. Thus, if we non-dimensionalize the system by scaling lengths by a^{-1} , the characteristic length of the wind profile,⁶ scale the velocity by the reference velocity U_1 , and normalize ϕ by setting $\phi|_{y=0} = 1$, we need to solve the following dimensionless equation in each layer:

$$\frac{d^2\phi}{dy^2} - \left(K^2 + \frac{d^2V/dy^2}{V - C}\right)\phi = 0, \quad \phi|_{y=0} = 1, \quad \phi|_{y=\infty} = 0 \quad (3.12)$$

⁶In oceanography, such a length scale is referred to as the “roughness” of the wind profile.

with boundary conditions at the interface of the form

$$KC^2 - r \left[(V_0 - C)^2 \frac{d\phi}{dy} - (V_0 - C) \frac{dV}{dy} \right] - G(1 - r) = 0, \quad (3.13)$$

where $C = c/U_1$, $K = k/a$, $G = g/U_1^2 a$, $V_0 = U(0)/U_1$, and $r = \rho_1/\rho_2$. These four parameters (K, G, V_0, r) thus characterize the system, for a given wind profile; the key control parameter turns out to be G , which measures the ratio of potential energy associated with the surface wave to the kinetic energy in the wind. In the case of accretion flow on the surface of a white dwarf $G \sim 1$, while in the case of oceanic waves driven by winds, $0.1 < G < 1$. In general, we would want to solve for C in the complex plane as a function of these four parameters, and to establish the stability boundaries in the space (K, G, r, V_o) . (Note that in our convention, $\text{Im}(C) > 0$ implies instability.)

However, for present purposes, it suffices to illustrate the nature of the solutions to this linear problem by fixing three of the control parameters, namely, (K, r, V_o) , and allowing G to vary; a specific example of the results of linear theory (based on (3.12) and (3.13)) obtained by Alexakis *et al.* (2002) for this case is shown in Figs. 2, 3.

Figure 2 immediately illustrates a number of key points about the nature of surface wave instability. First, one immediately sees that indeed, as expected from previous work, Kelvin-Helmholtz instability shows a small wavenumber cutoff, as well as a cutoff for sufficiently

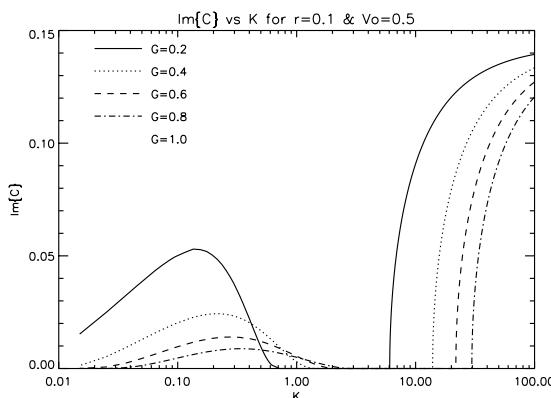


Figure 2. An example of computing the linear growth rate of unstable modes near the white dwarf/envelope interface (from Alexakis *et al.*, 2002). Note that the Kelvin-Helmholtz modes peak at large wavenumber, show a low- k cutoff, and exhibit $\omega/k \sim$ constant, while the resonant (critical layer) surface waves peak at low wavenumbers, have a high- k cutoff, and show a $\omega/k \sim k^{-1/2}$ behavior.

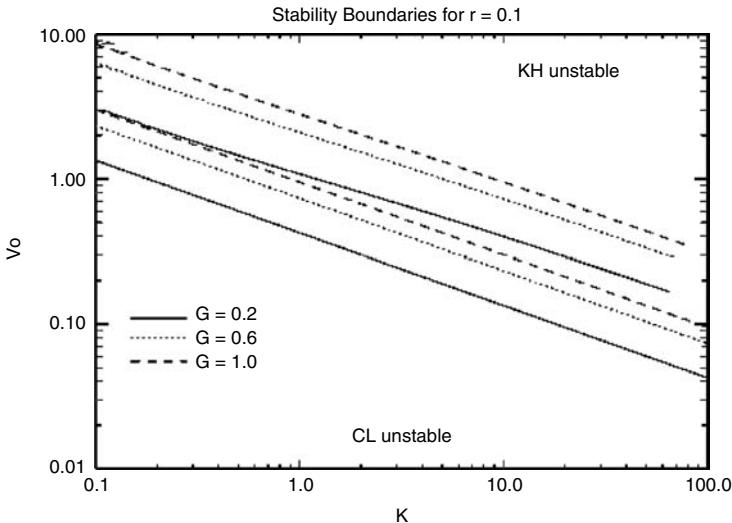


Figure 3. The stability boundaries corresponding to Fig. 2; we show the boundaries for three values of G , but for the same value of the density ratio r . Note especially the fact that Kelvin-Helmholtz modes show not only the previously mentioned low wavenumber cutoff, but also exhibit a cutoff at low wind speeds.

small wind speeds or wavenumber.⁷ It was the latter result which led to the realization that simple Kelvin-Helmholtz instability could not by itself explain the generation of observed ocean surface waves. In contrast, the resonantly driven waves can show growth at low imposed wind speed, and their small wavenumber cutoff is smaller than the corresponding cutoff for Kelvin-Helmholtz modes by a factor of ρ_1/ρ_2 ; thus, they are ideal candidates both for the observed ocean waves and for perturbing the interface between the H/He accreted envelope and C/O surface of white dwarfs.

3.3. The underlying physics and its relation to the astrophysics

Some brief comments regarding the underlying physics are now in order. This is most readily done in the context of a particularly simple

⁷Note that in the zero surface tension limit, one always obtains instability for sufficiently large wavenumbers. For the nova case, this fact is not particularly helpful since one requires instability at small wavenumbers in order to obtain effective mixing at the H/He-C/O interface.

model for the interface, in which the shear flow has step discontinuity across the density interface between the C/O white dwarf surface and the bottom of the H/He envelope, and where we ignore stratification. The physics of the classical Kelvin-Helmholtz instability follows from the observation that the flow must speed up over the “hills” of the perturbation and slow over the “valleys” – Bernoulli’s law then tells us that a low-pressure region develops over the “hills”, and a high-pressure region over the “valleys”, thus pulling up the “hills” and pushing down the “valleys”, leading to an instability whose growth rate $\gamma \sim kU$, where k is the wavenumber of the interface perturbation and U is the shear amplitude, i.e., the shortest wavelength modes grow fastest. As a consequence, the shear flow is doing work on the material near the interface, and since the available free energy in the flow near the interface is bounded, one would expect a corresponding bound on the growth of deformations of the interface. Indeed, one can understand the Richardson criterion (Chandrasekhar, 1962), as well as the two cut-offs seen in Fig. 3, as consequences of this bound. That is, there exist interface perturbations of sufficiently large horizontal scale (with wind speed fixed), or perturbations at fixed horizontal scale, but driven by a sufficiently weak wind, that are not supplied with sufficient energy by the mean shear flow to be able to lift and overturn the material near the interface. Furthermore, it is well known that the nonlinear development of Kelvin-Helmholtz instability (for wind speeds and horizontal wavenumbers that allow for linear growth) leads to the formation of vortex-like structures at the interface whose amplitude is of order the wavelength of the perturbation (viz., Drazin and Reid, 1981). As a result, the mixing region tends to be narrow, and is in any case bounded from above by the inverse wavenumber cutoff.

In contrast, consider the interaction of the same wind with the normal modes supported by the interface between the stellar surface and the accreted envelope. These normal modes – interfacial, or surface, gravity waves – are akin to “deep water waves” seen at the surface of terrestrial oceans; thus, unlike the interface perturbations associated with the Kelvin-Helmholtz instability, these modes are neutrally stable in the absence of any driving or dissipation. Now, the perturbation associated with a neutrally stable surface gravity wave extends vertically above and below the interface (with an amplitude that roughly varies exponentially with distance away from the interface), and moves at the same phase velocity

$$v_{\text{phase}} \equiv \omega/k \sim (Ag/k)^{1/2}$$

as the interfacial distortion itself. (Here we assume that surface tension is negligible, a fair assumption for the case of the gaseous media characterizing stars.) If, in the presence of a (horizontal) wind, there is a height at which this perturbation moves with the same speed as the local wind speed, i.e.,

$$v_{\text{phase}} \sim (Ag/k)^{1/2} = U(z),$$

coupling can occur, with the wind energy flowing to the (growing) interfacial gravity wave; the layer at which this resonance occurs is the critical layer (and the fact that a resonance lies at the heart of the physics of this interaction can be recognized directly from the presence of the singular term in the Rayleigh equation (3.12)). That is, a wave with wavenumber $k \sim Ag/U(y)^2$ will be driven resonantly unstable. (For typical values of A , g , and U characteristic of a white dwarf surface,⁸ one finds that the wavelength of unstable modes lies in the range 0.01–1 km.)

3.4. Modeling the mixing limit: the nonlinear regime

As pointed out earlier, the efficacy of mixing C/O into the H/He atmosphere will depend crucially on the nature of the boundary layer separating the two compositional phases. Thus, the key issue remaining is how one is to determine the mixing layer width once the unstable modes cease their growth and finally saturate. One would normally presume that some physical process intervenes to limit the ultimate mode amplitude, and thereby determine the width of the mixing layer; and the question is then what that saturation mechanism is. In the case of interfacial gravity modes, saturation is believed to occur via wavebreaking (see Chen *et al.*, 1999 and references therein; also Massel, 1996, and Longuet-Higgins, 1997); it is the resulting spray and overturning motions that then determine (from a statistical point of view) the effective mean width of the mixing layer – this width can be somewhat larger than the mode amplitude at saturation, as is well known in the case of wind-driven spray from breaking ocean waves.

Alexakis *et al.* (2004a,b) have recently carried out extensive direct numerical simulations in order to measure the effective width of this mixing layer; this is a difficult task for a number of reasons, but principally because numerical resolution effects continue to play a role

⁸We adopt values of $A \sim 1/3$, $g_{WD} \sim 2.7 \times 10^8 \text{ cm s}^{-1}$, and $U \sim 10^5 \text{ km s}^{-1}$.

unless one enters a fully developed turbulent regime. For this reason, we shall assume for present purposes that we have determined this layer width, independent of any concerns regarding the numerics; we denote this layer width as λ . In the following, we will now outline (after Rosner *et al.*, 2000) a simplified model that allows us to estimate the sought-after mixing efficacy.

We start with the amount of carbon and oxygen in the breaking wave mixing layer, which we write in the form

$$M_{\text{C+O}}^{\text{mixing layer}} \sim \alpha \rho_0^{\text{mixing layer}} \lambda \Lambda \xi,$$

where α is the coefficient for the C + O mass fraction in the mixing layer, λ is the mixing layer width (both α and λ are to be determined from simulations, e.g., Alexakis *et al.*, 2004a,b), $\rho_0^{\text{mixing layer}}$ is the density of the envelope at its base (i.e., in the breaking wave mixing layer), Λ is the characteristic length scale of the large-scale circulation (which can be identified with the outer scale of motions in the envelope – due either to convection or to accretion), and ξ is a length scale transverse to the wind direction (this dimension will drop out of our formulation). By design, the remaining parameters appearing in this relation can be obtained from extant (1-D) nova models. Now, as noted earlier, the amount of C/O to be mixed into the envelope is roughly 1/3 by mass of the ejecta mass $M_{\text{total}}^{\text{envelope}}$, or

$$M_{\text{C+O}}^{\text{envelope}} \sim \frac{1}{3} M_{\text{total}}^{\text{envelope}} \sim \frac{1}{3} \cdot \left(\frac{2}{3} \Lambda \rho_0^{\text{mixing layer}} \cdot \Lambda \xi \right).$$

The “sweepout time”, i.e., the timescale on which the boundary mixing layer is swept out by a penetrating convective roll, is just $\tau_{\text{sweep}} \sim \Lambda/U$, so that the time needed to mix the necessary amount of carbon and oxygen into the envelope is just $\tau_{\text{mixing}} \sim M_{\text{C+O}}^{\text{envelope}} / (M_{\text{C+O}}^{\text{mixing layer}} / \tau_{\text{sweep}})$, or

$$\begin{aligned} \tau_{\text{mixing}} &\sim \frac{2}{9} \Lambda^2 / \alpha \lambda U \\ &\sim 5 \alpha^{-1} (\Lambda / 10^8 \text{cm})^2 (\lambda / 10^2 \text{cm})^{-1} \\ &\quad \cdot (U / 10^5 \text{cm s}^{-1})^{-1} \text{ years}, \end{aligned}$$

with $\alpha \sim 0.3 - 1$. Evidently, the evolution timescale for the envelope prior to nova runaway (roughly of the order of the time between onset of envelope convection and runaway, or $\sim 10^3$ years) is much longer than the mixing timescale. This is a gratifying result: It suggests that resonantly driven mixing at the star-envelope boundary can be an efficient mixing process during the pre-nova star evolution.

3.5. Extending our results

The next steps are clear: first, we need to verify these results via simulations of weakly compressible fluids subject to these mixing instabilities; the recent calculations of Alexakis *et al.* (2004a,b) are an important first step in this direction, using the fully compressible, explicit adaptive mesh hydrodynamics code Flash (Fryxell *et al.*, 2000). An example of a two-dimensional simulation of the saturation process, e.g., wave breaking, is shown in Fig. 4 (see color insert following page 234);

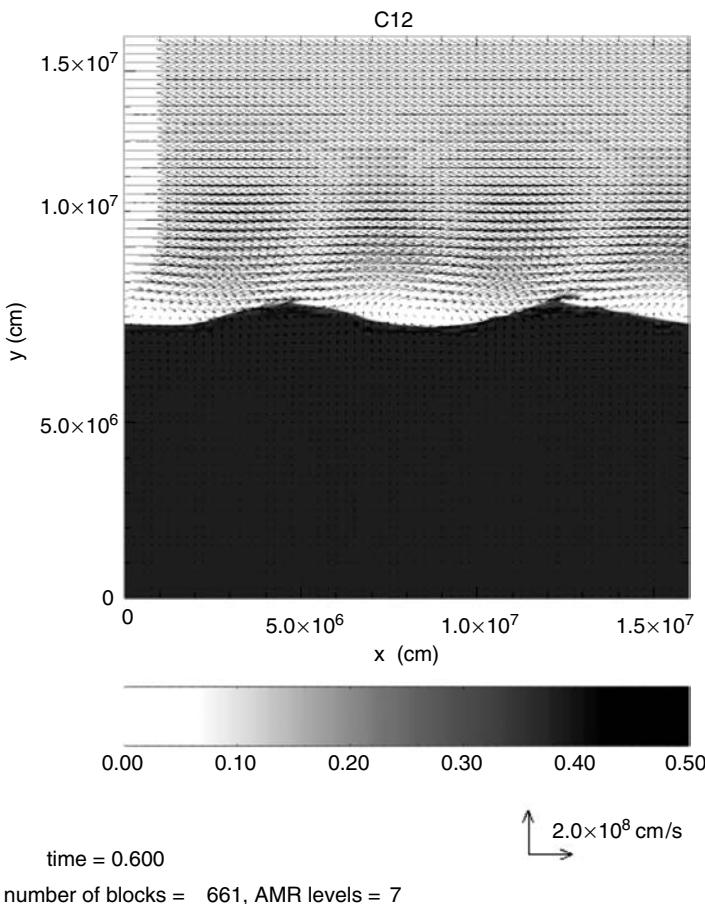


Figure 4. A snapshot of a two-dimensional simulation of a breaking gravity wave on the surface of a white dwarf, using the Flash code (from Alexakis *et al.*, 2004a). (See color insert.) The lines and arrows are meant to indicate the nature of the flow/wind overlying the perturbed interface.

Mixing Due to Wind-Wave Interaction on the Surface of a White Dwarf

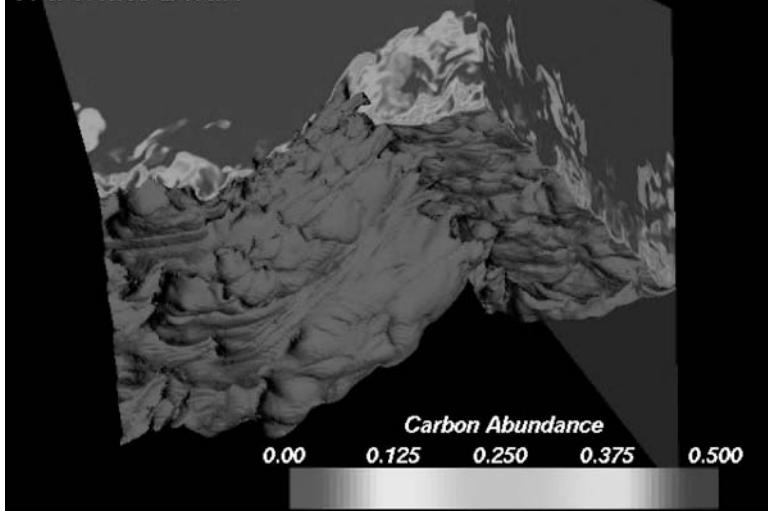


Figure 5. A snapshot of a three-dimensional simulation of a breaking gravity wave on the surface of a white dwarf. (See color insert.) The color scale indicates carbon abundance.

this calculation shows explicitly what we have previously presumed, namely the extensive mixing of material from the lower, denser layer into the upper, more tenuous layer.

Similarly, it is essential that the previously discussed results – based largely on analytical calculations and two-dimensional simulations – be extended to three dimensions. In particular, the nature of mixing in two and three dimensions can be dramatically different because the energy cascade critically depends upon dimensionality. For this reason, Alexakis *et al.* (2004a,b) have carried out a number of studies using the aforementioned Flash code; and a snapshot from one of their recent calculations is shown in Fig. 5 (see color insert following page 234). This calculation was designed to mimic as much as is practicable the physical conditions of the white dwarf surface, and provides a visual indication of the extent of mixing of C/O into the overlying H/He atmosphere.

One remaining significant issue relates to the possible effects of magnetic fields on the C/O mixing process; that is, one might be concerned that turbulent mixing may be suppressed if local magnetic fields in the envelope become large as convection sets on $\sim 1,000$ years before runaway. We are not currently in a position to resolve this possible

problem definitively, but only note that because the conservative mixing timescale $\tau_{\text{mixing}} \ll 1,000$ yrs, substantial mixing suppression by magnetic fields could be accommodated within this model without vitiating the main point, namely that resonant instability of the C/O envelope boundary can lead to effective mixing across that boundary. [Preliminary calculations by Linde (2004, private communication) of the magnetic counterpart of the calculations done by Alexakis *et al.* (2004) suggest that magnetic field indeed can significantly modify the timescale of effective mixing.] This is a critical point for any nova model because novae have been observed for white dwarfs with relatively strong magnetic fields (e.g., V1500 Cygni, 1975; Stockman *et al.*, 1988).

4. Conclusions

To conclude, by using the results of linear stability theory, as well as extrapolating from existing numerical simulations of wave breaking and nova outbursts, we have estimated the mixing zone parameters, and can show that pre-nova erosion of the wave-breaking mixing layer by slow convection could mix sufficient C/O into the accreted H/He envelope to satisfy observations. We have illustrated this point by using the simple mixing length subgrid prescription developed by Rosner *et al.* (2001) to describe this mixing process prior to the nova runaway. One consequence of our simple model is that further mixing during the outburst is no longer required. Furthermore, because the C/O abundance in the H/He envelope builds up gradually during the pre-nova slow convective phase, we expect the nova envelope mass attained before outburst can be substantially larger than in standard models assuming a “pre-seeded” envelope.

Acknowledgments

This work has been supported by the DOE-funded ASCI/Alliances Center for Astrophysical Thermonuclear Flashes at the University of Chicago. We appreciate the many fruitful discussions and collaborations with our colleagues, especially A. Calder, J. Dursi, S. A. Glasner, A. Heger, W. Hillebrandt, D. Lamb, J. Truran, and Y.-N. Young.

References

- Alexakis, A., Young, Y.-N. and Rosner, R., “Shear instability of fluid interfaces: Stability analysis,” *Phys. Rev. E* **65**, 026313 (2002). (Erratum: *Phys. Rev. E*, **65**, 059904 [2002])

- Alexakis, A., Young, Y.-N. and Rosner, R., "Weakly non-linear analysis of wind-driven gravity waves," *J. Fluid. Mech.*, to appear (2004).
- Alexakis, A., Calder, A. C., Heger, A., Brown, E. F., Dursi, L. J., Truran, J. W., Rosner, R., Lamb, D. Q., Timmes, F. X., Fryxell, B., Zingale, M., Ricker, P. M. and Olson, K., "On heavy element enrichment in classical novae," *Astrophys. J.*, to appear (2004a).
- Alexakis, A., Calder, A. C., Dursi, L. J., Rosner, R., Truran, J. W., Fryxell, B., Timmes, F. X., Olson, K. and Ricker, P. M., "On the nonlinear evolution of wind-driven gravity waves," *Phys. Fluids*, to be submitted (2004b).
- Brüggen, M. and Hillebrandt, W., "Mixing through shear instabilities," *Mon. Not. R. Astron. Soc.* **320**, 73–82 (2001a).
- Brüggen, M. and Hillebrandt, W., "Three-dimensional simulations of shear instabilities in magnetized flows," *Mon. Not. R. Astron. Soc.* **323**, 56–66 (2001b).
- Chandrasekhar, S., *Hydrodynamics and Hydromagnetic Stability*, Dover Press (1962).
- Chen, G., Kharif, C., Zaleski, S. and Li, J., "Two-dimensional Navier-Stokes simulation of breaking waves," *Phys. Fluids* **11**, 121–133 (1999).
- Drazin, P. G. and Reid, W. H., *Hydrodynamic Stability*, Cambridge University Press (1981).
- Fryxell, B., et al., "Flash: an adaptive mesh hydrodynamics code for modeling astrophysical thermonuclear flashes," *Astrophys. J.* **131**, 273–334 (2000).
- Fujimoto, M. Y., "The theory of hydrogen shell flashes on accreting white dwarfs. I. Their progress and the expansion of the envelope," *Astrophys. J.* **257**, 752–779 (1982).
- Gehrz, R. D., Truran, J. W., Williams, R. E. and Starrfield, S., "Nucleosynthesis in classical novae and its contribution to the interstellar medium," *Publ. Astron. Soc. Pacific* **110**, 3–26 (1998).
- Glasner, S. A. and Livne, E., "Convective hydrogen burning during a nova outburst," *Astrophys. J.* **445**, L149–L151 (1995).
- Glasner, S. A., Livne, E. and Truran, J. W., "Reactive flow in nova outbursts," *Astrophys. J.* **475**, 754–762 (1997).
- Goldreich, P. and Schubert, G., "Differential rotation in stars," *Astrophys. J.* **150**, 571–587 (1967).
- Howard, L. N., "Note on a paper of John W. Miles," *J. Fluid. Mech.* **10**, 509–512 (1961).
- Kercek, A., Hillebrandt, W. and Truran, J. W., "Two-dimensional simulations of the thermonuclear runaway in an accreted atmosphere of a C + O white dwarf," *Astron. Astrophys.* **337**, 379–392 (1998).
- Kercek, A., Hillebrandt, W. and Truran, J. W., "Three-dimensional simulations of classical novae," *Astron. Astrophys.* **345**, 831–840 (1999).
- Kippenhahn, R. and Thomas, H.-C., "Accretion belts on white dwarfs," *Astron. Astrophys.* **63**, 265–272 (1978).
- Kovetz, A. and Prialnik, D., "CNO abundances resulting from diffusion in accreting nova progenitors," *Astrophys. J.* **291**, 812–821 (1985).

- Lighthill, M. J., "Physical interpretation of the mathematical theory of wave generation by wind," *J. Fluid. Mech.* **14**, 385–398 (1962).
- Longuet-Higgins, M. S., "Progress toward understanding how waves break," in *Proc. 21st Symp. on Naval Hydrodynamics*, pp. 7–27, US Govt. Printing Office (1997).
- MacDonald, J., "Mixing by shear instabilities in differentially rotating inhomogeneous stars with application to accreting white dwarf models for novae," *Astrophys. J.* **273**, 289–298 (1983).
- Massel, S. R., *Ocean Surface Waves: Their Physics and Prediction*, World Scientific (1996).
- Miles, J., "On the generation of surface waves by shear flows," *J. Fluid. Mech.* **3**, 185–204 (1957).
- Miles, J., "On the generation of surface waves by shear flows. 2.," *J. Fluid. Mech.* **6**, 568–582 (1959a).
- Miles, J., "On the generation of surface waves by shear flows. 3. Kelvin-Helmholtz instability," *J. Fluid. Mech.* **6**, 583 (1959b).
- Miles, J., "On the stability of heterogeneous shear flows," *J. Fluid. Mech.* **10**, 496–508 (1960).
- Miles, J., "On the generation of surface waves by shear flows. 4.," *J. Fluid. Mech.* **13**, 433–448 (1962).
- Miles, J., "On the generation of surface waves by shear flows. 5.," *J. Fluid. Mech.* **30**, 163—and Part 1 (1967).
- Miles, J., "Surface-wave generation revisited," *J. Fluid. Mech.* **256**, 427–441 (1993).
- Miles, J. and Ierley, G., "Surface-wave generation by gusty wind," *J. Fluid. Mech.* **357**, 21–28 (1998).
- Munk, W. H., "A critical wind speed for air sea boundary processes," *J. Marine Research* **6**, 203–218 (1947).
- Peltier, W. R., "Mixing efficiency in stratified shear flows," *Annu. Rev. Fluid Mech.* **35**, 135–167 (2003).
- Phillips, O. M., "On the generation of waves by turbulent wind," *J. Fluid. Mech.* **2**, 417–445 (1957).
- Rosner, R., Alexakis, A., Young, Y.-N., Truran, J. W. and Hillebrandt, W., "On the C/O enrichment of nova ejecta," *Astrophys. J.* **562**, L177–L179 (2001).
- Shankar, A. and Arnett, W. D., "Thermonuclear runaways in nova outbursts. 2: Effect of strong, instantaneous local fluctuations," *Astrophys. J.* **433**, 216–228 (1994).
- Shankar, A., Arnett, W. D. and Fryxell, B. A., "Thermonuclear runaways in nova outbursts," *Astrophys. J.* **394**, L13–L15 (1992).
- Starrfield, S., Sparks, W. M. and Truran, J. W., "CNO abundances and hydrodynamic models of the nova outburst. II–1.00 solar-mass model with enhanced carbon and oxygen," *Astrophys. J. Suppl.* **28**, 247–270 (1974).
- Starrfield, S., Truran, J. and Sparks, W. M., "CNO abundances and hydrodynamic models of the nova outburst. V–1.00-solar mass model with small mass envelopes," *Astrophys. J.* **226**, 186–202 (1978).

- Starrfield, S., Truran, J., Sparks, W. M. and Kutter, G. G., "CNO abundances and hydrodynamic models of the nova outburst," *Astrophys. J.* **176**, 169–176 (1972).
- Stix, T. H., *Waves in Plasmas*, American Institute of Physics, Chapters 13.3–13.7 (1992).
- Stockman, H. S., Schmidt, G. D. and Lamb, D. Q., "V1500 Cygni-Discovery of a magnetic nova," *Astrophys. J.* **332**, 282–286 (1988).
- Sung, C.-H., "The stability of differential rotation with non-axisymmetric perturbations. I. The sufficient conditions," *Astron. Astrophys.* **33**, 99–104 (1974).
- Truran, J. W., "Nucleosynthesis in novae," in *Nucleosynthesis* (Eds. W. D. Arnett and J. W. Truran), p. 292, Univ. of Chicago Press (1985).
- Wallace, R. K. and Woosley, S. E., "Explosive hydrogen burning," *Astrophys. J. Suppl.* **45**, 389–420 (1981).
- Williams, J. M., "Limiting gravity waves in water of finite depth," *Phil. Trans. R. Soc. Lond. A* **302**, 139–188 (1981).
- Woosley, S. E., "Nucleosynthesis and stellar evolution," in: *Nucleosynthesis and Chemical Evolution* (Eds. B. Hauck, A. Maeder and G. Magnet), p. 1, Sauverny: Geneva Observatory (1986).

4 Pulsar magnetospheres

Leon Mestel

*Astronomy Centre, SciTech, University of Sussex, Falmer,
Brighton BN1 9QH, UK*

The seminal paper by Goldreich and Julian showed that a strongly magnetic rotating neutron star will be surrounded by a charged magnetosphere. In an axisymmetric model, with rotation and magnetic axes aligned, spin-down of the star will occur via the torques exerted by current flowing in and out of the star. Within classical physics, the angular momentum could be carried off by gamma-rays emitted beyond the light-cylinder. Magnetospheric generation of electron-positron pairs allows the angular momentum to be lost primarily via the relativistic analogue of a stellar wind. A model in which the perfect conductivity approximation remains valid has a wind zone satisfying the relativistic force-free equation, a pinched equatorial sheet, and a dead zone of field lines closing within the light-cylinder. More realistic models may allow for trans-field flow of gas via a resistivity due to macro-instabilities. The observed gamma-ray emission may originate in local gap regions in the outer magnetosphere.

1. Introduction

The standard model of a radio pulsar is pictured schematically in Fig. 1. A neutron star of radius $R \simeq 10^6$ cm has a magnetic field with polar field strength typically $\simeq 10^{12}$ G, frozen into the solid crust which rotates with angular velocity $\alpha \equiv 2\pi/P_1$. The rotation period P_1 inferred from the recurring radio emission is typically close to 1 s, but can be smaller, e.g., $\approx 1/30$ s for the pulsar in the Crab Nebula. There is also the class of “millisecond pulsars,” thought to have fields closer to 10^8 G.

Field strengths were first estimated by Gold (1969), who equated the rate of loss of rotational energy as inferred from the observed, normally monotonic, increase in P_1 , to the classical rate of magnetic dipole

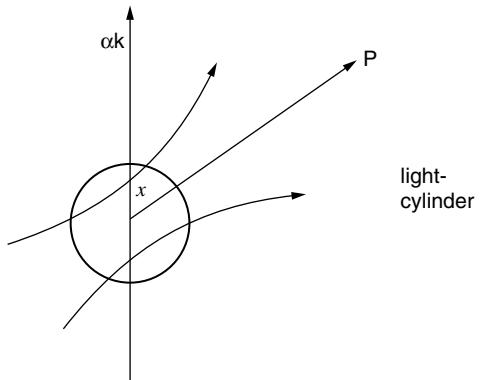


Figure 1. The obliquely rotating magnetic neutron star.

radiation with frequency α through a surrounding vacuum:

$$\frac{4\pi^2 I}{P_1^3} \frac{dP_1}{dt} = \frac{B_s^2 R^6 \alpha^4}{6c^3} \sin^2 \chi, \quad (1)$$

where I is the star's moment of inertia and χ is the angle of obliquity between the rotation axis \mathbf{k} and the magnetic dipole axis \mathbf{p} and is assumed not very small. As will be seen below, the “pulsar magnetosphere problem” arises because the vacuum description can be shown to be inadequate: in a realistic model, a particle current as well as the displacement current contributes to the energy flow. However, expression (1) with $\sin \chi$ replaced by unity is expected to remain a good estimate.

A departure from axisymmetry, as in the oblique rotator, is clearly essential for the rotational periodicity to be directly observable. However, in this report most attention is given to the mathematically simpler steady, axisymmetric state — the “non-pulsar,” with the obliquity angle $\chi = 0$, for which the vacuum radiation loss vanishes. It may be that strictly aligned bodies will always prefer to be “dead,” with zero magnetospheric currents and so vanishing spin-down, as argued, e.g., by Smith *et al.* (2001). Nevertheless, study in depth of a “live” aligned model is justified, since one can expect some of the problems that arise to occur also in the more complicated geometry of the oblique rotator.

The energy loss rate inferred from the spin-down is far greater than that observed directly as electromagnetic radiation. The broadband pulsed emission of radio waves by which pulsars are discovered is in fact a dramatic diagnostic of much more powerful processes of energy and angular momentum loss. This is itself a justification for study of the structure of the pulsar magnetosphere, a well defined but formidable

problem in relativistic electrodynamics and plasma physics: how does a rotating magnetic neutron star come to terms with its environment, within the framework of either classical or quantum physics? But one should not then infer that the weak, localized radio emission cannot supply at least some hints about the global magnetospheric structure (e.g., Wright, 2003; Rankin and Wright, 2003).

2. Basic electrodynamics

We use the cylindrical polar coordinate system (ϖ, ϕ, z) based on the rotation axis \mathbf{k} . In this essentially relativistic problem, a crucial role is played by the *light-cylinder* (l-c) of radius $\varpi_c = c/\alpha$ on which a point corotating with the star has the velocity c . Field lines closing well within the l-c form a *dead zone*, with no outflowing currents. In a “live magnetospheric model,” braking of the stellar rotation occurs through transport of angular momentum by currents flowing along the *open* field lines that emanate from the polar caps to cross the l-c.

In general, an axisymmetric magnetic field is written as the superposition of a poloidal (meridional) component \mathbf{B}_p and a toroidal (azimuthal) component \mathbf{B}_t . The component \mathbf{B}_p is conveniently described by the flux function $P(\varpi, z)$:

$$\mathbf{B}_p = -\nabla P \times \frac{\mathbf{t}}{\varpi} = \frac{1}{\varpi} \left(\frac{\partial P}{\partial z}, 0, -\frac{\partial P}{\partial \varpi} \right), \quad (2)$$

with \mathbf{t} the unit toroidal vector. For definiteness and simplicity, the magnetic flux distribution over the star’s surface is assumed to be that of a dipole located at the star’s center. The function P will then be everywhere symmetric with respect to the equator, and it will suffice to study just the hemisphere $z \geq 0$. Near the star the poloidal field is effectively curl-free, with

$$P = -\frac{B_s R^3}{2} \frac{\varpi^2}{(\varpi^2 + z^2)^{3/2}}, \quad (3)$$

where R is the stellar radius and B_s the polar field strength.

By Ampère’s law, in general \mathbf{B}_p is maintained by the toroidal current density

$$\mathbf{j}_t = \frac{c}{4\pi} \nabla \times \mathbf{B}_p = \frac{c}{4\pi} (\nabla \times \mathbf{B})_\phi \mathbf{t} \quad (4)$$

$$= \frac{c}{4\pi} \left(\frac{1}{\varpi} \nabla^2 P - \frac{2}{\varpi^2} \frac{\partial P}{\partial \varpi} \right) \mathbf{t}. \quad (5)$$

In the wind zone of an active magnetosphere the toroidal component \mathbf{B}_t is conveniently written

$$\mathbf{B}_t = B_\phi \mathbf{t} = -\frac{4\pi}{c} \frac{S}{\varpi} \mathbf{t}, \quad (6)$$

where S clearly must vanish on the axis. Again by Ampère's law, the field (6) is maintained by the poloidal current density

$$\mathbf{j}_p = \frac{c}{4\pi} \nabla \times \mathbf{B}_t = -\frac{1}{\varpi} \nabla S \times \mathbf{t}. \quad (7)$$

Thus $S(\varpi, z)$ is a Stokes stream function, constant on the poloidal current lines, with $-2\pi S$ measuring the total outflow of charge between the axis and the current line S . In a steady state the total outflow of charge must be zero, so there must be a current-closing streamline on which S vanishes.

In the stellar crust, assumed a classical perfect conductor, the electric field is given by the "corotational field"

$$\mathbf{E} = -\frac{\alpha}{c} (\mathbf{k} \times \mathbf{r}) \times \mathbf{B} = -\frac{\alpha}{c} \varpi \mathbf{t} \times \mathbf{B}_p = \frac{\alpha}{c} \nabla P = \frac{\alpha \varpi}{c} (-B_z, 0, B_x). \quad (8)$$

In the surrounding magnetosphere, the time-independent and therefore curl-free vector \mathbf{E} is conveniently written

$$\mathbf{E} = -\frac{\alpha}{c} (\mathbf{k} \times \mathbf{r}) \times \mathbf{B}_p - \nabla \psi = \frac{\alpha}{c} \nabla P - \nabla \psi, \quad (9)$$

i.e., the sum of the field (8) and the "non-corotational" field $-\nabla \psi$, which includes any electric component parallel to \mathbf{B} . (This break-up is valid also for a "quasi-steady" model, e.g., the oblique rotator, for which changes in time at a point fixed in the inertial frame are due to the rigid rotation $\alpha \mathbf{k}$ of a non-axisymmetric structure.) Equivalently, the electric potential ϕ is written

$$\phi = -\alpha P / c + \psi. \quad (10)$$

From the Poisson-Maxwell equation, the charge-density ρ_e required to maintain the electric field (9) is

$$\begin{aligned} \rho_e &= \frac{\nabla \cdot \mathbf{E}}{4\pi} = -\frac{\alpha}{2\pi c} \mathbf{k} \cdot \left[\mathbf{B} - \frac{1}{2} \mathbf{r} \times (\nabla \times \mathbf{B}) \right] - \frac{1}{4\pi} \nabla^2 \psi \\ &= -\frac{\alpha}{2\pi c} \left[B_z - \frac{1}{2} \varpi (\nabla \times \mathbf{B})_\phi \right] - \frac{1}{4\pi} \nabla^2 \psi \\ &= \frac{1}{4\pi} \left(\frac{\alpha}{c} \nabla^2 P - \nabla^2 \psi \right). \end{aligned} \quad (11)$$

The function ψ is determined by the physics of the magnetosphere. Inside the star, by (8), ψ is a constant. The earliest studies argued that the smallness of the thermal scale-height due to the enormous gravitational field implied that both the mass-density and so also the net charge-density would exponentiate down to values so small that the magnetosphere would be effectively an electrodynamic vacuum, with both ρ_e and \mathbf{j} zero. By (11) and (5), ψ would satisfy $\nabla^2\psi = (\alpha/c)\nabla^2P = (2\alpha/c)(P_\varpi/\varpi)$, with P given by the curl-free form (3). Solution subject to the boundary conditions that ψ be continuous at the star's surface and finite at infinity would then yield a normal discontinuity in $\nabla\psi$, maintained by a surface charge density.

In a classical paper, Goldreich and Julian (1969) (“GJ”) pointed out that with the canonical estimate of $\approx 10^{12}$ G for the pulsar field, the electric force on these surface charges would be far larger than the restraining gravitational forces. The “open circuit” vacuum model will in fact lead to spontaneous charging up of the magnetosphere by current flow out of the star. GJ suggested that the zero charge density condition should be replaced by the simple plasma condition $\mathbf{E} \cdot \mathbf{B} = 0$, whence from (9), $\mathbf{B} \cdot \nabla\psi = 0$: the uniform value for ψ inside the star would be propagated into the magnetosphere. Under this “GJ-hypothesis,” the magnetospheric \mathbf{E} would also satisfy the corotation relation (8). By (11), the charge-density is then given by

$$(\rho_e)_{\text{GJ}} = -\frac{\alpha}{4\pi c} [2B_z - \varpi(\nabla \times \mathbf{B})_\phi] \quad (12)$$

(symmetric in the equator).

In magnitude, the macroscopic electric force density $|\rho_e E|$ acting on the GJ charges will typically be of order $(\alpha\varpi/c)^2(B^2/4\pi\varpi) \simeq (\alpha\varpi/c)^2$ (magnetic force density). In normal MHD, one is not only allowed but one is almost morally obliged to neglect the electric force; in Schlüter’s words, it makes no sense to retain terms of order $(v/c)^2$ when one has put all the relativistic γ terms equal to unity. But in this essentially relativistic problem, we may anticipate that both parts of the electromagnetic stress tensor should be retained, whether or not the GJ approximation is locally valid.

Much of the “pulsar magnetosphere problem” is concerned with the determination of the function ψ and the associated charge-current field in different domains. It is however convenient and instructive to begin by following the spirit of GJ, in particular by assuming initially purely classical physics.

We now define dimensionless coordinates (x, \bar{z}) , and normalize $P, S, \mathbf{E}, \mathbf{B}$ in terms of a standard light-cylinder field strength $B_{lc} = (B_s/2)(\alpha R/c)^3$:

$$(x, \bar{z}) = (\alpha/c)(\varpi, z), \quad P = \bar{P} B_{lc}(c/\alpha)^2, \\ S = \bar{S} B_{lc}(c^2/4\pi\alpha), \quad (\phi, \psi) = B_{lc}(c/\alpha)(\bar{\phi}, \bar{\psi}), \quad (\mathbf{B}, \mathbf{E}) = B_{lc}(\bar{\mathbf{B}}, \bar{\mathbf{E}}). \quad (13)$$

(Once defined, the dimensionless quantities are again immediately written without the bars.) The normalized fields have the form

$$\mathbf{B}_p = -\nabla P \times \mathbf{t}/x, \quad B_\phi = -S/x, \quad \phi = -P + \psi, \quad \mathbf{E} = \nabla P - \nabla \psi. \quad (14)$$

As $(\varpi, z) \rightarrow 0$, P must reduce to

$$-x^2/(x^2 + z^2)^{3/2}, \quad (15)$$

the normalized form of (3).

2.1. An illustrative example

An immediate consequence of the change from a vacuum to a charged magnetosphere is on the detailed structure of the poloidal field. To illustrate this, consider the non-active case with $S = 0$ – no poloidal currents maintaining a toroidal field – but with the domain within the l-c filled with the GJ charge density supposed corotating with the star. The solution has been discussed by Michel (1973, 1991), and later by Mestel and Wang (1979) and Mestel and Pryce (1992), and is referred to as the MMWP field. From (4), (5) and (11), Ampère's equation

$$(\nabla \times \mathbf{B})_\phi = 4\pi\rho_e\alpha\varpi/c \quad (16)$$

yields

$$(1 - x^2)\nabla^2 P = \frac{2}{x} \frac{\partial P}{\partial x} \quad (17)$$

for the flux function P , equivalent to

$$\frac{\partial^2 P}{\partial x^2} + \frac{\partial^2 P}{\partial z^2} - \frac{1}{x} \frac{1+x^2}{1-x^2} \frac{\partial P}{\partial x} = 0. \quad (18)$$

The rotating charges manifest themselves through the terms in $x^2 \equiv (\alpha\varpi/c)^2$. Well within the l-c these terms are small, and (18) reduces to the equation satisfied by P yielding curl-free fields, such as (15).

The most convenient method of solution writes P as a Fourier cosine integral

$$P = \int_0^\infty f(x, k) \cos kz dk, \quad f = \frac{2}{\pi} \int_0^\infty P(x, z) \cos kz dz. \quad (19)$$

This ensures that the field lines forming a closed domain within the l-c satisfy the required equatorial boundary condition

$$B_x(x, 0) = 0 \quad (20)$$

with no equatorial current sheet. The transform function f satisfies

$$\frac{d^2 f}{dx^2} - \frac{1}{x} \frac{1+x^2}{1-x^2} \frac{df}{dx} - k^2 f = 0. \quad (21)$$

If neither the field nor the volume current density are to be singular at the l-c, (18) requires that at $x = 1$, $B_z = -\partial P / \partial x = 0$: the field lines reach the l-c normally, and there is a neutral point at the intersection (1, 0) of the l-c and the equator. As $x \rightarrow 0$, f must reduce to the Fourier transform of (15)

$$-\frac{2}{\pi} kx K_1(kx) \quad (22)$$

where K_1 is the modified Bessel function of the second kind. For each value of k , the solution near $x = 1$ is begun as a series expansion with $df/dx(1, k) = 0$ and continued inward numerically, to link up smoothly near $x = 0$ with a local extension of (22). Typical lines of the resulting MMWP field given by the numerical construction of (19) are shown in Fig. 2.

The solution is clearly incomplete unless continued over all space. For example, if one postulates that the domain beyond the l-c is pure vacuum, then a solution well behaved at infinity and continuous in B_x and E_z has to be maintained by surface charges and currents on the l-c, and is again subject to unbalanced stresses. The conditions for viable inactive axisymmetric models, such as large vacuum gaps within the l-c, are discussed in Smith *et al.* (2001). Some features of the MMWP field should persist in a realistic active model, such as the large dead zone, and the marked deviation from the vacuum dipolar field caused by the rotating charges as the l-c is approached.

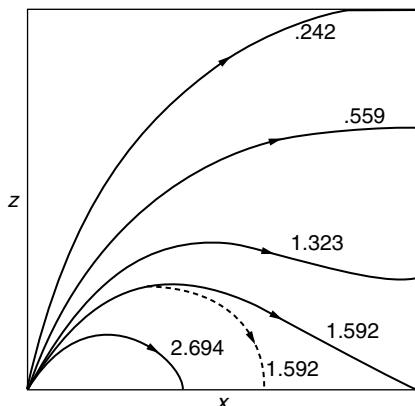


Figure 2. The field as modified by the corotating GJ charge density within the l-c. The numbers refer to the value of P . The dashed line belongs to the vacuum extension of the star's dipolar field.

2.2. A live classical model

In a series of studies aimed at the construction of a live classical model (Mestel *et al.*, 1985 – MRWW; Fitzpatrick and Mestel, 1988a,b – FM; Mestel, 1999, Chapter 13), the “GJ approximation” is supposed adequate out to and slightly beyond the l-c. Further, in the dead zone and in much of the active zone, the gas is “charge-separated”: the charge density ρ_e , given by (12), is equal to $-n_e e$ in a negatively charged domain, and $n_i Z e$ in a positively charged domain consisting of ions with charge $Z e$. This is in contrast to a normal plasma, in which ρ_e is the difference (usually small) between the charge densities of a mixture of ions and electrons.

When the magnetic axis is parallel rather than anti-parallel to the rotation axis \mathbf{k} – the aligned rather than the counter-aligned case – a current of initially sub-relativistic, negatively charged electrons leaves the star at high latitudes. Within the l-c the poloidal velocity \mathbf{v}_p is nearly but not quite parallel to \mathbf{B}_p , and the angular velocity Ω initially close to but always less than α . The outflowing gas acquires angular momentum $\gamma m \Omega \varpi^2$ from the non-zero magnetic torque. The associated “inertial drift” takes the negatively charged electrons to neighboring field lines, at a higher corotational potential $-\alpha P/c$, so the relativistic kinetic energy γmc^2 increases through the work done by the field

(8) – the “centrifugal slingshot effect.” Within the l-c, in spite of the steadily increasing γ -values, the trans-field inertial drifts remain small enough for MHD-type flow (corotation plus flow parallel to the total field $\mathbf{B}_p + \mathbf{B}_t$ – familiar from studies of magnetic stellar winds) to remain an excellent approximation. Further, in a cold, charge-separated gas, the bulk equation of motion of the gas is just the equation of motion of a single electron, multiplied by the local electron density: a “perfectly conducting” gas, in which inertial force and radiative damping are negligible, is also one which satisfies the “force-free” condition, in which the net bulk electromagnetic force density is zero, without any correction from non-electromagnetic forces.

However, beyond the l-c the geometry of the joint poloidal-toroidal magnetic field forces γ up until the electrons, moving along their curved trajectories, emit incoherent radiation at the rate $\mathcal{P} = (2e^2/3c^3)\gamma^4 [\mathbf{v} \times (\nabla \times \mathbf{v})]^2$. Because of the strong relativistic beaming in the direction of the electron motion, each electron feels the dynamically significant drag $-(\mathcal{P}/c^2)\mathbf{v}$. In particular, the radiation of angular momentum yields a “dissipative drift” equatorward across the poloidal field, analogous to the inertial drift, with a similar gain in kinetic energy. In order that this energy gain exceeds the radiation loss, the dissipation must occur beyond the l-c. It is found that the trans-field drift is small — so that the flow is still nearly parallel to \mathbf{B}_p — until near the equator, where B_ϕ vanishes and so $E > B$. The electron stream is then driven across \mathbf{B}_p toward the star by the dominant electric field (Fig. 3).

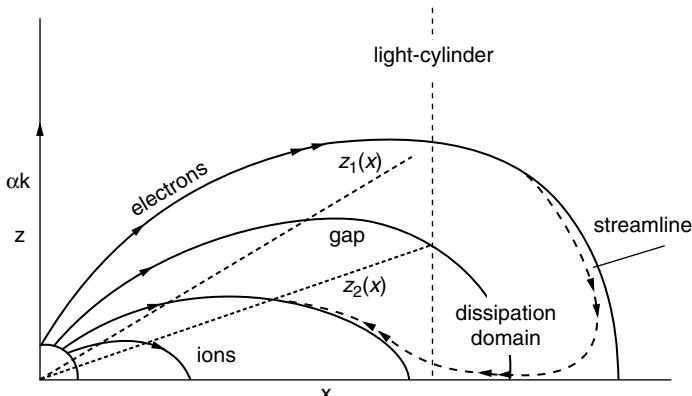


Figure 3. Schematic diagram of the live classical model.

Returning to first principles: the energy source that drives current through the magnetosphere is the potential variation $-\alpha P/c$ on the stellar surface due to the rotation of the magnetic star. In order to pick up this energy, particles must be able to cross field lines, achieved in this highly idealized model via the dissipative drift. In this way, radiation of angular momentum — acquired from the star via magnetic torques — enables the electrons to pick up electrical energy supplied from the rotational energy of the star. Some of this energy is radiated along with the angular momentum, and the rest is retained by the returning electrons. As emphasized by Gold (private communication) and Holloway (1977), there is a global constraint on any model for the spin-down of a rigidly rotating star: the time-derivatives of the kinetic energy $T = I\alpha^2/2$ and angular momentum $H = I\alpha$ must be related by $-\dot{T} = \alpha(-\dot{H})$, requiring that much of the radiation loss must occur beyond the l-c. In the above model, there is a “Gold-Holloway condition” that is automatically satisfied for each electron-current trajectory.

The potential difference across a polar cap — defined by the last field line to reach the l-c — is $\Delta\phi \simeq \alpha^2 R^3 B_s/2c^2$. The maximum γ -value attainable by a circulating electron is then $\gamma_m \simeq e\Delta\phi/mc^2 \simeq [eB_s(\alpha R/c)^3/2mc\alpha] \simeq 1.25 \times 10^7 (B_s/10^{12})(R/10^6)^3 P_1^{-2}$. The value of γ required to yield significant trans-field motion is $\gamma_d \simeq 1.4 \times 10^7 [(B_s/10^{12})(R/10^6)^3/P_1]^{1/4}$, comparable with γ_m for normal pulsar periods $P_1 \simeq 1$, and markedly less for the Crab pulsar. By the standard synchrotron-type argument, the emitted photons would be gamma-rays with typical frequency $\nu \simeq \gamma_d^3 \alpha$ and energy $\simeq 5 \times 10^7 / P_1^{7/4}$ eV.

A feature of this model is the presence of a near-vacuum gap within the dead zone, centered on the curve at which the GJ charge density (12) changes from being negative at higher latitudes to positive at lower. The possibility and indeed the likelihood of such a gap was noted first by Holloway (1973) and discussed further by Holloway and Pryce (1981). Well within the l-c, where the departure from a curl-free structure is small, $(\rho_e)_{GJ}$ and B_z vanish together, along the line $\theta = \arccos(1/3^{1/2}) \simeq 55^\circ$. The model gap is wedge-shaped, with walls defined by angles $\theta_1 < 55^\circ < \theta_2$. Within the gap, $\mathbf{E} \cdot \mathbf{B} \neq 0$, and with the sign so as to repel a negative particle approaching from the northerly, circumpolar domain, and similarly a positive particle approaching from the southerly, equatorial domain. However, an electron which has crossed the l-c, performed the dissipative equatorward drift and is now flowing back to the star, as described above, will cross the gap from south to north, and so will feel a further strong acceleration. A fully self-consistent calculation would need to take account of the returning electrons as additional sources of the electric field.

3. Models with pair production

3.1. The electron-positron plasma

The virtue of the tentative classical model of Section 2.2 is its attempt to link up the basic problem of the return current with the spontaneous radiation of energy and angular momentum. As seen, the potential difference available across the polar cap is so large that one might indeed expect the star to try to turn into a machine for generating gamma-rays. The obvious drawback is that it tries to prove too much: even those pulsars which do emit copious gamma rays still manage to convert most of their spin-down energy into other forms. But if relativistic acceleration occurs instead near the star, then as pointed out originally by Goldreich in a conference report and by Sturrock (1971), emitted gamma-rays moving across the very strong magnetic field convert spontaneously into electron-positron (\pm) pairs, which can lead to a pair production cascade. Much of the spin-down energy could then be lost as a relativistic e^\pm -wind.

As an example, consider the above model with an aligned dipolar magnetic field. The velocity $v_0(P)$ with which the electrons leave the star along the field-streamline P – typically $c/2$ – is fixed by the requirement that the GJ approximation remain valid, so that the electron stream can pass through the l-c with γ remaining moderate. However, if v_0 is markedly higher, then at some point close to the star there will be a breakdown in the GJ approximation: from the conservation of charge and of magnetic flux along the field-streamlines, curved away from the rotation axis, the Poisson-Maxwell equation will require a steadily increasing component of \mathbf{E} parallel to \mathbf{B}_p , which will enforce a rapid acceleration to high γ -values. Because of the same magnetic field curvature, the strongly relativistic particles will emit curvature radiation, again in the form of gamma-rays, followed by pair production.

This generation of a mixed plasma from high- γ primary electrons is not restricted to axisymmetric geometry. Arons and colleagues considered the oblique rotator and have argued rather for acceleration along those field lines that curve toward the rotation axis. They concentrated on the special “self-screening” case: near the star there is a slightly super-GJ charge density with its associated accelerating E_{\parallel} , but with $E_{\parallel} \rightarrow 0$ far from the star (Scharlemann *et al.*, 1978). A potentially important modification is the general relativistic “dragging of inertial frames,” which yields a significant modification to $(\rho_e)_{GJ}$, with a consequent increase in E_{\parallel} in the acceleration domain near the star (Muslimov and Tsygan, 1992).

The basic picture has been applied to different models and discussed in many papers, reviews and books (e.g., Arons, 1981, 1983, 1992; Arons and Scharlemann, 1979; Beskin *et al.*, 1993; Daugherty and Harding, 1982; Fawley *et al.*, 1977; Jones, 1980; Mestel, 1999; Mestel and Shibata (MS), 1994; Scharlemann *et al.*, 1978; Shibata, 1991, 1997; Shibata *et al.*, 1998, 2002; Usov, 1996; and many others.)

In an early, highly influential paper, Ruderman and Sutherland (1975) concentrated on the “anti-pulsar,” with the magnetic and rotation axes counter-aligned. They argued that an estimated large work function would inhibit the emission of ions required to build up the *positive* GJ charge-density (11), but the consequent strong vacuum electric field would again initiate local pair production. Many of the ideas in this paper have persisted, even though the newer calculations have much reduced the ionic work function.

More recent studies have focused on the generation of γ -rays through inverse Compton scattering of thermal photons from the pulsar surface, itself heated by returning positrons (Harding and Muslimov, 2001, 2002; Harding *et al.*, 2002; Hirschman and Arons, 2001; Hirschman, 2002).

There is not yet a consensus on the very important details, especially on the process by which a small fraction of the spin-down energy supplied to the plasma is emitted as coherent radio emission, nor on the location of the sources of the similarly pulsed emission in optical, ultra-violet, X-ray and gamma-ray frequencies, observed in some pulsars. (For a recent review, see Graham-Smith, 2003.) However, for study of the global magnetospheric structure, it is eminently plausible that in a realistic model, the active magnetospheric domain is well approximated by an electron-positron gas, described by the appropriate modification of MHD. It can be argued that even in the live classical model of Section 2.2, which assumes quasi-GJ flow without strong acceleration until the outflowing stream has crossed the l-c, the high- γ electrons returning to the strong field neighborhood of the star will inevitably generate pairs, so that the classical model would spontaneously transform itself into a quantum model.

A systematic treatment of the dynamics of a “cold,” dissipation-free electron-positron gas has been given by Melatos and Melrose (1996) (see also Blackman and Field, 1993). The significant differences from a normal plasma are due in part to the particles being relativistic, and also through both species having the same rest-mass, instead of differing by a factor $A m_H/m_e$ in standard notation. The two-fluid, non-dissipative equations, written in terms of “lab-frame” number densities n^\pm and velocities \mathbf{v}^\pm , and Lorentz factors $\gamma^\pm = [1 - (v^\pm/c)^2]^{-1/2}$, are reduced to

one-fluid equations in terms of

$$n = (n^+ + n^-), \quad \mathbf{U} = \frac{n^+ \mathbf{v}^+ + n^- \mathbf{v}^-}{n^+ + n^-}, \quad \Gamma = (1 - U^2/c^2)^{-1/2}$$

$$\rho_e = e(n^+ - n^-), \quad \mathbf{j} = e(n^+ \mathbf{v}^+ - n^- \mathbf{v}^-). \quad (23)$$

The particle continuity and current continuity equations retain their standard forms

$$\frac{\partial n}{\partial t} + \nabla \cdot (n \mathbf{U}) = 0; \quad \frac{\partial \rho_e}{\partial t} + \nabla \cdot \mathbf{j} = 0. \quad (24)$$

The equation of bulk motion with velocity \mathbf{U} reduces to

$$mn \frac{\partial}{\partial t} (\Gamma \mathbf{U}) + mn \mathbf{U} \cdot \nabla (\Gamma \mathbf{U}) = \frac{(\mathbf{j} - \rho_e \mathbf{U}) \times \mathbf{B}}{c} \quad (25)$$

provided $j \ll neU$, even if, as expected in a relativistic problem, the “convection current” $\rho_e U$ due to bulk motion of the net charge density is not small compared with the current due to relative motion of the two species. The generalized Ohm’s law reduces to

$$\frac{m}{e^2} \frac{\partial}{\partial t} \left(\frac{\Gamma \mathbf{j}}{n} \right) + \frac{m}{e^2} \mathbf{U} \cdot \nabla \left(\frac{\Gamma \mathbf{j}}{n} \right) + \frac{m}{ne^2} (\mathbf{j} - \rho_e \mathbf{U}) \cdot \nabla (\Gamma \mathbf{U}) = \mathbf{E} + \frac{\mathbf{U} \times \mathbf{B}}{c}. \quad (26)$$

We are interested primarily in steady states that are force-free over the bulk of the domain. In contrast to the charge-separated, classical model, in an $\pm e$ -plasma the total particle number density will normally greatly exceed the GJ number density, so the “perfect conductivity” and the “force-free” conditions are distinct: a gas satisfying the “perfect conductivity” condition need not be force-free. However, one can easily check that if inertial terms on the left of (25) are small compared with either term on the right, then the inertial corrections in (26) are *a fortiori* small. Thus in the absence of collisions, radiation damping or electron-positron annihilation, the force-free condition implies the ideal MHD condition $\mathbf{E} + \mathbf{U} \times \mathbf{B}/c = \mathbf{0}$, which in turn shows that the Lorentz force term on the right of (25) is $\rho_e \mathbf{E} + \mathbf{j} \times \mathbf{B}/c$.

It should be noted that in the more general, non-axisymmetric case, with \mathbf{k} and \mathbf{p} nonaligned, the time-dependent terms will not be negligible everywhere. As discussed by Melatos and Melrose (1996), at a finite distance far beyond the l-c, the displacement current will dominate over the particle current, and the MHD approximations must break down.

3.2. The relativistic force-free equation

With realistic pair production efficiency near the star, the electromagnetic energy density will still be much larger than the kinetic energy density except for very high γ -values, so it is certainly a reasonable working hypothesis to adopt both the force-free and the dissipation-free approximations as valid over much of the magnetosphere. The analysis should show in which (probably localized) domains there is a breakdown, with possible interesting observable consequences. A recent paper that has been a stimulus to much of the following discussion is by Contopoulos *et al.* (1999) – CKF (see also Mestel, 2001).

As a further approximation, we limit discussion here to the case in which the simple plasma condition $\mathbf{E} \cdot \mathbf{B} = 0$ is supposed to hold throughout the magnetosphere, both within and beyond the l-c. Then all field lines corotate with the star, and

$$\mathbf{E} = -\frac{\alpha\varpi}{c}\mathbf{t} \times \mathbf{B} = \frac{\alpha\varpi}{c}(-B_z, 0, B_x) = \frac{\alpha}{c}\nabla P. \quad (27)$$

As noted in Section 3.1, in models in which pair production is presumed to occur in the open field line domain near the star, there has to be a locally non-trivial component of \mathbf{E} along \mathbf{B}_p in the acceleration domain in which the primary electrons reach γ -values high enough for pair production to occur. In a hypothetical steady state, the simple plasma condition $\mathbf{E} \cdot \mathbf{B} = 0$ will again be set up beyond the acceleration domain in the electron-positron plasma, but now with $\mathbf{E} = -\tilde{\alpha}(P)(\varpi/c)\mathbf{t} \times \mathbf{B}$ – i.e., with the field lines in the wind zone having individual rotation rates $\tilde{\alpha}(P)$ that differ somewhat from the rotation α of the star (cf., MS, Section 4). However, at least for the more rapid rotators this effect will be fairly small, and so is left for inclusion in a subsequent generalization of the present treatment.

This local acceleration occurs through the density of the primary electrons being forced up to exceed the GJ value. Recall that there can exist also domains in which the GJ hypothesis breaks down in the opposite sense – *vacuum gaps*. A feature of the original GJ picture that survives in much of the later discussion is the charge-separated dead zone, bounded by the last field line to close within the l-c, and with local density given by (12). As noted in subsection 2.2, some models are forced to include a vacuum gap, separating negatively and positively charged domains, leading necessarily to a departure from universal corotation. Recently, there have been many studies of outer magnetospheric gaps (cf., Section 9). Again, these potentially important modifications are for the moment put on one side.

Since in the aligned case, the primary outflowing particles are negatively charged electrons, the stream function S defined by (7) begins by increasing from zero on the axis, so that B_ϕ is negative – the field lines are twisted backward with respect to the axis \mathbf{k} . As the electric force density

$$\rho_e \mathbf{E} = -\rho_e \frac{\alpha \varpi}{c} \mathbf{t} \times \mathbf{B}_p \quad (28)$$

is purely poloidal, in a force-free magnetosphere the toroidal component of the magnetic force density $\mathbf{j}_p \times \mathbf{B}_p/c$ vanishes (the “torque-free” condition), so \mathbf{j}_p must be parallel to \mathbf{B}_p , yielding from (2) and (7) the functional relation

$$S = S(P), \quad \mathbf{j}_p = \frac{dS}{dP} \mathbf{B}_p: \quad (29)$$

the poloidal current streamlines are identical with the poloidal field lines. The respective contributions of \mathbf{B}_p and \mathbf{B}_t to the poloidal force density are

$$\frac{1}{c} \mathbf{j}_t \times \mathbf{B}_p = \frac{1}{4\pi} (\nabla \times \mathbf{B})_\phi \mathbf{t} \times \mathbf{B}_p \quad (30)$$

and

$$\frac{1}{c} \mathbf{j}_p \times \mathbf{B}_t = \frac{4\pi S}{c^2 \varpi} \frac{dS}{dP} \mathbf{t} \times \mathbf{B}_p = -\frac{4\pi}{c^2 \varpi^2} S \frac{dS}{dP} \nabla P, \quad (31)$$

on use of (6), (7) and (2). By (12), (28), (30) and (31), the poloidal component of the force-free equation

$$\rho_e \mathbf{E} + \frac{1}{c} \mathbf{j}_t \times \mathbf{B}_p + \frac{1}{c} \mathbf{j}_p \times \mathbf{B}_t = \mathbf{0} \quad (32)$$

reduces to

$$\frac{1}{\varpi} (\nabla \times \mathbf{B})_\phi \left[1 - \left(\frac{\alpha \varpi}{c} \right)^2 \right] + \left(\frac{\alpha \varpi}{c} \right)^2 \frac{2B_z}{\varpi^2} + \left(\frac{4\pi}{c \varpi} \right)^2 S \frac{dS}{dP} = 0, \quad (33)$$

with $(\nabla \times \mathbf{B})_\phi$ given by (4) and (5). Note that the first term in (33) combines part of the electric force with the force due to \mathbf{B}_p ; the second again comes from the electric force, while the third is due to \mathbf{B}_t .

In normalized form, defined by (13) and (14), the force-free equation (33) becomes

$$(x^2 - 1) \frac{\partial^2 P}{\partial x^2} + \frac{(1+x^2)}{x} \frac{\partial P}{\partial x} + (x^2 - 1) \frac{\partial^2 P}{\partial z^2} = S \frac{dS}{dP}. \quad (34)$$

Note that the light-cylinder $x = 1$ is a singularity of this differential equation. When $x \ll 1$, $|\rho_e \mathbf{E}| \ll |\mathbf{j} \times \mathbf{B}|/c$, and (34) becomes the reduced form of the non-relativistic force-free condition $\mathbf{j} \times \mathbf{B} = \mathbf{0}$.

An essential part of the whole problem is the search for appropriate forms for the function $S(P)$, which should emerge from the construction of mutually consistent fields for the inner and outer domains, respectively, within and without the light-cylinder, and behaving properly near the star and at infinity. With the sign convention in (2), P decreases from zero on the axis, so that initially dS/dP is negative, and the force density (31) due to the toroidal field \mathbf{B}_t acts toward the axis. However, since S vanishes on the current closing streamline, there must be an intermediate field line on which dS/dP changes sign; below this field line the force (31) acts toward the equator. Closure via a sheet current is not excluded.

4. Live models

4.1. The outer domain

As noted, a live magnetospheric model, suffering systematic spin-down, must have $S \neq 0$. Let us suppose provisionally that everywhere beyond the l-c the γ -values remain small enough for trans-field motion of the plasma to be negligible – i.e., the simple “perfect conductivity” condition (27) continues to hold. The net current \mathbf{j}_p is then nearly parallel to \mathbf{B}_p . Then the field beyond the l-c must be topologically “open,” with no field lines crossing the equator. As noted, as in the non-relativistic wind problem, we expect a live model to have a corotating dead zone, with field lines crossing the equator normally but the zone must be confined within the l-c, as in the illustrative MMWP field just discussed. Beyond the l-c, the boundary condition $B_z = 0$ is imposed at the equator, with $B_x(x, 0)$ positive in the northern hemisphere and negative in the southern, and the associated toroidal components B_ϕ respectively negative and positive, implying a locally positive current density j_ϕ and so a magnetic force density that acts toward the equator. Such “pinching” magnetic fields are familiar from flare theory (e.g., Priest and Forbes, 2000). However, in a relativistic problem, again the electric field contribution to the macroscopic dynamics cannot be ignored. By (8), in the northern hemisphere, near the equator the electric field $x B_x$ is in the positive z -direction. Thus the poloidal magnetic field pinches, as in familiar non-relativistic problems, but the combined electric force (28) and the \mathbf{B}_p force (30) is $[(x^2 - 1)(\nabla \times \mathbf{B})_\phi/4\pi]\mathbf{k}$ and so acts away

from the equator, since $x > 1$. The proposed force-free equilibrium near the equator must therefore be maintained by the pinching effect of \mathbf{B}_t . We have seen above that $S dS/dP$ does indeed become positive at low latitudes, so that (31) has the required direction.

Consider first the idealized case, with B_x , B_ϕ and E_z abruptly reversing sign at the equator, and so with positive sheet currents J_ϕ , J_x and positive surface charges σ . Over the surface $z = +\varepsilon$, there is a net electromagnetic stress $T_{ij}k_j$ where

$$T_{ij} = \frac{1}{4\pi} \left[-\frac{1}{2} \left(|\mathbf{E}|^2 + |\mathbf{B}|^2 \right) \delta_{ij} + (E_i E_j + B_i B_j) \right] \quad (35)$$

is the Maxwell stress tensor. The electric terms yield $(x^2 B_x^2 / 8\pi) \mathbf{k}$ and the magnetic terms $-[(B_x^2 + B_\phi^2) / 8\pi] \mathbf{k}$, so the net electromagnetic pressure, acting toward the equator, is

$$\frac{1}{8\pi} \left[B_\phi^2 - (x^2 - 1) B_x^2 \right]. \quad (36)$$

There is an equal electromagnetic pressure, also acting toward the equator on the surface $z = -\varepsilon$. Provided

$$|B_\phi/B_x| > \sqrt{x^2 - 1}, \quad (37)$$

then an equal particle pressure at the equator is both necessary and sufficient to maintain equilibrium.

The force-free equation (33) can be written succinctly as $\partial T_{ij}/\partial x_j = 0$. Near the equator, $|B_z| \ll |B_x|$ in this model, and the force-free condition becomes

$$\frac{\partial}{\partial x_3} T_{33} = \frac{1}{8\pi} \frac{\partial}{\partial z} \left[-B_\phi^2 + (x^2 - 1) B_x^2 \right] = 0, \quad (38)$$

whence

$$\frac{1}{8\pi} \left[B_\phi^2 - (x^2 - 1) B_x^2 \right] = p_{\text{eq}}, \quad (39)$$

where p_{eq} is independent of z . If (37) holds, then p_{eq} is identified as a particle pressure at the equator, where B_x and B_ϕ change sign.

More realistically, the transition to zero field at the equator will be continuous, with the particle pressure p increasing steadily across half of the thin equatorial sheet. The force-free condition (38) is replaced by

$$\frac{\partial}{\partial z} \left\{ \frac{1}{8\pi} \left[-B_\phi^2 + (x^2 - 1) B_x^2 \right] - p \right\} = 0 \quad (40)$$

and (39) by

$$\frac{1}{8\pi} \left[B_\phi^2 - (x^2 - 1)B_x^2 \right] + p = p_{\text{eq}}. \quad (41)$$

As the equator is approached, p steadily increases from the uniform value (implicit in the force-free assumption), which could in principle be zero, to the value p_{eq} . *Thus the model necessarily includes a thin domain with a non-force-free electromagnetic field: the system appears to build up spontaneously its own boundary layer, in which a non-electromagnetic force is locally essential.*

The limiting case, with the inequality sign in (37) replaced by equality, appears consistent with zero equatorial pressure p_{eq} . However, condition (33) assumes that all non-electromagnetic forces, including inertial forces, are small compared with the dominant terms in the Lorentz force. With the electromagnetic field subject to the condition (8), the kinematics of the flow has the form familiar from stellar wind theory, with \mathbf{v}_p parallel to \mathbf{B}_p , and the rotation velocity given by

$$v_\phi = c \left(x + \frac{v_p}{c} \frac{B_\phi}{B_p} \right). \quad (42)$$

The energy integral in a pressure-free system is (MRWW; MS; Contopoulos, 1995)

$$\gamma [1 - x(v_\phi/c)] = \text{constant}, \quad (43)$$

showing that γ would become infinite if $v_\phi = c/x$, $v_p = c\sqrt{(x^2 - 1)}/x$; and from (42) this is equivalent to

$$|B_\phi/B_p| = \sqrt{x^2 - 1}. \quad (44)$$

Thus the seemingly exceptional case, with zero equatorial pressure and so with the force-free equation holding all the way to the equator, in fact requires that the flow outside the equatorial zone be so highly relativistic that the neglected inertial terms are not small, so violating an essential condition for the force-free approximation to hold. We conclude that in all cases, the equilibrium conditions for this model will require at least a local breakdown in force-free conditions. In analogous non-relativistic problems, Lynden-Bell (1996) has pointed out that a thermal pressure is again required to balance magnetic pinching forces, exerted locally by an otherwise force-free field. However we again emphasize the important difference, that whereas in a non-relativistic problem, the electric stresses are normally smaller than the magnetic by the factor $(v/c)^2$, in the present problem, the opposing electric stresses exceed

the pinching poloidal field stresses by the factor x^2 , and equilibrium is possible only through the pressure exerted by the toroidal field.

In the domain near the equator, with $|B_x/B_z| \gg 1$, the relativistic flow likewise has $|v_x/v_z| \gg 1$, so that $v_x^2 + v_\phi^2 \approx c^2$, whence from (42), (remembering that B_ϕ is negative),

$$\frac{v_\phi}{c} = \frac{x - b\sqrt{(1 + b^2 - x^2)}}{(1 + b^2)}, \quad \frac{v_x}{c} = \frac{bx + \sqrt{(1 + b^2 - x^2)}}{(1 + b^2)}, \quad (45)$$

where $b \equiv |B_\phi/B_x|$. (The algebraically allowed choice of the opposite signs before the two radicals would yield $v_\phi/c = 1$ at $x = 1$, implying infinite γ , and so is rejected.) The outflow of angular momentum from the star in both hemispheres across a closed surface Σ with local outward unit normal \mathbf{n} is (e.g., Mestel, 1999)

$$-\int \frac{\varpi B_\phi}{4\pi} \mathbf{B}_p \cdot \mathbf{n} d\Sigma = -\left(\frac{c}{\alpha}\right)^3 B_{lc}^2 \int_0^{P_c} S(P) dP, \quad (46)$$

on use of (2), (6) and (13).

5. The $S(P)$ relation

As emphasized above, if there do exist magnetospheric models that are everywhere dissipationless, with the field force-free outside singular regions such as the equatorial sheet, then allowed relations $S(P)$ should emerge as part of the solution. An early model of the whole magnetosphere by Michel (1974, 1991) has no dead zone, but a poloidal field that is radial all the way from the star to infinity. In our notation, the Michel field is

$$\begin{aligned} P &= P_c \left(1 - \frac{z}{(x^2 + z^2)^{1/2}}\right), & B_x &= -P_c \frac{x}{(x^2 + z^2)^{3/2}}, \\ B_z &= -P_c \frac{z}{(x^2 + z^2)^{3/2}} \end{aligned} \quad (47)$$

with $P_c < 0$ in the northern hemisphere, and with the critical field line $P = P_c$ coinciding with the equator. Michel's $S(P)$ relation is

$$S = -2P + \frac{P^2}{P_c} = -P_c \frac{x^2}{(x^2 + z^2)}, \quad B_\phi = -\frac{S}{x}, \quad (48)$$

so that

$$\frac{dS}{dP} = -2 \left(1 - \frac{P}{P_c} \right) = -\frac{2z}{(x^2 + z^2)^{1/2}}, \quad S \frac{dS}{dP} = 2P_c \frac{x^2 z}{(x^2 + z^2)^{3/2}}. \quad (49)$$

One can easily verify that (34) is satisfied. Note that the poloidal field (47) is radial and (away from the equator) independent of the spherical polar angle θ , and so has no curl: the equilibrium condition is a balance between the electric force given by $\rho_e \mathbf{E}$ and the force due to the toroidal field (48). However, since P_c changes sign on the equator, there are again both toroidal and poloidal equatorial sheet currents that respectively maintain the jumps in B_x and B_ϕ ; also $B_\phi^2 - (x^2 - 1)B_x^2 = P_c^2/x^4$ when $z = 0$, so that by the above discussion, $p_{eq} > 0$. Note also that $dS/dP = 0$ on the critical line $P = P_c$.

The field (47) will need modification near the star in order to link up smoothly with the dipolar poloidal field on the star. We anticipate that there will also be an associated dead zone within the l-c, similar to that found in the MMWP, FM and MS fields, and analogous to that in the non-relativistic wind problem (e.g., Mestel and Spruit, 1987). The dead zone terminates at the point $(x_c, 0)$; the value of x_c (≤ 1) will be seen to be an extra parameter, fixing the global field structure. Within the dead zone the field lines close, crossing the equator normally so that the appropriate equatorial boundary condition is $\partial P / \partial z \propto x B_x(x, 0) = 0$ for $x < x_c$. The dead zone is bounded by the separatrix field line $z = z_c(x)$ with $P = P_c$. In the wind zone outside P_c , the wind flow is along the poloidal field, and so the equatorial boundary condition is $\partial P / \partial x \propto x B_z = 0$. Note that this condition holds not just beyond the l-c but from $x_c < x < \infty$. Again there is a charge-current sheet maintaining the discontinuous sign changes in E_z , B_x and B_ϕ . Thus the critical field line $P(x, 0) = P_c$, extending along the equator from x_c outward, is the continuation of the separatrix between the wind and dead zones.

Just outside the equatorial charge-current sheet and the separatrix, (34) holds all the way in from ∞ . At the intersection $(1, 0)$ with the l-c and just outside the sheet, with $\partial P / \partial x \propto x B_z = 0$ and the second derivatives $\partial^2 P / \partial x^2$ and $\partial^2 P / \partial z^2$ non-singular, (34) requires that the constant value of $S dS/dP$ along P_c must be zero. But $S(P_c) \neq 0$, because as seen above, beyond the l-c the pinching force exerted by B_ϕ is necessary for equilibrium; we therefore require, as in the Michel field,

$$\left(\frac{dS}{dP} \right) = 0 \quad (50)$$

on the critical field line P_c . This condition is propagated inward along P_c on to its continuation, the separatrix P_c between the wind and dead

zones: the equatorial equilibrium conditions beyond the l-c impose a constraint on the global $S(P)$ relation. From (7), condition (50) requires that the poloidal volume current density \mathbf{j}_p falls to zero on P_c , but the finite value for $S(P_c)$ requires a poloidal current sheet at the equator beyond x_c and along the separatrix within it.

6. The inner domain

We have emphasized that beyond the neutral point $(x_c, 0)$, there is a finite thermal pressure on the equator, necessary for equilibrium. Likewise, it is not obvious that within the l-c, the boundary condition on the field line separatrix between the wind zone (labeled 2) and the dead zone (labeled 1) can be satisfied without a thermal pressure within the dead zone.

It is in fact easy to generalize (33) and (34) to include a pressure gradient: with gravity and inertia still negligible,

$$\begin{aligned} & -\frac{\nabla P}{4\pi} \left\{ \frac{1}{\varpi} (\nabla \times \mathbf{B})_\phi \left[1 - \left(\frac{\alpha\varpi}{c} \right)^2 \right] + \left(\frac{\alpha\varpi}{c} \right)^2 \frac{2B_z}{\varpi^2} + \left(\frac{4\pi}{c\varpi} \right)^2 S \frac{dS}{dP} \right\} \\ & - \nabla p = \mathbf{0}. \end{aligned} \quad (51)$$

Thus $p = p(P)$ – the constant pressure surfaces must coincide with the poloidal field lines. In normalized form, (51) becomes

$$(1-x^2) \frac{\partial^2 P}{\partial x^2} - \frac{(1+x^2)}{x} \frac{\partial P}{\partial x} + (1-x^2) \frac{\partial^2 P}{\partial z^2} = -S \frac{dS}{dP} - \frac{1}{2} x^2 \frac{dp}{dP}. \quad (52)$$

Prima facie, there is no obvious objection to the adoption of the simplest case, with p the same constant on all the field lines within the dead zone and zero in the wind zone, i.e., with a discontinuity in both p and $S(P)$ on the separatrix. The equation for P within the wind zone then remains (34) (with signs reversed for convenience):

$$(1-x^2) \frac{\partial^2 P}{\partial x^2} - \frac{(1+x^2)}{x} \frac{\partial P}{\partial x} + (1-x^2) \frac{\partial^2 P}{\partial z^2} = -S \frac{dS}{dP}; \quad (53)$$

and in the dead zone we have the same equation with $S = 0$. However, along the separatrix P_c , extending inward from the equatorial point $(x_c, 0)$, the equilibrium conditions require a discontinuity in B_p as well as those in p and $S(P)$. Writing

$$n_i = -[\mathbf{t} \times (\mathbf{B}_p / B_p)]_i, \quad (54)$$

the unit normal to the separatrix, we require continuity of

$$[-(8\pi p + E^2 + B_p^2 + B_\phi^2)\delta_{ij} + 2E_i E_j + 2B_i B_j]n_j \quad (55)$$

with \mathbf{E} given by (8). The components of (55) parallel to \mathbf{t} and to the separatrix are automatically zero. The component normal to P_c reduces to continuity of

$$8\pi p + B_p^2 \left[1 - \left(\frac{\alpha\varpi}{c} \right)^2 \right] + B_\phi^2, \quad (56)$$

i.e., to

$$8\pi p + B_{p1}^2(1 - x^2) = B_{p2}^2(1 - x^2) + B_{\phi2}^2. \quad (57)$$

It is convenient to normalize p in units of $B_{lc}^2/8\pi$.

For $x > x_c$, the wind zone 2 extends from the equator $z = 0$ to $z = \infty$; for $x < x_c$, zone 2 extends from $z_c(x)$ defined by the separatrix

$$P(x, z_c) = P_c \quad (58)$$

to $z = \infty$. The separatrix function (58) is not known *a priori* but must emerge as part of the solution by iteration. We assume provisionally that at the point $(x_c, 0)$, the poloidal field $B_p = 0$ both just outside and just inside the separatrix. From (57), the pressure of the toroidal field is balanced by the thermal pressure p , so that

$$p = S^2(P_c)/x_c^2. \quad (59)$$

For $x < x_c$, with use of (13) and (14), (57) then becomes

$$|\nabla P_1|^2 = |\nabla P_2|^2 + S^2(P_c) \frac{(1 - x^2/x_c^2)}{(1 - x^2)}. \quad (60)$$

The discontinuity in $|\nabla P|$ grows from zero as x moves in from x_c but will become a small fraction of $|\nabla P|$ for small x .

When $x_c < 1$, the function P has a simple analytical behavior near the neutral point $(x_c, 0)$. The separatrix leaves $(x_c, 0)$ making an angle $\theta = 2\pi/3$ with the outward-pointing equator. To leading order, in both the dead and wind zones, $\partial^2 P / \partial x^2 + \partial^2 P / \partial z^2 = 0$, which has the local solution

$$P/P_c = 1 + A_{2,1}R^{3/2} \sin(3\theta/2), \quad R^2 = z^2 + (x - x_c)^2, \quad (61)$$

where the coefficients $A_{2,1}$ apply respectively to the wind zone $0 < \theta < 2\pi/3$ and the dead zone $2\pi/3 < \theta < \pi$. It is seen that along $\theta = 0$,

$P = P_c$, and on $\theta = \pi$, $P_\theta \propto P_z = 0$, as required. Across the separatrix $z = 3\sqrt{3}(x_c - x)$, P is continuous, while the jump condition (60) then yields

$$A_1^2 = A_2^2 + \frac{4S^2(P_c)}{9P_c^2 x_c (1 - x_c^2)}. \quad (62)$$

At this point, it is instructive to make a comparison with the analogous non-relativistic problem, in which the electric stresses are small by factors $O(v/c)^2$ and so are negligible, and also inertial forces are still neglected. Suppose that there is again a separatrix passing through the poloidal field neutral point at $(x_c, 0)$. The balance equation across the separatrix is now continuity of $8\pi p + B_p^2 + B_\phi^2$ – the $(1 - x^2)$ factors in (57) are replaced by unity. If again p is negligible in the wind zone, then the constant value of p along the separatrix within the dead zone is again fixed by the condition at the neutral point – $p = S^2(P_c)/x_c^2$ in normalized form. The balance condition is then $S^2(P_c)(1 - x^2/x_c^2) + |\nabla P_2|^2 = |\nabla P_1|^2$. The presence of the factor $(1 - x^2/x_c^2)$ in the S^2 term enables \mathbf{B}_p to be continuous (and zero) at the neutral point $(x_c, 0)$, but with $S^2(P_c)$ still non-zero. (Clearly, in the non-relativistic problem, the numerical value of x_c is of no significance.)

By contrast, in the relativistic problem, near the l-c the electric field strength approaches the poloidal magnetic field strength, so that the factor $(1 - x^2)$ now appears multiplying both the $|\nabla P|^2$ terms in (60). If now one were to take $x_c = 1$, then the non-vanishing factor $(1 - x^2)$ would cancel, and (60) would reduce to

$$S^2(P_c) + |\nabla P_2|^2 = |\nabla P_1|^2. \quad (63)$$

along the separatrix. At the point $(x_c, 0)$, the simultaneous vanishing of ∇P_1 and ∇P_2 would then require that $S(P_c) = 0$ (and so also by (59) $p = 0$). But if beyond the l-c the equatorial toroidal component $B_\phi = -S(P_c)/x$ were zero, then the balance condition could not be satisfied (cf., (39)).

An essential step in all the arguments is the condition $p = p(P)$, following from (51). If instead the thermal pressure p were (illicitly) allowed to vary so as to balance the pressure $B_{\phi 2}^2 = S^2(P_c)/x^2$ exerted by the external toroidal field, then from (57) there would be no discontinuity in B_p at the equator or indeed anywhere along the separatrix.

It is appropriate also to re-emphasize just how crucial to the discussion is the outer domain equatorial boundary condition

$$B_z(x, 0) = 0 \quad (64)$$

for $x > 1$; for it is the consequent local equilibrium requirement $S(P_c) > 0$ that implies a poloidal sheet current within the l-c along the separatrix, leading to the conditions (57) and (60), with a non-vanishing $B_{\phi 2}$. It is then clear that a non-zero $p(P_c)$ is required for the balance condition to hold at the neutral point.

7. Computational results

For computational purposes, the functions $P(x, z)$, $S(P)$ are renormalized:

$$\bar{P} = P/P_c, \quad \bar{S} = S/P_c, \quad (65)$$

where $P = P_c$ on the separatrix $z_c(x)$ and along the segment of the equator $z = 0$, extending from the point $x_c < 1$, terminating the dead zone, through the l-c $x = 1$ to ∞ . With our sign convention, \bar{P} is positive and \bar{S} negative. In the following, the bar will be dropped, all quantities being assumed normalized. Thus on the equator beyond $x = x_c$, $P = 1$, and the normalized dipole at the origin now has the form $P = x^2/|P_c|(x^2+z^2)^{3/2}$.

With the parameters x_c and P_c chosen, we need to find a wind zone function $S(P)$ that yields a solution P which crosses the l-c smoothly. P must also be well behaved at infinity, reduce to the dipolar form near $(x, z) = 0$ and satisfy (60) on the separatrix. The requirement that P_{xx} be non-singular on the l-c yields the constraint

$$2P_x = S \frac{dS}{dP} \quad (66)$$

on $x = 1$. The postulated uniform pressure within the dead zone is then given by (59). Equivalently, a suitable choice of the pressure will fix the extent x_c of the dead zone.

The numerical technique (developed by A.J. Mestel at Imperial College), attacks the fully non-linear problem globally, using finite difference techniques. $P(x, z)$ is found iteratively for fixed values of P_c and x_c . After each iteration, the position of the dead zone boundary is advanced in accordance with the jump condition (60), using under-relaxation. The function $S(P)$ is updated from the continuity condition (66). As $(x^2 + z^2) \rightarrow \infty$, the field is required to be radial with $(xP_x + zP_z) \rightarrow 0$, and approaches the Michel form (47).

Solutions are found for some but not all (x_c, P_c) pairs. As x_c increases the possible P_c values decrease. As expected, lower pressure in the dead zone requires the neutral point x_c to be further out, and the separatrix field line emerges closer to the stellar poles.

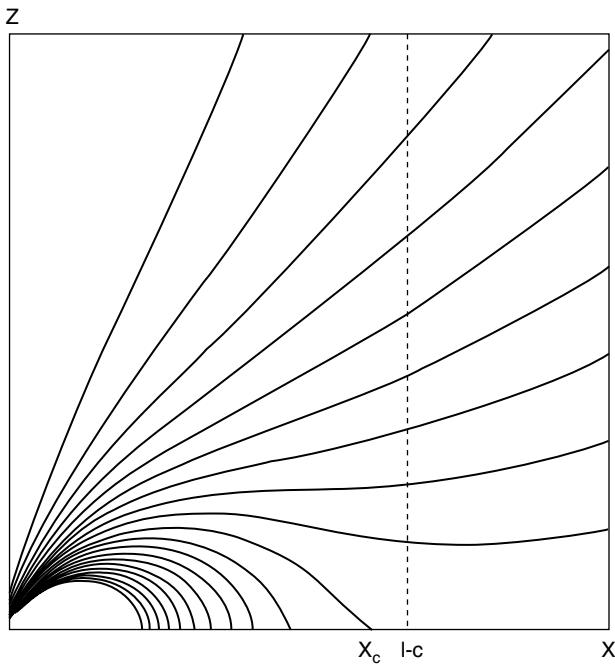


Figure 4. Poloidal field contours for $x_c = 0.9$, $P_c = 1.6$.

Figure 4 portrays the field for $x_c = 0.9$ and $P_c = 1.6$. The $S(P)$ relation is found to deviate only slightly from the Michel form.

Further possible solutions are presented in Goodwin *et al.* (2004) (“GM²W”).

8. Alternative models

We have seen how the imposition beyond x_c of the “perfect conductivity” equatorial boundary condition – $B_z = 0$, $B_x \neq 0$ – reacts back strongly on the conditions that must hold on the separatrix between the dead and wind zones within the l-c. Within the dead zone the field lines cross the equator normally, so that $\partial P / \partial z = 0$ for $x < x_c$, $z = 0$; whereas the condition that the wind zone be perfectly conducting enforces fully open field lines, so that $\partial P / \partial x = 0$ for $x > x_c$, $z = 0$. This in turn yields a non-zero S both at the outer domain equator and on its continuation as the separatrix P_c between the wind and dead zones, so that much of the current returns to the star as a sheet. Simultaneously, the conditions of

equilibrium require a gas pressure both at the equator beyond x_c and within the dead zone.

We now outline how a change in the equatorial physics in the outer domain can enforce qualitative changes in the global model, both within and without the l-c. Suppose that the dead zone extends to a neutral point at $(1, 0)$, but the inner boundary condition $\partial P / \partial z \propto x B_x(x, 0) = 0$ holds in the outer domain $x > 1$ also, allowing some of the field lines in the wind zone to cross the equator normally rather than extend to infinity. Away from the equator the force-free and MHD approximations continue to hold, with both the wind and the negative poloidal current flowing along \mathbf{B}_p . However, in a steady state the flow into the equator must be balanced by a trans-field flow of both gases along the equator, requiring dissipation through either a microscopic resistivity or a dynamically-driven macro-resistivity. Likewise, the return current to the star along the equator must cross \mathbf{B}_p , yielding a magnetic torque, to be balanced by toroidal inertial and/or drag forces. Again (33) yields $S dS/dP = 0$ at the neutral point $(1, 0)$, and so also along the separatrix P_c between the wind and dead zones within the l-c, but now the simple choice $S(P_c) = 0$ is not ruled out, as there is no longer a requirement that S be constant on the equator beyond $x = 1$: $S(P)$ can go to zero continuously, without there needing to be a sheet current along P_c , nor need we introduce a pressure into the dead zone.

These apparent simplifications, however, come at a price: at least for the cases studied, with the equatorial boundary condition $\partial P / \partial z \propto x B_x = 0$ holding everywhere, a solution that is well behaved and continuous in both P and P_x at the l-c blows up before it can reach infinity. As a simple example, consider again (18), the force-free equation with $S = 0$, and extrapolate it to apply beyond the l-c, imposing continuity at $x = 1$ of both B_x and B_z . (This requires the GJ charge density to have a notional velocity greater than c , so clearly the domain $x > 1$ could not be charge-separated: $(\rho_e)_{GJ}$ would need to be as in a normal plasma, the small difference between two large densities n_+, n_- , each with appropriate sub-luminal toroidal velocities. It is not suggested that this is a realistic model: it is purely illustrative.) Again P is written as a Fourier cosine transform (19) which automatically satisfies the boundary condition $B_x(x, 0) = 0$. From standard asymptotics, it is easily shown (Mestel, 1999, p. 547) that when $x > 1$ and for large k , $f_k \propto k^{1/2} \exp k(x - 2)$ – the inner solution cannot be extrapolated beyond two l-c radii.

In Mestel and Shibata (1994) (MS), the live model satisfying $S(P_c) = 0$ was studied, with the particular choice $S(P) = -2P + 2P^2/S(P_c)$ was studied, again adopting the equatorial boundary condition $\partial P / \partial z = 0$,

and assuming continuity of P and P_x on the l-c. A similar breakdown was found to occur, somewhat nearer the l-c, at $x = 1.4$. Significantly, with this changed field line geometry, the assumption of MHD flow also breaks down: along each field line the equations predict $\gamma \rightarrow \infty$ close to the point where the constructed field begins to diverge.

The conclusion drawn in MS is that if the equatorial boundary condition $B_x = 0$ is imposed, then there has to be a breakdown in the simple plasma condition $\mathbf{E} \cdot \mathbf{B} = 0$, not only on the equator, but also in a thin volume domain, idealized as a cylindrical shell symmetric about the rotation axis. Plausibly, the necessary dissipative process will be dynamically driven. The studied examples of force-free fields predict toroidal field components which become large compared with the poloidal as the surface of breakdown is approached. A likely spontaneous hydromagnetic instability of such a field would be equivalent to a large macroscopic resistivity (Mestel and Weiss, 1987; Lyubarskii, 1992; Begelman, 1998).

These contrasting models focus attention on the appropriate equatorial boundary condition. Consider the problem discussed in Sections 4–7. From experience of non-relativistic MHD, one is led to wonder whether in an electromagnetically dominant system — as is implied by the force-free assumption — the field beyond the l-c would persist in a quasi-radial structure, which is associated rather with a field pulled out by a quasi-radial wind in a kinetically dominant system. However, if one adopts the alternative picture, with field lines attempting to cross the equator normally, then as noted, MHD-type flow predicts that the particles acquire high γ -values, invalidating locally the force-free and possibly also the dissipation-free assumptions.

The tentative conclusion is that neither of the two extreme cases is realistic, but that the field lines will approach the equator at a finite angle less than $\pi/2$, yielding both non-zero B_x and B_z , and also large γ -values as gas approaches the equator. In the thin equatorial zone, the z -equation will again require a pressure to balance both the pinching electromagnetic stresses and the momentum carried by the inflowing wind particles. Further, the continuity condition will demand there be a dissipative process that will allow radial equatorial flow across B_z . Significantly, with B_z non-zero, the associated electric field yields a strong radial electric force that must be included in the dynamics of the equatorial zone. It is not yet clear whether the preferred model, satisfying all the boundary conditions, will — as in the GM²W model but unlike the MS model — limit dissipation to the equatorial zone, retaining some of the return current as a sheet on the separatrix between the dead and wind zones.

9. Magnetospheric gaps

It was remarked in Section 3.2 that aligned magnetosphere models can include gap domains in which the charge-density is zero, with a consequent electric field component along \mathbf{B} , affecting strongly the motion of individual migratory particles. The Holloway-type gap is stable, in the sense that an electron straying in from the circumpolar domain will be driven back, and likewise a positive particle coming from the equatorial domain. However, in the live classical model of Section 2.2, an electron that has crossed field lines beyond the l-c through radiation of angular momentum (the “dissipative drift”), and is making its way back to the star, will receive an extra boost as it crosses the gap. And, as noted, pair creation by the returning high- γ electrons in the strong magnetic field near the star could in fact convert the classical model into a quantum model.

More recently, a series of papers have studied the structure of gaps forming in the outer magnetospheres, especially of non-aligned models (a representative sample includes Cheng *et al.*, 1986; Chiang and Romani, 1994; Romani, 1996; Hirotani, 2000; Hirotani and Shibata, 2001; and references therein). In the Chiang and Romani model, the gap is bounded by the last closed field line, and extends from the neutral surface $B_z = 0$ to the l-c. Migratory electrons and/or positrons — with a net charge density well below the local GJ density — will again be accelerated to γ -values high enough to yield curvature gamma-rays. Electron-positron pairs are now produced by collision with thermal X-ray photons. Besides being accelerated directly by the electric field, these new pairs will be boosted to high energy by inverse Compton interaction with high-energy photons. As with the polar cap cascade models, a self-consistent treatment must allow for the partial screening by the pairs of the accelerating electric field.

References

- Arons, J., “Pair creation above pulsar polar caps: steady flow in the surface acceleration zone and polar cap X-ray emission,” *Astrophys. J.* **248**, 1099–1116 (1981).
- Arons, J., “Pair creation above pulsar polar caps: geometrical structure and energetics of slot gaps,” *Astrophys. J.* **266**, 215–241 (1983).
- Arons, J., “Magnetospheric structure of rotation-powered neutron stars,” in: *Magnetospheric Structure and Emission Mechanisms of Radio Pulsars* (Eds. J. A. Gil, T. M. Hankins and J. M. Rankin), IAU Colloquium **128**, pp. 56–77. Pedagogical University Press, Zielona Gora, Poland (1992).

- Arons, J. and Scharlemann, E. T., "Pair formation above pulsar polar caps: structure of the low altitude acceleration zone," *Astrophys. J.* **231**, 854–879 (1979).
- Begelman, M. C., "Instability of toroidal magnetic fields in jets and plerions," *Astrophys. J.* **493**, 291–300 (1998).
- Beskin, V. S., "General relativity effects on electrodynamic processes in radio pulsars," *Sov. Astron. Lett.* **16**, 286–289 (1990).
- Beskin, V. S., Gurevich, A. V. and Istomin, Ya. N., *Physics of the Pulsar Magnetosphere*, Cambridge University Press (1993).
- Blackman, E. G. and Field, G. B., "Ohm's law for a relativistic pair plasma," *Phys. Rev. Lett.* **71**, 3481–3484 (1993).
- Cheng, K. S., Ho, C. and Ruderman, M., "Energetic radiation from rapidly spinning pulsars. I. Outer magnetospheric gaps," *Astrophys. J.* **300**, 500–521 (1986).
- Chiang, J. and Romani, R. W., "An outer gap model of high-energy emission from rotation-powered pulsars," *Astrophys. J.* **436**, 754–761 (1994).
- Contopoulos, J., "Force-free self-similar magnetically driven relativistic jets," *Astrophys. J.* **446**, 67–74 (1995).
- Contopoulos, J., Kazanas, D. and Fendt, C., "The axisymmetric pulsar magnetosphere," *Astrophys. J.* **511**, 351–358 (1999) (CKF).
- Daugherty, J. K. and Harding, A. K., "Electromagnetic cascades in pulsars," *Astrophys. J.* **252**, 337–347 (1982).
- Fawley, W. M., Arons, J. and Scharlemann, E. T., "Potential drops above pulsar polar caps: acceleration of non-neutral beams from the stellar surface," *Astrophys. J.* **217**, 227–243 (1977).
- Fitzpatrick, R. and Mestel, L., "Pulsar electrodynamics I, II," *Mon. Not. R. Astron. Soc.* **232**, 277–302 and 303–321 (1988a,b)(FM).
- Gold, T., "Rotating neutron stars and the nature of pulsars," *Nature* **221**, 25–27 (1969).
- Goldreich, P. and Julian, W. H., "Pulsar electrodynamics," *Astrophys. J.* **157**, 869–880 (1969) (GJ).
- Goodwin, S. P., Mestel, J., Mestel, L. and Wright, G. A. E., "The relativistic force-free equation," *Mon. Not. R. Astron. Soc.* **349**, 213–224 (2004) (GM²W).
- Graham-Smith, F., "The radio emission from pulsars," *Rep. Prog. Phys.* **66**, 173–238 (2003).
- Harding, A. K. and Muslimov, A. G., "Pulsar polar cap heating and surface thermal X-ray emission. I. Curvature radiation pair fronts," *Astrophys. J.* **556**, 987–1001 (2001).
- Harding, A. K. and Muslimov, A. G., "Pulsar polar cap heating and surface thermal X-ray emission. II. Inverse Compton radiation pair fronts," *Astrophys. J.* **568**, 862–877 (2002).
- Harding, A. K., Muslimov, A. G. and Zhang, B., "Regimes of pulsar pair formation and particle energetics," *Astrophys. J.* **576**, 366–375 (2002).
- Hibschman, J. A., "Gamma-ray emission from rotation-powered pulsars," *Astrophys. J.* **565**, 1183–1190 (2002).

- Hibschman, J. A. and Arons, J., "Pair production multiplicities in rotation-powered pulsars," *Astrophys. J.* **560**, 871–884 (2001).
- Hirotani, K., "Gamma-ray emission from pulsars: strength of the acceleration field in the outer gap," *Mon. Not. R. Astron. Soc.* **317**, 225–233 (2000).
- Hirotani, K. and Shibata, S., "Electrodynamic structure of an outer gap accelerator: location of the gap and the gamma-ray emission from the Crab pulsar," *Astrophys. J.* **558**, 216–227 (2001).
- Holloway, N. J., "p-n junctions in pulsar magnetospheres?," *Nature Phys. Sci.* **246**, 6–9 (1973).
- Holloway, N. J., "Angular momentum and energy loss from pulsars," *Mon. Not. R. Astron. Soc.* **181**, 9P–12P (1977).
- Holloway, N. J. and Pryce, M. H. L., "Properties of gaps in pulsar magnetospheres," *Mon. Not. R. Astron. Soc.* **194**, 95–110 (1981).
- Jones, P. B., "Pair formation and electric field boundary conditions at neutron star magnetic poles," *Mon. Not. R. Astron. Soc.* **192**, 847–860 (1980).
- Lynden-Bell, D., "Magnetic collimation by accretion discs of quasars and stars," *Mon. Not. R. Astron. Soc.* **279**, 389–401 (1996).
- Lyubarskii, Yu. E., "Energy release in strongly magnetized relativistic winds," *Sov. Astron. Lett.* **18**, 5, 356–359 (1992).
- Melatos, A. and Melrose, D. B., "Energy transport in a rotation-modulated pulsar wind," *Mon. Not. R. Astron. Soc.* **279**, 1168–1190 (1996).
- Mestel, L., *Stellar Magnetism*, Clarendon Press, Oxford (1999). (Paperback edition 2003)
- Mestel, L., "The relativistic force-free equation," *Publ. Astron. Soc. Aust.* **18**, 1–10 (2001).
- Mestel, L. and Pryce, M. H. L., "On the construction of a simple model pulsar magnetosphere," *Mon. Not. R. Astron. Soc.* **254**, 355–360 (1992) (MP).
- Mestel, L. and Shibata, S., "The axisymmetric pulsar magnetosphere: a new model," *Mon. Not. R. Astron. Soc.* **271**, 621–638 (1994) (MS).
- Mestel, L. and Spruit, H. C., "On magnetic braking of late-type stars," *Mon. Not. R. Astron. Soc.* **226**, 57–66 (1987).
- Mestel, L. and Wang, Y.-M., "The axisymmetric pulsar magnetosphere II," *Mon. Not. R. Astron. Soc.* **188**, 799–812 (1979) (MW).
- Mestel, L. and Weiss, N. O., "Magnetic fields and non-uniform rotation in stellar radiative zones," *Mon. Not. R. Astron. Soc.* **226**, 123–135 (1987).
- Mestel, L., Robertson, J. A., Wang, Y.-M. and Westfold, K. C., "The axisymmetric pulsar magnetosphere," *Mon. Not. R. Astron. Soc.* **217**, 443–484 (1985) (MRWW).
- Michel, F. C., "Rotating magnetospheres: a simple relativistic model," *Astrophys. J.* **180**, 207–225 (1973).
- Michel, F. C., "Rotating magnetospheres: acceleration of plasma from the surface," *Astrophys. J.* **192**, 713–718 (1974).
- Michel, F. C., *Theory of Neutron Star Magnetospheres*, University of Chicago Press, Chicago (1991).
- Muslimov, A. G. and Tsygan, A. I., "General relativistic electric potential drops above pulsar polar caps," *Mon. Not. R. Astron. Soc.* **255**, 61–70 (1992).

- Priest, E. R. and Forbes, T., *Magnetic Reconnection*, Cambridge University Press (2000).
- Rankin, J. M. and Wright, G. A. E., "Circulating sub-beam systems and the physics of pulsar emission," *Astron. Astrophys. Rev.* **12**(1), 43–69 (2003).
- Romani, R. W., "Gamma-ray pulsars: radiation processes in the outer magnetosphere," *Astrophys. J.* **470**, 469–478 (1996).
- Ruderman, M. A. and Sutherland, P. G., "Theory of pulsars: polar gaps, sparks and coherent microwave radiation," *Astrophys. J.* **196**, 51–72 (1975).
- Scharlemann, E. T. and Wagoner, R. V., "Aligned rotating magnetospheres. I. General analysis," *Astrophys. J.* **182**, 951–960 (1973).
- Scharlemann, E. T., Arons, J. and Fawley, W. M., "Potential drops above pulsar polar caps: ultrarelativistic particle acceleration along the curved magnetic field," *Astrophys. J.* **222**, 297–316 (1978).
- Shibata, S., "Magnetosphere of the rotation-powered pulsar: a DC circuit model," *Astrophys. J.* **378**, 239–254 (1991).
- Shibata, S., "The field-aligned accelerator in the pulsar magnetosphere," *Mon. Not. R. Astron. Soc.* **287**, 262–270 (1997).
- Shibata, S., Miyazaki, J. and Takahara, F., "On the electric field screening by electron-positron pairs in the pulsar magnetosphere -I, -II," *Mon. Not. R. Astron. Soc.* **295**, L53–58 (1998); **336**, 233–240 (2002).
- Smith, I. A., Michel, F. C. and Thacker, P. D., "Numerical simulations of aligned neutron-star magnetospheres," *Mon. Not. R. Astron. Soc.* **322**, 209–217 (2001).
- Sturrock, P. A., "A model of pulsars," *Astrophys. J.* **164**, 529–556 (1971).
- Usov, V. V., "Physical processes in pulsar magnetospheres and non-thermal high-frequency emission," in: *Pulsars: Problems and Progress* (Eds. S. Johnston, M. A. Walker and M. Bailes), IAU Colloquium **160**, 323. Astronomical Society Pacific Conference Series, **105** (1996).
- Wright, G. A. E., "An empirical model for the radio emission from pulsars," *Mon. Not. R. Astron. Soc.* **344**, 1041–1056 (2003).

5 Magnetic fields in galaxies

Ellen G. Zweibel

Departments of Astronomy and Physics, and Center for Magnetic Self-Organization, University of Wisconsin, Madison, WI 53706, USA

Disk galaxies, of which the Milky Way is a typical example, possess large scale magnetic fields. The interstellar gas in which the fields reside is removed and replaced on a timescale of about 1/10 of the age of the Galaxy, and the new gas must acquire a magnetic field of the correct strength and direction on this timescale in order to maintain a steady state. The outstanding problem facing all theories for the origin and evolution of galactic magnetic fields is identifying the mechanisms for magnetic diffusion. Diffusion is necessary to prevent the buildup of magnetic energy at the resistive scale, to magnetically assimilate new material, and to remove the field from the galaxy. Some of these processes require topological changes in the field, and necessitate resistive or collisionless effects. Others only require that the field diffuse with respect to the center of mass, and could occur through ion-neutral drift. In addition to diffusion, there must be a mechanism for maintaining the large scale field.

1. Introduction

The theory of astrophysical dynamos was developed to explain the solar cycle: the observed redistribution and polarity reversals of the surface magnetic field on a timescale much shorter than the Ohmic time demands a dynamical explanation. Although alternatives involving long term storage of magnetic flux in the solar core were proposed, there is now overwhelming consensus that the solar activity cycle is the manifestation of a dynamo.

The situation with regard to galaxies is different. Nearby galaxies with detectable quantities of interstellar gas emit polarized radio frequency radiation which is thought to be synchrotron radiation generated by relativistic electrons (so-called “cosmic ray” electrons) gyrating in a magnetic field. The radio luminosity is well correlated with the

far infrared luminosity, which is a measure of the rate at which massive stars are forming. There is no evidence for cyclic behavior, or for variability independent of the star formation rate. Moreover, since the Ohmic decay time of the magnetic field is about 10 orders of magnitude longer than the age of the Universe, one might simply assume that galactic magnetic fields are primordial, or at least were generated before galaxies formed. Indeed, it is not straightforward to prove otherwise.

This paper concerns the origin and evolution of galactic magnetic fields as revealed by their observed properties and by related aspects of galactic structure and evolution. The array of topics is so broad that it appeared impossible to make the paper bibliographically complete. Relevant review articles and books are cited in the Appendix so that readers new to the field can pursue particular topics, and literature is cited up to 2004. Most of the discussion focuses on our own Galaxy, the Milky Way, for which the most detailed observations are available, and which is thought to be a typical example of its kind.

Section 2 is a brief introduction to those aspects of galactic structure and evolution which seem relevant to understanding galactic magnetic fields. Section 3 summarizes the observed properties of magnetic fields in galaxies at the present epoch, while Section 4 is a history of magnetic fields in the Universe. Section 5 is concerned with the basic physical processes which are relevant to the generation and evolution of galactic magnetic fields. Section 6 is a summary and discussion of future work.

2. Galactic structure

The Milky Way is the second largest member of a group of ~ 20 galaxies called the Local Group. M31, the Andromeda galaxy, is the dominant galaxy, while the Magellanic Clouds, the two best known dwarf members, are satellites of the Milky Way. Most known galaxies belong to small groups or to large clusters.

The Milky Way is an example of a large class of galaxies known as spirals, named for the strong suggestion of spiral structure in their optical appearance. At visual wavelengths, which probe the visible stars, a spiral galaxy consists of a disk with a radial distribution of light that declines exponentially over a scale of several kiloparsecs (kpc) and vertically over a few hundred parsecs. The innermost kpc is dominated by a spheroidal distribution of stars known as the bulge. There is probably an oval distortion known as a bar (galaxies with long, prominent bars are known as barred spirals). There is also a faint, spheroidal

component of stars several tens of kpc in radius which is considered part of the galactic halo. The orbits of disk stars are nearly circular, with a strong prevailing sense of rotation and a well-defined plane. Bulge and halo stars follow fully three-dimensional orbits. The disk stars have higher metallicity (astronomers consider all elements except H and He to be metals) than the halo stars, and are thought to be younger. The age of the oldest stars is estimated to be more than 10^{10} years, and star formation continues in the present epoch.

The mass distribution in galaxies is determined by measuring the orbital velocities of stars and gas and assuming they are in dynamical equilibrium. This reveals that the mass does not follow the light: more than 90% of the matter in disk galaxies is in unseen form. This so-called dark matter follows an extended spheroidal distribution, its density declining with radius as r^{-2} . The differential rotation profile of the disk material in the potential well of the halo is $\Omega \propto r^{-1}$, or $V \sim \text{constant}$. In the Milky Way, $V \sim 220 \text{ km s}^{-1}$, with a rotation period at the solar circle of about 2×10^8 years. The dynamics in the innermost few pc of the Galaxy are dominated by a $10^6 M_\odot$ black hole.

In our discussion of galactic magnetic fields, we are primarily concerned with the gaseous component, which is known as the interstellar medium (ISM). The mass of the ISM is about $10^9 M_\odot$, only 1% of the mass in visible stars, but since new stars are continually forming from interstellar gas, and evolved stars are enriching it with heavy elements, the ISM is a major driver of galactic evolution. The kinematics of interstellar gas is similar to that of the stars: there is a rotationally supported disk component and tenuous halo component supported primarily by turbulent and thermal pressure.

In addition to the thermal component of the interstellar medium, there is a relativistic, or cosmic ray component. We have already alluded to cosmic ray electrons as the source of the polarized radio continuum, but about 98% of cosmic ray particles are nuclei, most of them protons. The energy density in cosmic rays is comparable to that in the thermal gas, and in the magnetic field. Their energy spectrum is well fitted by a broken power law — i.e., highly nonthermal — but their angular distribution is highly isotropic. The nuclear component of cosmic rays is highly enriched in the light elements Li, Be, and B compared to other cosmic sources. These elements are thought to be produced by spallation reactions between cosmic rays and ambient interstellar material. Measurements of the abundances of the light elements, including the relative abundances of radioactive isotopes, show that individual cosmic ray particles remain within the Galaxy for $1\text{--}2 \times 10^7$ years, and that the average density of the medium in which they are stored is

about 0.1 cm^{-3} . The latter value is representative of the density in the lower Galactic halo.

Interstellar gas has a wide range of thermal states, from tenuous and hot ($n \sim 10^{-2} - 10^{-3} \text{ cm}^{-3}$; $T > 10^6 K$) to dense and cold ($n \sim 10^3 - 10^6 \text{ cm}^{-3}$; $T \sim 10 K$). Most of the mass is weakly ionized, cold atomic or molecular gas, but most of the volume is fully or significantly ionized, with the filling factor of ionized gas increasing above the Galactic plane. A given parcel of interstellar gas switches thermal states, which is itself a fairly rapid process, on timescales of tens of millions of years through exposure to heating or to strong compression. The different thermal states are intermingled, with the largest coherent structures being hot supernova and stellar wind-driven superbubbles of order 100–200 pc in size.

The cooler gas is kept in a state of partial ionization by ambient starlight and by low energy cosmic rays, while gas hotter than about $10^4 K$ is collisionally ionized. The resulting degree of ionization, together with the lengthscales that characterize galactic structure, result in a Lundquist number of $10^{15} - 10^{21}$. Galaxies are superconducting beyond the wildest dreams of any condensed matter physicist.

Although the dynamical equilibrium of the disk gas is determined primarily by rotation, the gas velocity also has a small random, or turbulent component of order 10 km s^{-1} . It is this random component, together with magnetic stresses and cosmic ray pressure, that determines the vertical thickness of the gas disk. The turbulent flow is typically supersonic in all but the hottest gas, and readily forms shocks. The radiative efficiency of the heated, compressed gas in these shocks is so high that if the turbulence were not driven it would decay in about 10^7 years. A few percent of the energy available from supernova explosions suffices to maintain interstellar turbulence. Thus, supernovae are believed to be the main energetic driver of the ISM.

The interstellar medium is an open system. Gas is converted to stars at the rate of a few $M_\odot \text{ yr}^{-1}$. Stars less massive than the Sun — which represent most of the mass in visible stars — evolve so slowly that they have returned little mass to the ISM over the age of the Galaxy. More massive stars evolve more quickly, and return most of their mass to the ISM in the form of stellar winds, planetary nebulae, and supernova explosions. Much of this recycled material has been chemically enriched by nuclear reactions. At the same time, previously extragalactic material is falling into the Milky Way at a rate comparable to the rate of star formation, while disk gas, heated by supernova explosions and powerful stellar winds, is ejected into the halo. Although some of this

gas probably cools and returns to the disk, some of the halo gas may be leaving the galaxy, in the form of a wind. There may also be a radial inflow of material in the plane of the disk, which accumulates near the Galactic Center or is coupled to a wind driven from the inner Galaxy. It is uncertain whether the mass of the ISM is steady over time.

Sources and sinks of interstellar material — each of which leads to the replacement of the ISM on a timescale of $\sim 10^9$ years, less than a tenth the age of the Galaxy — have important implications for the evolution of the Galactic magnetic field. The mass which is returned to the ISM by stars is generally undermagnetized by several orders of magnitude. This is a consequence of the so-called magnetic flux problem in star formation, which refers to the enormous magnetic fields stars would have if almost all the magnetic flux threading their constituent gas were not lost during star formation (Mestel and Spitzer, 1956). There are important exceptions to this rule. About 1% of supernova remnants (known as plerions; the Crab nebula is the prototype) have magnetic fields about 3 orders of magnitude larger than the ambient interstellar field. The solar wind at 1 AU is magnetized at roughly the interstellar level, and the fieldstrength/density ratio B/ρ increases outward as r . This is a general property of rotating, magnetized stellar winds: the azimuthal field B_ϕ decreases as r^{-1} while ρ decreases as r^{-2} . Solar type winds are too tenuous to contribute much to the Galactic field, but if winds from massive stars are magnetized (a point still not settled; see Cohen *et al.*, 2003 and references therein), their contribution could be significant. The magnetization of infalling extragalactic clouds is also unknown. These clouds are observed to contain metals, so they cannot be cosmologically pristine (Richter *et al.*, 2001, Tripp *et al.*, 2003). If they originated wholly or partially within galaxies, one would expect them to be magnetized, and thus a source of magnetic field. On the other hand, a galactic wind, if it exists, would carry away magnetic field.

The direction of the magnetic field associated with interstellar sources and sinks is as important as its intensity. The fields from stellar sources and infalling extragalactic clouds are probably randomly oriented. The azimuthal component of field expelled by supernova explosions tends to be directed opposite to the prevailing Galactic field, because as loops of field rise, they are twisted by Coriolis forces. If the field is efficiently ejected by this mechanism, it could be an important component of a Galactic dynamo (Kulsrud, 1999; Rafikov and Kulsrud, 2000).

An important clue about the facility with which magnetic fields from stellar and extragalactic sources are mixed with the ambient gas comes

from measurements of chemical abundances in the ISM. Abundances are measured spectroscopically, through absorption features created by interstellar gas along the lines of sight to bright stars. Recent measurements suggest that although abundances are quite uniform within ~ 0.5 kpc of the Sun, there is a deficit of heavy elements in the solar vicinity relative to the remainder of the disk (Meyer *et al.*, 1998; André *et al.*, 2003). One explanation for this is recent contamination of the ISM by infalling metal-poor gas (Meyer *et al.*, 1998). Independent dynamical evidence that this occurred roughly 10^7 years ago is presented by Comeron and Torra (1994). These results, taken together, are consistent with a picture in which mixing is efficient, although far from instantaneous.

3. Observations of Galactic magnetic fields

The magnetic field of the Milky Way, and magnetic fields in external galaxies, are detected through their radiative signatures. The polarized radio continuum intensity measures the strength and orientation of the projection of the magnetic field on the plane of the sky, B_{\perp} . It is sensitive to the orientation of B_{\perp} , not its signed direction. As mentioned in Section 1, the polarized continuum is thought to be synchrotron radiation emitted by cosmic ray electrons. If the field orientation varies along the line of sight and within the telescope beam, the polarization is lower than that produced by a uniform field. Denoting the average and random components of the projected field by $B_{\perp u}$ and $B_{\perp r}$, respectively, the observed polarization p is related to the maximum possible polarization p_{max} by $p/p_{max} = B_{\perp u}^2/(B_{\perp u}^2 + B_{\perp r}^2)$. Measurements of p provide an important probe of B_u/B_r .

Nonthermal continuum radiation originates from more or less the entire ISM. Polarized radio frequency radiation is also emitted by point-like sources such as pulsars and extragalactic background sources such as quasars and active Galactic nuclei. As this radiation propagates through the foreground ISM, its plane of polarization is rotated due to the slightly different propagation speeds of left and right circularly polarized electromagnetic waves in magnetized plasma, with the rotation angle increasing with wavelength λ as λ^2 . When combined with an estimate of the electron density, this so-called Faraday rotation gives a measure of \mathbf{B}_{\parallel} , the signed component of \mathbf{B} along the line of sight to the source. Recently, Faraday rotation of the radio continuum has also been mapped in selected fields of view. This has given some insight into the small scale structure of the magnetic field and thermal electron distribution (Haverkorn *et al.*, 2003).

Magnetic fields can also be detected through Zeeman splitting of lines in the spectra of interstellar atoms and molecules. In principle, the full vector magnetic field can be measured through the Zeeman effect, but in practice this has been carried out only for the Sun, the brightest source. Zeeman detections of the Galactic magnetic field yield only \mathbf{B}_{\parallel} , convolved with the density of the tracer species, integrated along the line of sight, and, except in the case of absorption of a background continuum point source, averaged over the telescope beam. Although Zeeman splitting occurs at all wavelengths, the magnitude of the shift is generally an increasing function of wavelength, and to date the interstellar Zeeman effect has only been detected at radio wavelengths. A related line polarization signature, the so-called Goldreich-Kylafis effect (Goldreich and Kylafis, 1982), has also been used to measure magnetic fields in dense clouds (Lai *et al.*, 2003).

A final set of diagnostics exploits the presence of small, aspherical dust grains in the ISM (see the review by Draine, 2003). The grains tend to become oriented with their shortest axes parallel to \mathbf{B} . Absorption and scattering of starlight by this dust — a strong effect in the Galactic plane, where most of the grains reside — results in a small linear polarization from which the orientation of \mathbf{B}_{\perp} can be deduced. The thermal radiation emitted by the dust, which generally peaks in the far infrared, is also polarized, and has been used to map \mathbf{B}_{\perp} in directions toward dense clouds. The relationship between the degree of polarization, the fieldstrength, and the spatial coherence of the field remain obscure, however, due to the influence of environmental factors such as gas density and temperature on grain alignment. For example, it is often observed that the ratio of polarized to continuum intensity in the far infrared decreases toward the centers of dense clouds. Whether this occurs because the dust is not well aligned in cloud centers, or because the field is tangled on small scales, remains an open question (Hildebrand *et al.*, 2000; Padoan *et al.*, 2001).

There are three main points to this brief summary of observational techniques. First, the measurements are always line of sight and sometimes angular averages. Second, the observations yield either the signed \mathbf{B}_{\parallel} or the unsigned B_{\perp} . Third, in each type of measurement \mathbf{B} is convolved with an interstellar tracer: cosmic ray electrons, thermal electrons, atomic or molecular species, or aligned dust grains. Thus, there is a complementarity to the measurements, but little redundancy. This, together with the coarse-graining induced by averaging over large scales, means that it is all but impossible to measure the full vector magnetic field with any precision anywhere in the ISM. It is a triumph of both

ingenuity and diligence that we have as good a picture of Galactic magnetic fields as we do.

With this as background, we now summarize the results. The magnetic field of the Milky Way is measured primarily within several kpc of the Sun. Within this volume, it has a coherent, or uniform component \mathbf{B}_u , and a small scale, or random component \mathbf{B}_r (Rand and Lyne, 1994; Han *et al.*, 1999). A lower limit to the scale of the coherent component is based on the distance to the sources (primarily pulsars). It should be emphasized that the coherence length of \mathbf{B}_u exceeds by more than an order of magnitude the driving scale of interstellar turbulence, and the scales of all observed flow except that due to differential rotation. The orientation of \mathbf{B}_u is parallel to the plane of the disk and nearly azimuthal, with an inclination angle similar to that of the spiral arms. This is similar to what is seen globally in many other spiral galaxies. There are also at least two reversals of the field with radius, possibly as many as five (Han, 2001). Similar reversals have been seen in some, but not all, other spiral galaxies. The vertical scale height of the field is about 1.5 kpc, several times thicker than the gas disk, but comparable to the ionized gas disk.

The fluctuations appear to exist on every spatial scale at which they can be detected. The smallest observable lengths come from mapping the polarization of stars in dense clusters (Serkowski, 1965), or comparing the Faraday rotation between pairs of radio sources (Clegg *et al.*, 1992; Minter and Spangler, 1994). The latter studies reveal fluctuations on subparsec scales.

Observational determinations of the partitioning of magnetic energy between the uniform and random components is somewhat method dependent. The following discussion is a summary of Heiles (1996). If one measures B_u and B_r from Faraday rotation of pulsars, their respective values are $1.4 \mu\text{G}$ and $5.0 \mu\text{G}$, with the ratio $B_u/B_r \sim 0.28$. If the fluctuations are isotropic, the synchrotron polarization implied by these values would be only about 1/3 of what is observed. Estimates of B_u/B_r from synchrotron polarization, on the other hand, suggest $B_u/B_r \sim 0.5$. Starlight polarization suggests larger values yet; $B_u/B_r \sim 0.7\text{--}1.0$, depending on whether one assumes the fluctuations are isotropic or entirely transverse to B_u . Heiles suggests that there are systematic differences between the field structure in the predominantly ionized regions sampled by the pulsar measurements and the denser, predominantly neutral gas which dominates the starlight polarization measurement. An alternative explanation is that the fluctuations in the field have a highly sheared or folded structure, like a hairpin—resulting in larger fluctuations in direction than in orientation (Brown and Taylor, 2001).

The mean fieldstrengths determined by these different methods are quite similar: Heiles adopts $5.2 \mu\text{G}$ from pulsar data, $4.2 \mu\text{G}$ for the azimuthally averaged synchrotron emission, and $5.9 \mu\text{G}$ for the synchrotron emission from spiral arms. Larger magnetic fieldstrengths in dense molecular gas are seen through the Zeeman effect, but there is a large scatter of values and many nondetections, possibly because of projection effects (Crutcher, 1999). The upper envelope of the fieldstrength — density relation in molecular gas is consistent with $B \propto \rho^{1/2}$. In contrast, the $B - \rho$ relation in less dense gas — $n \sim 0.1 - 100 \text{ cm}^{-3}$ — is essentially flat (Troland and Heiles, 1986). The mean fields in molecular gas appear to be 2–3 times larger than the fields in diffuse atomic gas, while the mean density contrast is about a factor of 50.

At the field strengths measured, the magnetic energy density is comparable to the energy in turbulent gas motions. This means that magnetic fields play an important dynamical role in the ISM. They provide substantial support against the large scale Galactic gravitational field in the vertical direction, and against self gravity in dense, star forming regions. They also influence the nature of turbulence and its modes of dissipation, especially on small scales.

An upper limit on the fraction of vertical support that can be provided by magnetic fields and cosmic rays arises from stability considerations. Since these nonthermal components are not directly confined by gravity, they can destabilize the gas layer to a Rayleigh–Taylor-like instability first discussed in the galactic context by Parker (1966). Although the exact stability boundary is model dependent, it is clear that the gas layer is unstable if the energy density of the magnetic field and cosmic rays exceeds the kinetic energy of the gas by more than a factor of order unity.

In addition to results obtained from direct detection, it is possible to infer some properties of Galactic magnetic fields from observations of cosmic rays. The mean density of the volume in which the cosmic rays are stored is representative of the density in the lower Galactic halo, into which we know the thick magnetic disk extends. Since the cosmic ray confinement time is about 3 orders of magnitude longer than naive estimates of the propagation time through the Galaxy, it seems likely that the propagation is diffusive, with the cosmic rays scattering from small scale fluctuations. Because scattering results in very little motion across fieldlines (Jokipii, 1987), it seems that the field must connect the disk to the halo, from which the cosmic rays escape.

The problem of cosmic ray escape raises the question of whether the Galactic magnetic field is connected to an intergalactic field. The magnetization of the intergalactic medium in the Local Group and beyond

is still an open question. Rich clusters of galaxies emit x-rays, which indicates that the cluster is filled with hot gas, and a polarized radio continuum, which demonstrates that the medium is pervaded by cosmic ray electrons and magnetic fields (Clarke *et al.*, 2001 and references therein). Because the hot gas is enriched in metals, it is thought to contain material stripped from galaxies. The magnetic field may likewise be interstellar field stripped from galaxies and modestly amplified up to equipartition with the turbulent gas motions in the cluster.

It is also possible that there is a pervasive intergalactic field, although none has yet been detected. Since the intergalactic medium is fully ionized and its cosmic ray population is negligible, such a field could be detected only through Faraday rotation of extragalactic sources. Thus, upper limits exist primarily on \mathbf{B}_u . These limits are somewhat model dependent, but lie in the vicinity of 10^{-9} G (Kronberg and Perry, 1982; Blasi *et al.*, 1999).

The structure of the magnetic field within about 200 pc of the Galactic Center appears to be qualitatively different from the field in the disk: there is a substantial vertical component, outlined in thin, nearly straight filaments of enhanced synchrotron radiation. The field strength is estimated to be of order 1 mG, much stronger than the field in the disk and somewhat above equipartition with the turbulence in the surrounding molecular gas (Morris and Serabyn, 1996). Whether this vertical field is a local structure or reflects a large scale field, perhaps concentrated near the Galactic Center by a planar accretion flow, as suggested by Chandran *et al.* (2000) and alluded to in the previous section, remains unknown.

4. Galaxy formation and magnetogenesis

Recent advances in the observation and interpretation of the Cosmic Background Radiation have elucidated many important features of the Universe and its evolution over time (Spergel *et al.*, 2003). The age of the Universe is thought to be about 13.4 ± 0.3 Gyr. The matter within it had cooled sufficiently to recombine after about $3.72 \pm 0.14 \times 10^5$ years. After about 2×10^8 years, sources of Lyman continuum radiation appeared which reionized the Universe, and have maintained the intergalactic medium as a fully ionized plasma up to the present time. The agents of reionization were probably a first generation of massive stars, or possibly hot accretion flows onto massive black holes (see, for example, the review by Madau, 2003).

The environments in which the first stars formed were probably quite different from galaxies at the present epoch. The theory of galaxy

formation is not yet complete, but current ideas favor a hierarchical scenario (e.g., Silk and Wyse, 1993) in which large galaxies like the Milky Way are built up by the accretion of dynamically distinct smaller units which have already undergone some star formation. The prevalence of a group or cluster environment for galaxies, as opposed to an isolated setting, and the large number of dwarf galaxies, support this hypothesis. Mergers are expected on theoretical grounds because of the gradual conversion of orbital energy to tides in the galaxies, or heat in the ambient medium. The Milky Way and M31, for example, are thought to be slowly spiraling together. Mergers still occur, although probably at a lower rate than in the past. Evidence for this is seen in the Milky Way in the form of kinematically defined streams of stars and gas torn from small companions by the Galactic tidal field, the Magellanic Stream being the best known example. There is also clear morphological and kinematic evidence for mergers of external galaxies (Gilmore *et al.*, 2002).

Direct evidence for Galactic magnetic fields does not go as far back in time as direct evidence for galaxies, but there are practical obstacles to early detection. The λ^2 dependence of the rotation measure means that Faraday rotation measured at a fixed wavelength would decrease with redshift z as $(1+z)^{-2}$, even for a population of identical sources. Despite this constraint, observations suggest that Galactic halos were sufficiently magnetized to produce Faraday rotation by the time the Universe was about a third its present age (Oren and Wolfe, 1995). There is indirect evidence for magnetic fields at earlier times: some of the oldest stars in the Galaxy contain measurable amounts of light elements in their atmospheres (Duncan *et al.*, 1992; Gilmore *et al.*, 1992). The most plausible origin for these elements is spallation reactions between CNO cosmic ray nuclei and the gas from which these stars formed. If cosmic rays were accelerated and confined in this medium, there must have been a magnetic field. However, it could have been very weak by present Galactic standards, as little as 10^{-12}G (Zweibel, 2003).

The general problem of magnetogenesis in the cosmos is still unsolved. In addition to exotic processes which rely on a modification of Maxwell's equations in the early Universe, and are still speculative (see the reviews by Grasso and Rubenstein, 2001 and Widrow, 2002), there is a robust mechanism for creating a magnetic field in a plasma *ab initio* which is known as Biermann's Battery. Whenever there is an electron pressure gradient, the plasma quickly becomes polarized because of the large electron mobility. The result is an electric field $\mathbf{E} = -(\nabla P_e)/en_e$ which opposes the pressure gradient. If T_e is not a function of n_e , \mathbf{E} is nonpotential, and generates a magnetic field by induction.

It can be shown that the rate of magnetogenesis is just such that the thermal ion gyroradius is about equal to the size l of the system after one thermal ion crossing time $l/(k_B T_e/m_p)^{1/2}$. Thus, a system in which the Biermann Battery has operated barely satisfies the assumptions of magnetohydrodynamics. Since astrophysical systems are normally much more strongly magnetized, it is inescapable that a magnetic field generated by the Biermann battery can only serve as “seed” which is amplified by another mechanism, such as the hydromagnetic dynamo.

One can classify theories of magnetogenesis as either “top down” or “bottom up,” depending on whether they generate large scale, coherent fields or small scale, incoherent fields. The Biermann Battery mechanism can be either top down or bottom up, depending on where it operates. On cosmological scales, where the effect arises in structures such as ionization fronts or shock fronts (Subrahmanian *et al.*, 1994; Kulsrud *et al.*, 1997; Gnedin *et al.*, 2000), the fieldstrength is exceedingly low, about $10^{-18}\text{--}10^{-21}$ G, but is coherent on galactic scales. In this setting, the Biermann Battery is top down. A variant which is effectively bottom up with respect to magnetic fields in galaxies involves generation of a magnetic field in an accretion disk surrounding a massive black hole in an active Galactic nucleus, combined with expulsion of the field in a magnetized jet. Such jets are observed to feed lobes of relativistic, magnetized plasma which can be as much as 2 Mpc in size. Once the active accretion phase is over, the lobes lose energy to synchrotron radiation and adiabatic expansion, eventually becoming invisible. If extinct radio sources filled a large volume of the Universe by the time galaxies formed, they could have been the initial source of Galactic magnetic fields (Daly and Loeb, 1990; Furlanetto and Loeb, 2001; Kronberg *et al.*, 2001).

In bottom up theories of magnetogenesis in galaxies, the field is generated first on small scales, such as in accretion disks or stars, amplified by a dynamo, and spewed into the interstellar medium. The resulting field is chaotic, and according to most estimates, too weak, to account for Galactic magnetic fields as they now exist. It must then be amplified and reconfigured by a dynamo. For example, Rees (1994) has discussed the properties of a seed field built up from strongly magnetized supernova remnants.

5. Building blocks of Galactic magnetic fields

A theory of Galactic magnetic fields must explain the magnitude of both the large scale and small scale components. The field must be

maintained in a steady state despite the effects of sources and sinks of gas and field, which act on $\sim 10^9$ yr time scales. It is difficult to determine the fundamental significance of the strong vertical field near the Galactic Center for the Galactic magnetic field problem as a whole, because it is the only known example (and also the only example detectable with present techniques) of its kind.

Galaxies have the standard ingredients necessary for a hydromagnetic dynamo: large scale differential rotation and small scale turbulence. Over the years there have been a number of mean field dynamo models in the geometry of disk galaxies (see Ferrière and Schmitt, 2000; Moss *et al.*, 2001 for recent examples, and Beck *et al.*, 1996 for a review), as well as a number of calculations of the mean field dynamo tensors α and β (see for example Ferrière, 1996; Hanasz and Lesch, 1997). The models capture the most ubiquitous magnetic feature seen in disk galaxies, namely the predominantly toroidal orientation. This is the outcome expected in any model with strong differential rotation. There are a number of additional properties which depend on the details of the underlying Galaxy and dynamo models, such as whether or not the toroidal field reverses with radius (bisymmetric or axisymmetric spirals, respectively), dipolar or quadrupolar symmetry with respect to the Galactic plane, and linear growth rate of the unstable dynamo modes. Thanks to these models we now have a good understanding of the relationship between the properties of the host galaxy and the predictions of mean field dynamo theory, as well as the timescales on which the field is generated and can respond to temporal changes in α and β (Ko and Parker, 1989).

However, it has become increasingly clear that mean field, kinematic dynamo theory is incomplete. The approximate equality of the magnetic and turbulent kinetic energies implies that the effect of the field on the small scale motions cannot be ignored. Perhaps a more serious problem is that it is difficult to reconcile the existence of the strong mean field \mathbf{B}_u with the enormous Lundquist number in galaxies. In the absence of diffusive effects, amplification of the field is accompanied by lengthening of the fieldlines by the same factor. If the field is confined to a constant volume, it must become folded, tangled, or otherwise random. The result is that the mean field is many orders of magnitude less than the small scale field. This was predicted on analytical grounds using mean field theory (Kulsrud and Anderson, 1992), and has also been seen in simulations of small scale dynamos at high Lundquist number (Schekochihin *et al.*, 2002).

In view of the problems with conventional dynamo theory, it is reasonable to ask whether Galactic fields could be primordial, i.e., whether

a coherent field could have been present in the material from which galaxies formed and been only minimally processed since that time. Howard and Kulsrud (1997) have investigated the evolution of a primordial field which is initially laid down coherently on a galactic disk. Differential rotation winds the field up, and the strength of the toroidal component B_θ at first increases linearly with time while acquiring an increasing number of reversals with radius. In a disk of characteristic radius R and angular velocity $\Omega(r)$, the field has been amplified at time t by a factor of order $t d\Omega / d\ln r$ and the distance between field reversals is of order $R(t d\Omega / d\ln r)^{-1}$. The growth of the field is eventually halted by ion-neutral drift, also called ambipolar diffusion, perpendicular to the Galactic plane. The ambipolar drift velocity \mathbf{v}_D is the velocity at which acceleration of the plasma component by Lorentz forces is balanced by frictional deceleration due to elastic collisions with neutrals. The drift, which we describe more fully in Section 5.2, is the inevitable outcome of partial hydrostatic support of the ISM by magnetic pressure. It slowly but inexorably removes the field from the galaxy. Because the magnitude of \mathbf{v}_D is proportional to B^2 , there is a value of B_θ at which the rate of windup equals the rate of removal. The toroidal field saturates at this level.

With the 10^{10} year age adopted by Howard and Kulsrud for the Galactic disk, the toroidal field would now reverse on a scale of ~ 100 pc, far less than the spacing between reversals which is observed. If the initial magnetic field had been completely uniform, the residual mean toroidal field $\langle B_\theta \rangle$ would be too small to be consistent with the observed Faraday rotation of distant pulsars. But a smooth initial field gradient over the lengthscale of the Galactic disk would lead to a net field of one sign, despite the folding of each fieldline with radius. With the proper choice of initial gradient, the result would be a field of toroidal orientation, with the rms field several times larger than the mean field.

The model of Howard and Kulsrud is based on simple, well understood physics, and as such is worth serious consideration. In particular, it deserves thorough comparison with the data. This requires a somewhat more complete theory, one which accounts for the vertical structure of the ISM in a more realistic way and which also includes a turbulent component of magnetic field. With more development, one could also explore the extent to which the model must be fine-tuned. The initial fieldstrength gradient must be chosen so as to give just the proper ratio $\langle B_\theta \rangle / \langle B_\theta^2 \rangle^{1/2}$ at the present epoch. At the same time that differential rotation is amplifying B_θ , turbulence will amplify the random component of the field to near equipartition with the kinetic energy. It is not obvious why the equipartition random field should be close to

the saturated value of B_θ , which is set by completely unrelated physics, such as the ionization fraction in the ISM and the fraction of the volume filled with weakly ionized clouds. Thus, the ratios $\langle B_\theta \rangle / \langle B_\theta^2 \rangle^{1/2}$ and $\langle B^2 \rangle / \langle B \rangle^2$ depend on parameters which are undetermined in the theory. Finally, since the amplification of the toroidal field is algebraic rather than exponential, the initial field must be quite strong, of order 10^{-8}G for Howard and Kulsrud's choice of parameters.

Because it now appears that both mean field dynamo models and the simple primordial field model can produce magnetic fields which are qualitatively similar to the magnetic fields observed in disk galaxies, it appears impossible to rule out either theory on the basis of the current direct observational evidence. Perhaps the best argument that Galactic magnetic fields are sustained by dynamos rather than being primordial relics comes indirectly, from our current picture of Galactic evolution. The continual flux of material in and out of the interstellar medium means that the field has not evolved in pristine isolation. Rather, there must be mechanisms which assimilate the new material and compensate for the loss of old material so as to maintain a steady state. Thus, we turn to a discussion of the main processes which are relevant to the generation and evolution of Galactic magnetic fields. Some of the ingredients are standard, while others are quite novel, and not encountered in systems other than galaxies.

5.1. Differential rotation and turbulence

The shearing time $\Gamma^{-1} \equiv (d\Omega/d\ln r)^{-1}$ is of order $4 \times 10^7 (r/8.5\text{kpc})\text{years}$.¹ Thus, a radial spur of magnetic field introduced into the disk would be sheared such that $B_\theta/B_r \sim 25$ within 10^9 years, more than enough to agree with the observed orientation. The oscillation frequency of a ballistic particle in the Galactic potential well is known as the epicyclic frequency $\kappa \equiv 2\Omega[1 + d\ln \Omega/d\ln r^2]^{1/2}$. The parameter which controls the influence of Coriolis forces on an eddy of wavelength λ and velocity v_λ is $\kappa\lambda/2\pi v_\lambda$. Taking $v_\lambda \sim 10 \text{ km s}^{-1}$, this parameter is unity at $\lambda = 1.7 (r/8.5 \text{ kpc}) \text{ kpc}$. Even for the largest interstellar structures, the superbubbles, rotation is generally less than a 10% effect.

As we mentioned in Section 2, supernova explosions and expanding superbubbles are thought to be the primary energy source of interstellar turbulence, but they are not the only sources. Instabilities driven by

¹We scale the estimates in this section by 8.5 kpc, the galactocentric distance of the solar orbit.

thermal cooling, self gravity, and magnetorotational effects can also play a role (Sellwood and Balbus, 1999; Kritsuk and Norman, 2002; Kim *et al.*, 2003). The latter could be particularly important at large galactocentric radii, where there is little star formation and hence few supernovae.

The interstellar magnetic field is strong enough to affect the turbulent cascade process on small scales. In particular, the cascade is thought to be primarily Alfvénic in character, and anisotropic, with the components of κ perpendicular to \mathbf{B} much larger than the parallel component (Goldreich and Sridhar, 1997). How this affects the steady state operation of a dynamo is not yet clear.

5.2. Microscopic state of the gas

The heterogeneous state of the ISM, which we briefly described in Section 2 poses challenges in treating the evolution of Galactic magnetic fields. On small scales, the properties of turbulence and its interaction with magnetic fields are expected to vary considerably from one type of region to another. On the other hand, the largest observed magnetic coherence length exceeds the largest observed thermal structures in the ISM—hot ionized bubbles and cold molecular clouds—by about an order of magnitude. Similarly, the characteristic growth time for the large scale field, which is about the Galactic rotation period, exceeds the thermal cycling time by a factor of several. Therefore, the evolution at the largest scales should be calculated from appropriate averages over the effects of small scale fluctuations in different types of regions as well as ISM properties themselves. Howard and Kulsrud (1997) have done the latter, for example, in arguing that the ion-neutral drift rate is enhanced by the reciprocal of f , the filling factor of cold clouds in Galactic disks.

Within the ionized component, the ion collision time τ_i and gyrofrequency ω_i satisfy the condition²

$$\omega_i \tau_i \sim 1.7 \times 10^4 \frac{B T^{3/2}}{n_i} \gg 1. \quad (5.1)$$

Because $\omega_i \tau_i$ is large, viscous momentum transport perpendicular to \mathbf{B} is drastically reduced (by 1–2 powers of $\omega_i \tau_i$), with the viscous stress

²In this and other numerical formulae, we assume the ions are protons, set the Coulomb logarithm to 10, and take $T_e = T_i$. We express B in G, n in cm^{-3} , and T in degrees K, and all other quantities in subsequent formulae in cgs units.

tensor taking the form given by Braginskii (1965). This anisotropy has important consequences for small scale turbulence. For example, viscous damping of shear Alfvén waves is strongly suppressed by anisotropy effects, while compressive waves are damped at nearly the same rate they would be in an isotropic plasma.

We remarked elsewhere that the Lundquist number for the ISM is extremely large; now we quantify it. The magnetic diffusivity λ is

$$\lambda \sim 9.7 \times 10^{13} T^{-3/2} \text{cm}^2 \text{s}^{-1}. \quad (5.2)$$

The corresponding Lundquist number S is

$$S \equiv \frac{Lv_A}{\lambda} \sim 6.0 \times 10^9 \left(\frac{L}{1 \text{pc}} \right) \left(\frac{B}{\mu \text{G}} \right) T^{3/2}. \quad (5.3)$$

Even on scales of 1 pc, with $B = 5 \mu \text{G}$, $S \sim 3 \times 10^{16}$.

The magnetic Prandtl number P_{rm} — the ratio of viscosity to resistivity — is large in interstellar plasmas

$$P_{rm} \sim 2.8 \times 10^{-2} \frac{T^4}{n_i} \gg 1. \quad (5.4)$$

The result is that the turbulent kinetic energy spectrum is truncated at a much smaller wave number than the magnetic energy spectrum. This is generally true even accounting for the suppression of cross-field viscosity.

In weakly ionized gas, Lorentz forces tend to separate the plasma from the neutrals, while collisions tend to keep them together.³ On timescales long compared to the ion-neutral collision time τ_{in} , the drift velocity \mathbf{v}_D is set by balancing the Lorentz force on the plasma against the frictional drag by the neutrals

$$\mathbf{v}_D \equiv \mathbf{v}_i - \mathbf{v}_n = \tau_{in} \frac{\mathbf{J} \times \mathbf{B}}{\rho_i c}. \quad (5.5)$$

When equation (5.5) holds, the ideal form of Faraday's Law can be written in terms of \mathbf{v}_D and the center of mass velocity $\mathbf{v} \sim \mathbf{v}_n$ as

$$\frac{\partial \mathbf{B}}{\partial t} = \nabla \times (\mathbf{v} \times \mathbf{B}) + \nabla \times \left[\left(\tau_{in} \frac{\mathbf{J} \times \mathbf{B}}{\rho_i c} \right) \times \mathbf{B} \right] \quad (5.6)$$

³The plasma itself can be treated as a single fluid for most applications in the interstellar medium, since the ion inertial length $\delta_i \equiv v_A/\omega_i$ is of order $200 \text{ km}/\sqrt{n_i}$, far below other dynamical scales.

(e.g., Shu, 1983). The second term on the right-hand side represents ambipolar drift. In the restricted case $\mathbf{J} \cdot \mathbf{B} \equiv 0$, it has the same form as Ohmic diffusion, with the magnetic diffusivity replaced by the so-called ambipolar diffusivity λ_{AD} (Zweibel and Brandenburg, 1997)

$$\lambda_{AD} \equiv v_A^2 \tau_{ni} = 2.7 \times 10^{31} \frac{B^2}{n_n n_i} \text{ cm}^2 \text{ s}^{-1}. \quad (5.7)$$

Equations (5.6) and (5.7) motivate the introduction of an ambipolar Reynolds number $R_{AD} \equiv Lv/\lambda_{AD}$, and the associated decoupling length L_{AD} below which ions and neutrals are decoupled

$$L_{AD} \equiv \frac{\lambda_{AD}}{v} = 1.3 \times 10^{20} \frac{B}{M_A n_n n_i} \text{ cm}, \quad (5.8)$$

where $M_A \equiv v/v_A$ is the Alfvén Mach number. The decoupling length is typically rather small by interstellar standards; for example, if $B = 3 \mu\text{G}$, $n_n = 20$, $n_i = 0.02$, $M_A = 1$, $L_{AD} \sim 3 \times 10^{-4} \text{ pc}$. While this is a small scale, it is much larger than the resistive scale (compare (5.2) and (5.7)).

In weakly ionized regions, one must distinguish between behavior above and below the ion-neutral decoupling scale. At the scales on which the plasma and neutrals are well coupled, viscosity is dominated by the neutrals, and hence is isotropic. The magnetic Prandtl number, however, remains large. On small scales, the ions and neutrals are decoupled, and it is possible to have hydromagnetic fluctuations in which the neutrals do not participate except for the frictional damping of ion motion. At these small scales, hydromagnetic turbulence is similar to the turbulence in fully ionized regions, except that due to the low inertia of the plasma, the Alfvén Mach number is extremely small. One result is that on scales where ion-neutral drift is significant, the magnetic field tends to relax to a force free state (Brandenburg *et al.*, 1995). This may have implications for the saturation of a small scale dynamo, since the fluid cannot do work on a force free field.

As we have already mentioned, ion-neutral drift leads to slow removal of the magnetic field from the galactic disk — this can be seen from (5.5), together with the fact that magnetic fields contribute to hydrostatic support. The mechanism is virtually inoperative above a few hundred pc because of the paucity of neutral gas at these heights, but it may be aided by buoyancy.

It has recently been shown that ambipolar drift is enhanced by turbulence (Zweibel, 2002; Kim and Diamond, 2002; Heitsch *et al.*, 2004). By analogy with other mixing problems, the small scale magnetic

gradients which develop in a turbulent flow produce large local drifts, which cause turbulent diffusion of the mean field at a rate close to the eddy rate. There may be an observational signature of this process: the weak correlation observed between magnetic fieldstrength and gas density in the ISM, which we alluded to in Section 3. Turbulent ambipolar diffusion could accelerate the rate of escape of magnetic fields from Galactic disks, and the transport of magnetic field into new material in the ISM.

Finally, it can be seen from (5.7) that the diffusion associated with ambipolar drift is nonlinear, and, like other forms of nonlinear diffusion, can mediate the formation of sharp fronts. These fronts occur in the vicinity of magnetic neutral sheets, and can initiate rapid magnetic merging. The basic physical picture is that ions flow toward the neutral sheet (say, the plane $x=0$), dragging the field with them. In a steady state, an ideal, pressureless plasma develops a singular current density $J \propto x^{-2/3}$ (Brandenburg and Zweibel, 1994). The singularity is resolved by a combination of finite resistivity and ion pressure, but under interstellar conditions the merging is still very fast⁴ (Heitsch and Zweibel, 2003a). In particular, ion pressure, which would build steadily since the plasma flow stagnates at the neutral sheet, is limited by rapid recombination of the ions. This makes it easy to see that the conditions under which rapid merging occurs are very strict: any sheared or transverse component of magnetic field is swept toward the neutral line. Its pressure limits the thinning of the sheet and drastically slows the rate of magnetic merging (Heitsch and Zweibel, 2003b).

5.3. *Cosmic rays*

As we discussed in Section 2, galaxies are pervaded by a relativistic plasma known as cosmic rays. Because of their high velocities, cosmic rays propagate virtually collisionlessly through the ISM (although as we mentioned, their infrequent nuclear collisions with ambient particles allow us to unravel their history, while low energy cosmic rays are an important source of ionization). They are dynamically coupled to the gas, however, through their interactions with the magnetic field.

Cosmic rays are measured to be nearly isotropic, and are thought to propagate diffusively, with the scattering centers being some form of

⁴The merging rate depends on the magnetic diffusivity, and so is not formally considered “fast” reconnection, but it is quantitatively rapid.

magnetic irregularities. Over large scales, they can be described as a fluid with an isotropic pressure tensor but without inertia. They transfer momentum to the gas through their contribution to the total current in the plasma, and through the momentum they lose through scattering (Wentzel, 1974; Skilling, 1975). These large scale cosmic ray forces play a role in the escape of the Galactic magnetic field through buoyancy instabilities (Parker, 1966), and/or by driving a Galactic wind (Ipavich, 1975; Zirakashvili *et al.*, 1996).

On small scales cosmic rays interact with the magnetic field through kinetic effects. If the cosmic ray drift anisotropy exceeds the Alfvén speed, cosmic rays destabilize Alfvén waves with a wavelength of order the cosmic ray gyroradius (Kulsrud and Pearce, 1969). At the present epoch, this is a small scale — about 2×10^{-6} pc for waves driven by cosmic rays near the mean energy — and the fluctuation amplitude is small; $\delta B / B \sim 10^{-3} - 10^{-4}$. Near strong shocks driven by supernova explosions, where Galactic cosmic rays are thought to be accelerated, the driving is stronger, and may significantly amplify the magnetic field (Lucek and Bell, 2000). But this mechanism may be particularly significant in young galaxies in which the magnetic field may be much less than its present value and the supernova rate was probably significantly higher. Under these conditions, the gyroradius is much larger, the instability grows much faster, and there is no significant source of linear damping (Zweibel, 2003). In view of the strong amplification of the field at the tiny resistive scale predicted by current calculations, any mechanism which drives fluctuations at a larger scale may be of interest.

5.4. Magnetic reconnection

Magnetic reconnection plays a crucial role in the operation of a dynamo and in the assimilation of the small scale magnetic fields introduced by stellar and extragalactic sources to the ISM. It is also one of the most problematic areas of study, because the high Lundquist number of interstellar gas requires that resistive phenomena can occur only on tiny scales compared to the global scale. The plasma and embedded magnetic field brought into the thin reconnection region must somehow be expelled. Since the velocity of expulsion is limited by the energy available to approximately the Alfvén speed, the efflux is small. In a steady state, the rate of reconnection is therefore slow.

As we discussed in Section 5, it has been shown that rapid merging occurs in magnetic neutral sheets when the ambient gas is weakly ionized. In this case, the continuity problem is solved by plasma

recombination. However, because these structures must be virtually perfect neutral sheets, with almost no sheared or transverse magnetic field components, this mechanism is unlikely to be a universal panacea for the reconnection problem. In fact, we know that in the opposite limit of a strongly sheared magnetic field, resistive tearing mode growth rates have the same dependence on resistivity that they have in a fully ionized plasma (Zweibel, 1989).

Two effects which are known to increase the reconnection rate in space and laboratory plasmas are anomalous resistivity and the Hall effect. Anomalous resistivity occurs when the electron drift velocity associated with the current flowing in the reconnection region exceeds a threshold for a kinetic instability. The current is then an energy source for fluctuations, which scatter the electrons and effectively increase the resistivity. In addition to reducing the Lundquist number, the resulting variable space dependence of the resistivity can permit a different and much faster mode of MHD reconnection called Petschek reconnection (Kulsrud, 1998). It is unlikely that anomalous resistivity can operate in the ISM, however, because the high current densities required are incompatible with the long lengthscales of interstellar magnetic fields. For example, so-called ion-acoustic waves are unstable when the current drift J/en_e exceeds the ion thermal speed $(k_B T / m_i)^{1/2}$. This requires that the plasma β , thermal ion gyroradius r_i , and magnetic scalelength L satisfy the inequality $\beta^{-1} r_i / L \geq 1$. Since $\beta \sim 1$, the magnetic field must be structured down to the scale of the ion gyroradius, which is only $\sim 10^3$ km.

In Hall, or Whistler mediated reconnection, the electrons and ions decouple, leading to a wider outflow channel and hence a faster reconnection rate (Mandt *et al.*, 1994; see Biskamp, 2000 for a pedagogical treatment). The potential importance of this type of reconnection in the ISM depends on how the rate scales with a parameter similar to that identified above in the criterion for anomalous resistivity, $\beta^{-1} r_i / L$. In order for Whistler reconnection to be important, the reconnection rate must be virtually independent of this ratio. This is presently a matter of some controversy, with arguments and numerical evidence for independence (Shay *et al.*, 2004) but also for a linear scaling (Fitzpatrick, 2004).

It is possible that the resolution of the reconnection problem lies in turbulence. Lazarian and Vishniac (1999) have argued that if anisotropic MHD turbulence exists down to the resistive scale, the reconnection speed can be as large as v_A itself. In their model, the reconnection rate is enhanced because there are many reconnection sites and because the outflow from each site follows a chaotic and hence highly divergent trajectory.

6. Summary and outlook

Magnetic fields appear to be universally present in disk galaxies. All the mechanisms proposed for their origin require that they be amplified, and that a coherent field be developed or preserved, through hydro-magnetic processes. Because the interstellar medium is continuously replenished and removed on timescales of order 10^9 years, the mechanisms which configure Galactic magnetic fields must operate on these timescales in order to maintain a steady state.

Neither dynamo theory as applied to galaxies nor the primordial theory fully accounts for the cycling of matter through the ISM. Bearing that in mind, it seems that neither theory is ruled out on observational grounds, and that both would benefit from more development. The standard kinematic mean field dynamo theory is incomplete for reasons stated more fully elsewhere in this volume: the breakdown of the kinematic approximation and the problem of saturation in a medium with high Lundquist number. The primordial theory appears to require some tuning of parameters to obtain agreement with observations.

In our view, the nature of magnetic field diffusion is the major unsolved problem facing theories of the origin and maintenance of Galactic magnetic fields. Diffusion, as we understand it here, is of two types. One is the breakdown of the frozen flux constraint in the *plasma* component of the medium, and is required in order to change the magnetic topology. This type of diffusion is necessary for the operation of a dynamo, and, more generally, to prevent the buildup of small scale field to the point that it overwhelms the large scale field. It may also contribute to removal of the field from the halo, by breaking the connection to the Galactic disk. Magnetic reconnection is an example of this type of diffusion. In Section 5.4 we briefly reviewed current theories of reconnection as they might operate in the ISM, and emphasized the need for a fast reconnection mechanism.

The second type of diffusion, known as ambipolar diffusion, occurs in the weakly ionized portions of the medium, and arises through ion-neutral drift (see Section 5.2). It preserves magnetic topology, but transports the field with respect to the center of mass. In a laminar medium, ambipolar diffusion is rather slow, except in the most strongly magnetized and weakly ionized interstellar gas (see (5.7)). However, in a turbulent medium, ambipolar diffusion provides the small scale drifts which lead to mixing of the field on the eddy timescale. This second type of diffusion may be important in magnetically assimilating new gas as it is introduced to the ISM, and in the vertical transport of field out of the Galaxy. These two types of diffusion are connected through

the tendency for ambipolar drift to steepen magnetic neutral sheets into current sheets. However, the significance of this process is open to question because it has been demonstrated only under rather restricted conditions.

There must also be a mechanism for producing a component of field on scales much larger than the current turbulent forcing scale. Whether this large scale field results from a classical inverse cascade, as might be expected in a differentially rotating system with helical turbulence, or from a large scale intergalactic field is still uncertain.

In addition to these theoretical issues, there are some impending observational developments that should influence the theory of Galactic magnetic fields. The planned deployment of a low frequency radio telescope array will make it possible to detect, or place meaningful upper limits on, magnetic fields in very young galaxies. A new survey of pulsar Faraday rotation measurements will enable mapping of the field in the Milky Way with improved accuracy. The prospective Square Kilometer Array would provide a new tool of unprecedented resolution and sensitivity for mapping magnetic fields in other galaxies.

Acknowledgments

The author is grateful for support provided by NSF awards AST-0328821 and PHY-0215581 and by the Graduate School of the University of Wisconsin, Madison.

Appendix: background reading

For an excellent introduction to Galactic astronomy and Galactic dynamics, see Binney and Merrifield (1998) and Binney and Tremaine (1987). The classic reference on the interstellar medium is Spitzer (1978); see also the newer book by Dopita and Sutherland (2003) and recent review by Ferrière (2001). Cosmic ray astrophysics is discussed in the comprehensive work by Schlickeiser (2002). The properties of Galactic magnetic fields, and theories of their origin, are discussed in the review articles by Beck *et al.* (1996), Zweibel and Heiles (1997), Kulsrud (1999), Beck (2001), and Han and Wielebinski (2002). For a cosmological slant on the topic of magnetic fields, see Grasso and Rubenstein (2001) or Widrow (2002).

Cosmology is in a stage of rapid development, and the references provided here may soon need revision. Nevertheless we recommend the

books by Liddle and Lyth (2000) and Dodelson (2003), and the review article by Peebles and Ratra (2003), for background in this field. Similarly, current theories of Galaxy formation are not yet encoded in a standard text. One can extract the flavor of the field from recent conference proceedings, such as the volume edited by Umemura and Susa (2001).

References

- André, M. K., Oliveira, C. M., Howk, J. C., Ferlet, R., Désert, J.-M., Hébrard, G., Lacour, S., des Étangs, A., Lecavelier, O., Vidal-Madjar, A. and Moos, H. W., *Astrophys. J.* **591**, 1000–1012 (2003).
- Beck, R., “Galactic and extragalactic magnetic fields,” *Sp. Sci. Rev.* **99**, 243–260 (2001).
- Beck, R., Brandenburg, A., Moss, D., Shukurov, A. and Sokoloff, D., “Galactic magnetism: recent developments and perspectives,” *Annu. Rev. Astron. Astrophys.* **34**, 155–206 (1996).
- Binney, J. and Merrifield, M., *Galactic Astronomy*, Princeton University Press, Princeton (1998).
- Binney, J. and Tremaine, S., *Galactic Dynamics*, Princeton University Press, Princeton (1987).
- Biskamp, D., *Magnetic Reconnection in Plasmas*, Cambridge University Press, Cambridge (2000).
- Blasi, P., Burles, S. and Olinto, A. V., “Cosmological magnetic field limits in an inhomogeneous universe,” *Astrophys. J.* **514**, L79–L82 (1999).
- Braginskii, S. I., “Transport processes in a plasma,” *Rev. Plasma Phys.* **1**, 205–311 (1965).
- Brandenburg, A. and Zweibel, E. G., “The formation of sharp structures by ambipolar diffusion,” *Astrophys. J.* **427**, L91–L94 (1994).
- Brandenburg, A., Nordlund, A., Stein, R. F. and Torkelsson, U., “Dynamo-generated turbulence and large-scale magnetic fields in a Keplerian shear flow,” *Astrophys. J.* **446**, 741–754 (1995).
- Brown, J. C. and Taylor, A. R., “The structure of the magnetic field in the Outer Galaxy from rotation measure observations through the disk,” *Astrophys. J.* **563**, L31–L34 (2001).
- Chandran, B. D. G., Cowley, S. C. and Morris, M., “Magnetic flux accumulation at the Galactic Center and its implications for the strength of the pregalactic magnetic field,” *Astrophys. J.* **528**, 723–733 (2000).
- Clarke, T. E., Kronberg, P. P. and Böhringer, H., “A new radio–x-ray probe of galaxy cluster magnetic fields,” *Astrophys. J.* **547**, L111–L114 (2001).
- Clegg, A. W., Cordes, J. M., Simonetti, J. M. and Kulkarni, S. R., “Rotation measures of low-latitude extragalactic sources and the magnetoionic structure of the Galaxy,” *Astrophys. J.* **386**, 143–157 (1992).
- Cohen, D. H., de Messiéres, G. E., MacFarlane, J. J., Miller, N. A., Cassinelli, J. P., Owocki, S. P. and Liedahl, D. A., “High resolution Chandra spectroscopy

- of τ Scorpii: a narrow-line x-ray spectrum from a hot star," *Astrophys. J.* **586**, 495–505 (2003).
- Comeron, F. and Torra, J., "The origin of the Gould Belt by the impact of a high velocity cloud on the Galactic disk," *Astron. Astrophys.* **281**, 35–45 (1994).
- Crutcher, R., "Magnetic fields in molecular clouds: observations confront theory," *Astrophys. J.* **520**, 706–713 (1999).
- Daly, R. A. and Loeb, A., "A possible origin of galactic magnetic fields," *Astrophys. J.* **364**, 451–455 (1990).
- Dodelson, S., *Modern Cosmology*, Academic Press, New York (2003).
- Dopita, M. A. and Sutherland, R. S., *Astrophysics of the Diffuse Universe*, Springer, Berlin (2003).
- Draine, B. T., "Interstellar dust grains," *Annu. Rev. Astron. Astrophys.* **41**, 241–289 (2003).
- Duncan, D. K., Lambert, D. L. and Lemke, M., "The abundance of boron in three halo stars," *Astrophys. J.* **403**, 584–595 (1992).
- Ferrière, K. M., "Alpha-tensor and diffusivity tensor due to supernovae and superbubbles in the Galactic disk near the Sun," *Astron. Astrophys.* **310**, 438–455 (1996).
- Ferrière, K. M., "The interstellar environment of our galaxy," *Rev. Mod. Phys.* **73**, 1031–1066 (2001).
- Ferrière, K. M. and Schmitt, D., "Numerical models of the galactic dynamo driven by supernovae and superbubbles," *Astron. Astrophys.* **358**, 125–143 (2000).
- Fitzpatrick, R., "Scaling of forced magnetic reconnection in the Hall-magnetohydrodynamic Taylor problem," submitted to *Phys. Plasmas* (2004).
- Furlanetto, S. R. and Loeb, A., "Intergalactic magnetic fields from quasar outflows," *Astrophys. J.* **556**, 619–634 (2001).
- Gilmore, G., Wyse, R. F. G. and Norris, J. E., "Deciphering the last major invasion of the Milky Way," *Astrophys. J.* **574**, L39–L42 (2002).
- Gilmore, G., Gustafsson, B., Edvardsson, B. and Nissen, P. E., "Is beryllium in metal-poor stars of galactic or cosmological origin?," *Nature* **357**, 379–384 (1992).
- Gnedin, N. Y., Ferrara, A. and Zweibel, E. G., "Generation of the primordial magnetic fields during cosmological reionization," *Astrophys. J.* **539**, 505–516 (2000).
- Goldreich, P. and Kylafis, N. D., "Linear polarization in radio frequency lines in molecular clouds and circumstellar envelopes," *Astrophys. J.* **253**, 606–621 (1982).
- Goldreich, P. M. and Sridhar, S., "Magnetohydrodynamic turbulence revisited," *Astrophys. J.* **485**, 680–688 (1997).
- Grasso, D. and Rubenstein, H. R., "Magnetic fields in the early universe," *Phys. Reports* **348**, 163–266 (2001).
- Han, J.-L., "Magnetic fields in our galaxy: how much do we know?," *Astrophys. Sp. Sci.* **278**, 181–184 (2001).
- Han, J.-L. and Wielebinski, R., "Milestones in the observations of cosmic magnetic fields," *Chin. J. Astron. Astrophys.* **2**, 293–324 (2002).

- Han, J. L., Manchester, R. N. and Qiao, G. J., "Pulsar rotation measures and the magnetic structure of our Galaxy," *Mon. Not. R. Astron. Soc.* **306**, 371–380 (1999).
- Hanasz, M. and Lesch, H., "The galactic dynamo effect due to Parker-shearing instability of magnetic flux tubes: I. general formulation and the linear approximation," *Astron. Astrophys.* **321**, 1007–1020 (1997).
- Haverkorn, M., Katgert, P. and de Bruyn, A. G., "Characteristics of the structure in the Galactic polarized radio background at 350 MHz," *Astron. Astrophys.* **403**, 1045–1057 (2003).
- Heiles, C., "A comprehensive view of the Galactic magnetic field, especially near the Sun," in: *Polarimetry of the Interstellar Medium* (Eds. W. G. Roberge and D. C. B. Whittet), pp. 457–474 Astronomical Society of the Pacific (1996).
- Heitsch, F. and Zweibel, E. G., "Fast reconnection in a two-stage process," *Astrophys. J.* **583**, 229–244 (2003a).
- Heitsch, F. and Zweibel, E. G., "Suppression of fast reconnection by magnetic shear," *Astrophys. J.* **590**, 291–295 (2003b).
- Heitsch, F., Zweibel, E. G., Slyz, A. and Devriendt, J. E. G., "Turbulent ambipolar diffusion: numerical studies in 2D," *Astrophys. J.* in press (2004).
- Hildebrand, R. H., Davidson, J. A., Dotson, J. L., Dowell, C. D., Novak, G. and Vaillancourt, J. E., "A primer in far-infrared polarimetry," *Publ. Astron. Soc. Pacific* **112**, 1215–1235 (2000).
- Howard, A. M. and Kulsrud, R. M., "The evolution of a primordial galactic magnetic field," *Astrophys. J.* **483**, 648–665 (1997).
- Ipavich, F. M., "Galactic winds driven by cosmic rays," *Astrophys. J.* **196**, 107–120 (1975).
- Jokipii, J. R., "Rate of energy gain and maximum energy in diffusive shock acceleration," *Astrophys. J.* **313**, 842–846 (1987).
- Kim, E.-J. and Diamond, P. H., "Turbulent diffusion of magnetic fields in weakly ionized gas," *Astrophys. J.* **578**, L113–L116 (2002).
- Kim, W.-T., Ostriker, E. C. and Stone, J. M., "Magnetorotationally driven galactic turbulence and the formation of giant molecular clouds," *Astrophys. J.* **599**, 1157–1172 (2003).
- Ko, C. M. and Parker, E. N., "Intermittent behavior of galactic dynamo activities," *Astrophys. J.* **341**, 828–831 (1989).
- Kritsuk, A. G. and Norman, M. L., "Interstellar phase transitions stimulated by time-dependent heating," *Astrophys. J.* **580**, L51–L55 (2002).
- Kronberg, P. P. and Perry, J. J., "Absorption lines, Faraday rotation, and magnetic field estimates for QSO absorption-line clouds," *Astrophys. J.* **263**, 518–532 (1982).
- Kronberg, P. P., Dufton, Q. W., Li, H. and Colgate, S. A., "Magnetic energy of the intergalactic medium from galactic black holes," *Astrophys. J.* **560**, 178–186 (2001).
- Kulsrud, R. M., "Magnetic reconnection in a magnetohydrodynamic plasma," *Phys. Plasmas* **5**, 1599–1606 (1998).
- Kulsrud, R. M., "A critical review of galactic dynamos," *Annu. Rev. Astron. Astrophys.* **37**, 37–64 (1999).

- Kulsrud, R. M. and Anderson, S. W., "The spectrum of random magnetic fields in the mean field dynamo theory of the Galactic magnetic field," *Astrophys. J.* **396**, 606–630 (1992).
- Kulsrud, R. M. and Pearce, W. P., "The effect of wave-particle interactions on the propagation of cosmic rays," *Astrophys. J.* **156**, 445–469 (1969).
- Kulsrud, R. M., Cen, Y., Ostriker, J. P. and Ryu, D., "The protogalactic origin for cosmic magnetic fields," *Astrophys. J.* **480**, 481–491 (1997).
- Lai, S.-P., Girart, J. M. and Crutcher, R. M., "Interferometric mapping of magnetic fields in star-forming regions. III. Dust and CO polarization in DR 21(OH)," *Astrophys. J.* **598**, 392 (2003).
- Lazarian, A. and Vishniac, E. T., "Reconnection in a weakly stochastic field," *Astrophys. J.* **517**, 700–718 (1999).
- Liddle, A. R. and Lyth, D. H., *Cosmological Inflation and Large Scale Structure*, Cambridge University Press, Cambridge (2000).
- Lucek, S. G. and Bell, A. R., "Nonlinear amplification of a magnetic field driven by cosmic ray streaming," *Mon. Not. R. Astron. Soc.* **314**, 65–74 (2000).
- Madau, P., "The era of reionization," in: *Galaxy Evolution: Theory and Observations* (Eds V. Avila-Rees, C. Firmani, C. S. Frenk and C. Allen), pp. 287–296, Revista Mexicana de Astronomía y Astronomía (Serie de Conferencias) **17** (2003).
- Mandt, M. E., Denton, R. E. and Drake, J. F., "Transition to Whistler mediated reconnection," *Geophys. Res. Lett.* **21**, 73–76 (1994).
- Mestel, L. and Spitzer, L., "Star formation in magnetic dust clouds," *Mon. Not. R. Astron. Soc.* **116**, 503–514 (1956).
- Meyer, D. M., Jura, M. and Cardelli, J. A., "The definitive abundance of interstellar oxygen," *Astrophys. J.* **493**, 222–229 (1998).
- Minter, A. H. and Spangler, S. R., "Observation of turbulent fluctuations in the interstellar plasma density and magnetic field on spatial scales of 0.01 to 100 parsecs," *Astrophys. J.* **458** 194–214 (1994).
- Morris, M. and Serabyn, E., "The Galactic Center environment," *Annu. Rev. Astron. Astrophys.* **34**, 645–702 (1996).
- Moss, D., Shukurov, A., Sokoloff, D., Beck, R. and Fletcher, A., "Magnetic fields in barred galaxies. II. Dynamo models," *Astron. Astrophys.* **380**, 55–71 (2001).
- Oren, A. L. and Wolfe, A. M., "A Faraday rotation search for magnetic fields in quasar damped Ly α systems," *Astrophys. J.* **445**, 624–641 (1995).
- Padoan, P., Goodman, A., Draine, B. T., Juvela, M., Nordlund, A. and Rögnvaldsson, Ö. E., "Theoretical models of polarized dust emission from protostellar cores," *Astrophys. J.* **559**, 1005–1018 (2001).
- Parker, E. N., "The dynamical state of the interstellar gas and field," *Astrophys. J.* **145**, 811–833 (1966).
- Peebles, P. J. E. and Ratra, B., "The cosmological constant and dark energy," *Rev. Mod. Phys.* **75**, 559–606 (2003).
- Rafikov, R. R. and Kulsrud, R. M., "Magnetic flux expulsion in powerful superbubble explosions and the $\alpha - \Omega$ dynamo," *Mon. Not. R. Astron. Soc.* **314**, 839–848 (2000).

- Rand, R. J. and Lyne, A. G., "New rotation measures of distant pulsars in the inner Galaxy and magnetic field reversals," *Mon. Not. R. Astron. Soc.* **268**, 497–505 (1994).
- Rees, M. J., "Origin of the seed magnetic field for a galactic dynamo," in: *Cosmical Magnetism* (Ed. D. Lynden-Bell), pp. 155–160, Kluwer, Dordrecht (1994).
- Richter, P., Sembach, K. R., Wakker, B. P., Savage, B. D., Tripp, T. M., Murphy, E. M., Kalberla, P. M. W. and Jenkins, E. B., "The diversity of high- and intermediate-velocity clouds: Complex C versus IV Arch," *Astrophys. J.* **559**, 318–325 (2001).
- Schekochihin, A. A., Maron, J. L., Cowley, S. C. and McWilliams, J. C., "The small-scale structure of magnetohydrodynamic turbulence with large magnetic Prandtl numbers," *Astrophys. J.* **576**, 806–813 (2002).
- Schlickeiser, R., *Cosmic Ray Astrophysics*, Springer, Berlin (2002).
- Sellwood, J. A. and Balbus, S. A., Differential rotation and turbulence in extended H I disks, *Astrophys. J.* **511**, 660–665 (1999).
- Serkowski, K., "Polarization of galactic clusters M25, NGC 869, 1893, 2422, 6823, 6871 and association VI Cygni," *Astrophys. J.* **141**, 1340–1361 (1965).
- Shay, M. A., Drake, J. F., Swisdak, M. and Rogers, B. N., "The scaling of embedded collisionless reconnection," submitted to *Phys. Plasmas* (2004).
- Shu, F. H., "Ambipolar diffusion in self-gravitating isothermal layers," *Astrophys. J.* **273**, 202–213 (1983).
- Silk, J. and Wyse, R. F. G., "Galaxy formation and the Hubble sequence," *Phys. Rep.* **231**, 295–365 (1993).
- Skilling, J., "Cosmic ray streaming. III. Self-consistent solutions," *Mon. Not. R. Astron. Soc.* **173**, 255–269 (1975).
- Spergel, D. N. and 16 coauthors, "First year Wilkinson Microwave Anisotropy Probe (WMAP) observations: determination of cosmological parameters," *Astrophys. J. Suppl.*, **148**, 175–194 (2003).
- Spitzer, L., *Physical Processes in the Interstellar Medium*, Wiley-Interscience, New York (1978).
- Subrahmanian, K., Narasimha, D. and Chitre, S. M., "Thermal generation of cosmological seed magnetic fields in ionization fronts," *Mon. Not. R. Astron. Soc.* **271**, L15–L18 (1994).
- Tripp, T. M., Wakker, B. P., Jenkins, E. B., Bowers, C. W., Danks, A. C., Green, R. F., Heap, S. R., Joseph, C. L., Kaiser, M. E., Linsky, J. L. and Woodgate, B. E., "Complex C: a low metallicity, high velocity cloud plunging into the Milky Way," *Astron. J.* **125**, 3122–3144 (2003).
- Troland, T. H. and Heiles, C., "Interstellar magnetic fieldstrengths and gas densities: observational and theoretical perspectives," *Astrophys. J.* **301** 339–345 (1986).
- Umemura, M. and Susa, H., *The Physics of Galaxy Formation*, Astronomical Society of the Pacific, San Francisco (2001).
- Wentzel, D. G., "Cosmic ray propagation in the Galaxy: Collective effects," *Annu. Rev. Astron. Astrophys.* **12**, 71–96 (1974).

- Widrow, L., "Origin of galactic and extragalactic magnetic fields," *Rev. Mod. Phys.* **74**, 775–823 (2002).
- Zirakashvili, V. N., Breitschwerdt, D., Ptuskin, V. S. and Völk, H. J., "Magnetohydrodynamic wind driven by cosmic rays in a rotating galaxy," *Astron. Astrophys.* **311**, 113–126 (1996).
- Zweibel, E. G., "Magnetic reconnection in partially ionized gases," *Astrophys. J.* **340**, 550–557 (1989).
- Zweibel, E. G., "Ambipolar drift in a turbulent medium," *Astrophys. J.* **567**, 962–970 (2002).
- Zweibel, E. G., "Cosmic ray history and its implications for galactic magnetic fields," *Astrophys. J.* **587**, 625–637 (2003).
- Zweibel, E. G. and Brandenburg, A., "Current sheet formation in the interstellar medium," *Astrophys. J.* **478**, 563–568 (1997).
- Zweibel, E. G. and Heiles, C., "Magnetic fields in galaxies and beyond," *Nature* **385**, 131–136 (1997).

6 Self-consistent mean field electrodynamics in two and three dimensions

P.H. Diamond

*Department of Physics, University of California, San Diego,
La Jolla, CA 92093-0319, USA*

D.W. Hughes

*Department of Applied Mathematics, University of Leeds,
Woodhouse Lane, Leeds LS2 9JT, UK*

Eun-jin Kim

*Department of Physics, University of California, San Diego,
La Jolla, CA 92093-0319, USA*

The status of the self-consistent theory of mean field electrodynamics for incompressible MHD is reviewed. This discussion treats the calculation of the transport of magnetic potential or flux in two and three dimensions, the α -effect in three dimensions, and the transport of momentum in two dimensions. Physical interpretations and connections to numerical experiments are emphasized.

1. Introduction

Mean field electrodynamics is concerned with the application of the techniques of mean field or quasi-linear theory to the derivation of local, turbulent transport equations for macroscopic quantities (such as the average magnetic field ($\langle \mathbf{B} \rangle$) in MHD and other plasma models. The most well-known products of mean field electrodynamics are the α and

β coefficients for the evolution of $\langle \mathbf{B} \rangle$ in 3D, i.e.,

$$\frac{\partial}{\partial t} \langle \mathbf{B} \rangle = \nabla \times \alpha \langle \mathbf{B} \rangle + (\eta + \beta) \nabla^2 \langle \mathbf{B} \rangle, \quad (1)$$

and the turbulent resistivity η_T in 2D, i.e.,

$$\frac{\partial}{\partial t} \langle A \rangle(x, t) = \eta_T \frac{\partial^2}{\partial x^2} \langle A \rangle, \quad (2)$$

where A is the magnetic potential. (In (1) and (2) we have considered the simplest case, in which the underlying turbulence is assumed to be homogeneous and isotropic, and for which α , β and η_T are constant.) Of course, α is the familiar “dynamo coefficient”, which aims to capture, in a local transport coefficient, the fundamental process of amplification of field by cyclonic turbulence. Here α is a pseudo-scalar and nearly always depends on the turbulence helicity; β and η_T typically depend on the turbulence energy, but may involve other quantities. In addition, α , β and η_T each involve a field-fluid correlation time.

The practice of mean field electrodynamics is a well-developed art form, the fundamentals of which are set forth in the classic monograph by Moffatt (1978). Until recently, mean field electrodynamics calculations were kinematic in character, and the fluid energy spectrum and, more subtly, the field-fluid correlation time, were taken as given. The effects of the small-scale magnetic field on either the transport coefficients or on the correlation time were almost always neglected. An important exception to this trend was the work of Pouquet *et al.* (1976, 1978). However, in a seminal paper, Cattaneo and Vainshtein (1991) convincingly suggested that for high R_m (the magnetic Reynolds number), small-scale magnetic field effects on mean field evolution are *not* negligible even for very weak values of the large-scale field, and that, consequently, diffusion of the mean flux is strongly reduced, or “quenched”, in comparison to its kinematic value. The quench reflects the presence of a “dynamical memory” on the part of the magnetic flux with respect to its location relative to the fluid. This type of analysis was rapidly extended to the α -effect in 3D incompressible MHD and various other systems. As the results of these investigations shook the foundations of the prevailing concepts of magnetic dynamo and field diffusion processes at high R_m , they quite naturally also engendered significant controversy.

It is the purpose of this paper to present a review of the theory of self-consistent mean field electrodynamics at high R_m — the regime of interest for astrophysical plasmas. There is no pretense of completeness, the “last word”, or “neutrality” in this review. Rather, we seek to

set forth our own understanding of the current status of the field, the principal ideas and results, and the main unresolved physics issues. We strive throughout to elucidate the basic physics and to make connections with relevant computational studies. In this paper, we discuss the application of mean field electrodynamics to the turbulent diffusion of magnetic fields in two and three dimensions, to the α -effect (necessarily in 3D) and to momentum transport and the effect of flow shear in 2D. In all cases, we assume a mean magnetic field (in most cases weak), which is distorted and stretched by turbulent motions. This distortion is the origin of the small-scale magnetic field, which grows rapidly in comparison to the timescale upon which the mean field evolves. The evolution of the mean field (either via diffusion or cyclonic distortion) is then considered in the presence of *both* small-scale fluid velocity and magnetic perturbations. Hence, we refer to these calculations as examples of *self-consistent* mean field electrodynamics.

The remainder of this review is organized as follows. Section 2 presents the theory of magnetic flux and field diffusion (i.e., the calculation of η_T and β) in 2D and 3D. We demonstrate that treating the small-scale magnetic turbulence on a footing equal to that of the velocity fluctuations, along with the constraint of mean square magnetic potential conservation, together imply that magnetic diffusion is quenched in comparison to kinematic predictions. We demonstrate that this quench is critically dependent on the magnetic Reynolds number. New results on the effects of boundaries and scale-to-scale coupling of magnetic potential are presented. Magnetic diffusion in reduced and full 3D MHD is also discussed. Section 3 discusses α -quenching in 3D MHD. The basic theory of the α -effect is reviewed. The incidence of the quench is related to the combined effects of the “back- α ” effect of small-scale magnetic fields and the conservation of magnetic helicity. The implications of related computational studies are discussed and assessed. We also address some of the contradictory claims and controversy surrounding this topic. Section 4 presents the mean field theory of momentum transport in 2D and the effect of mean shear flow on flux diffusion in 2D. Scalings of the effective turbulent viscosity and resistivity are derived for the strong shear and strong field limits. The crucial role of “Alfvénization” is identified. The relevance of these calculations to the mechanism of the interface dynamo at the boundary of the solar convection zone and tachocline is discussed. Although effects such as rotation and compressibility are obviously of importance in astrophysical contexts, many of the fundamental issues of turbulent transport can be considered within the framework of incompressible MHD, and hence we limit our discussion to this case throughout.

2. Turbulent diffusion of magnetic fields

2.1. Overview

In this section, we review the status of the mean field theory of diffusion of magnetic fields in 2D and 3D incompressible MHD. Attention is focused primarily on the simpler 2D problem, for which the effects of diffusion are not entangled with those of field growth through dynamo action. Also, and perhaps surprisingly, there are many close analogies between mean flux diffusion in 2D and that of the mean field α -effect in 3D. We discuss the similarities and differences between the two problems, with the goal of developing insight into the more interesting (but difficult!) 3D α -effect problem from the simpler 2D diffusion problem. Throughout our discussion, magnetic Prandtl number (P_m) of unity (i.e., $\nu = \eta$) and periodic boundary conditions are assumed, unless otherwise explicitly noted, although it is worth mentioning that interesting questions arise as to the resulting behavior when either of these assumptions are relaxed.

2.2. Flux diffusion in 2D—basic model and concepts

The familiar equations of 2D MHD are

$$\frac{\partial A}{\partial t} + (\nabla \psi \times \hat{\mathbf{z}}) \cdot \nabla A = \eta \nabla^2 A, \quad (3)$$

$$\frac{\partial}{\partial t} \nabla^2 \psi + (\nabla \psi \times \hat{\mathbf{z}}) \cdot \nabla \nabla^2 \psi = (\nabla A \times \hat{\mathbf{z}}) \cdot \nabla \nabla^2 A + \nu \nabla^2 \nabla^2 \psi, \quad (4)$$

where A is the magnetic potential ($\mathbf{B} = \nabla \times A \hat{\mathbf{z}}$), ψ is the velocity stream function ($\mathbf{v} = \nabla \times \psi \hat{\mathbf{z}}$), η is the resistivity, ν is the viscosity and $\hat{\mathbf{z}}$ is the unit vector orthogonal to the plane of motion. We shall consider the case where the mean magnetic field is in the y -direction, and is a slowly varying function of x . Equations (3) and (4) have non-dissipative quadratic invariants, the energy $E = \int [(\nabla A)^2 + (\nabla \psi)^2] d^2x$, mean-square magnetic potential $H_A = \int A^2 d^2x$ and cross helicity $H_c = \int \nabla A \cdot \nabla \psi d^2x$. Throughout this paper, we take $H_c = 0$ *ab initio*, so there is no net Alfvénic alignment in the MHD turbulence considered here. The effects of cross helicity on MHD turbulence are discussed by Grappin *et al.* (1983).

The basic dynamics of 2D MHD turbulence are well understood. For large-scale stirring, energy is self-similarly transferred to small scales and eventual dissipation via an Alfvénized cascade, as originally suggested by Kraichnan (1965) and Iroshnikov (1964), and clearly

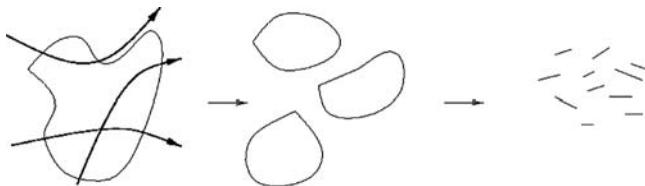


Figure 1. Forward transfer: fluid eddies chop up scalar A .

demonstrated in simulations by Biskamp and Welter (1989). The Kraichnan-Iroshnikov spectrum for the MHD turbulence cascade is the same in 2D as in 3D. This cascade may manifest anisotropy in the presence of a strong mean field in 3D, as predicted by Goldreich and Sridhar (1995, 1997). Mean square magnetic potential H_A , on the other hand, tends to accumulate at (or cascade toward) large scales, as is easily demonstrated by equilibrium statistical mechanics for non-dissipative 2D MHD (Fyfe and Montgomery, 1976). Here, H_c is the second conserved quadratic quantity (in addition to energy), which thus suggests a dual cascade. In 2D, the mean field quantity of interest is the spatial flux of magnetic potential $\Gamma_A = \langle v_x A \rangle$. An essential element of the physics of Γ_A is the competition between advection of scalar potential by the fluid, and the tendency of the flux A to coalesce at large scales. The former is, in the absence of back-reaction, simply a manifestation of the fact that turbulence tends to strain, mix and otherwise "chop up" a passive scalar field, thus generating small-scale structure (see Fig. 1). The latter manifests the fact that A is *not* a passive scalar, and that it resists mixing by the tendency to coagulate on large scales (see Fig. 2) (Riyopoulos *et al.*, 1982). The inverse cascade of A^2 , like the phenomenon of magnetic island coalescence, is ultimately rooted in the fact that like-signed current filaments attract. Not surprisingly then, the velocity field drives a positive potential diffusivity (turbulent resistivity), while the magnetic field perturbations drive a *negative* potential diffusivity. Thus, we may anticipate a relation for the turbulent resistivity of the form $\eta_T \sim \langle v^2 \rangle - \langle B^2 \rangle$, a considerable departure from expectations based upon kinematic models. A similar competition between mixing and

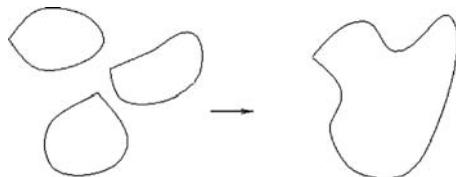


Figure 2. Inverse transfer: current filaments and A -blobs attract and coagulate.

coalescence appears in the spectral dynamics. Note also that η_T vanishes for turbulence at Alfvénic equipartition (i.e., $\langle v^2 \rangle = \langle B^2 \rangle$). Since the presence of even a weak mean magnetic field will naturally convert some of the fluid eddies to Alfvén waves, it is thus not entirely surprising that questions arise as to the possible reduction or “quenching” of the magnetic diffusivity relative to expectations based upon kinematics. Also, note that any such quenching is intrinsically a synergistic consequence of both:

- (i) the competition between flux advection and flux coalescence intrinsic to 2D MHD;
- (ii) the tendency of a mean magnetic field to “Alfvénize” the turbulence.

The close correspondence between the problems of 2D flux diffusion and that of the 3D mean field electromotive force, summarized in [Table 1](#), is remarkable. Both seek a representation of a mean product of fluid and magnetic fluctuations (i.e., the mean e.m.f. \mathcal{E} in 3D, Γ_A in 2D) in terms of local transport coefficients, namely α and β in 3D and η_T in 2D. In each case, the magnetic dynamics are critically constrained by the conservation, up to resistive dissipation, of magnetic helicity in 3D and of H_A in 2D. Both magnetic helicity and H_A inverse cascade to large scales, and thus produce an interesting dual cascade, since energy flows to small scales in each case. The inverse cascade of magnetic helicity and mean-square potential underpin the appearance of magnetic “back-reaction” contributions to α and η_T . In particular, $\alpha \sim (\langle \mathbf{v} \cdot \boldsymbol{\omega} \rangle - \langle \mathbf{B} \cdot \mathbf{J} \rangle)$, while $\eta_T \sim \langle v^2 \rangle - \langle B^2 \rangle$. Thus, both tend to vanish for fully Alfvénized

[Table 1](#). Table of analogies between calculations of 3D mean e.m.f. and 2D mean potential transport

3D Mean EMF	2D Mean Potential Flux
$\mathcal{E} = \langle \mathbf{v} \times \mathbf{B} \rangle$	$\Gamma_A = \langle v_x A \rangle$
$\mathcal{E} = \alpha \langle \mathbf{B} \rangle - \beta \langle \mathbf{J} \rangle$	$\Gamma_A = -\eta_T \partial \langle A \rangle / \partial x$
invariant \rightarrow Helicity	invariant \rightarrow
$\int \mathbf{A} \cdot \mathbf{B} d^3x$	mean-square magnetic potential
inverse cascade of magnetic helicity	$\int A^2 d^2x$
back $\alpha \sim \langle \mathbf{B} \cdot \mathbf{J} \rangle$	inverse cascade of H_A
$\langle \mathbf{B} \cdot \mathbf{J} \rangle$ from helicity balance	negative diffusivity $\sim \langle B^2 \rangle$
α quenching	$\langle A^2 \rangle$ from H_A balance
β -quenching	η_T quenching

turbulence. This trend, then, naturally suggests the possibility of both α -quenching in 3D, and magnetic diffusivity quenching in 2D. Of course, there are crucial *differences* between the two problems. Obviously, in 2D only decay of the magnetic field is possible, whereas 3D admits the possibility of dynamo growth. Furthermore, magnetic helicity and α (the pertinent quantities in 3D) are pseudo-scalars while H_A and η_T are scalars; thus, the effect of helicity conservation on β , the magnetic diffusivity in three dimensions, remains far from clear.

An important element of the basic physics, common to both problems, is the process of “Alfvénization”, whereby eddy energy is converted to Alfvén wave energy. This may be thought of as a physical perspective on the natural trend of MHD turbulence toward an approximate balance between fluid and magnetic energies, for $P_m \sim 1$. Note also that Alfvénization may be thought of as the development of a *dynamical memory*, which constrains and limits the cross-phase between v_x and A . This is readily apparent from the fact that $\langle v_x A \rangle$ vanishes for Alfvén waves in the absence of resistive dissipation. For Alfvén waves then, flux diffusion is directly proportional to resistive dissipation, an unsurprising conclusion for cross-field transport of flux which is, in turn, frozen into the fluid, up to η . As we shall soon see, the final outcome of the quenching calculation also reveals an explicit proportionality of η_T to η . For small η , then, Γ_A will be quenched. Another perspective on Alfvénization comes from the studies of Lyapunov exponents of fluid elements in MHD turbulence (Cattaneo *et al.*, 1996). These showed that as small-scale magnetic fields are amplified and react back on the flow, Lyapunov exponents drop precipitously, so that chaos is suppressed. This observation is consistent with the notion of the development of a dynamical memory, discussed above.

2.3. Mean field electrodynamics for $\langle A \rangle$ in 2D

In this section, we discuss the mean field theory of flux diffusion in 2D. In the discussion of the calculation of Γ_A , we do not address the relationship between the turbulent velocity field and the mechanisms by which the turbulence is excited or stirred. However, a weak large-scale field (the transport of which is the process to be studied) will be violently stretched and distorted, resulting in the rapid generation of a spectrum of magnetic turbulence. As discussed above, magnetic turbulence will likely tend to retard and impede the diffusion of large-scale magnetic fields. This, of course, is the crux of the matter, as Γ_A depends on the full spectrum arising from the external excitation and the back-reaction of the magnetic field, so the net imbalance of $\langle v^2 \rangle$ and $\langle B^2 \rangle$ determines

the degree of η_T quenching. Leverage on $\langle B^2 \rangle$ is obtained by considering the evolution of mean-square magnetic potential density \mathcal{H}_A . In particular, the conservation of $H_A = \int \mathcal{H}_A d^2x$ straightforwardly yields the identity

$$\frac{1}{2} \frac{\partial H_A}{\partial t} = -\Gamma_A \frac{\partial \langle A \rangle}{\partial x} - \eta \langle B^2 \rangle, \quad (5)$$

where the surface terms vanish for periodic boundaries. For stationary turbulence, then, this gives

$$\langle B^2 \rangle = -\frac{\Gamma_A}{\eta} \frac{\partial \langle A \rangle}{\partial x} = \frac{\eta_T}{\eta} \left(\frac{\partial \langle A \rangle}{\partial x} \right)^2, \quad (6)$$

which is the well-known Zeldovich (1957) theorem ($\langle B^2 \rangle / \langle B \rangle^2 = \eta_T / \eta$) for 2D MHD. The physics of the Zeldovich theorem is discussed further in the Appendix. The key message here is that when a weak mean magnetic field is coupled to a turbulent 2D flow, a *large mean-square fluctuation level can result*, on account of stretching iso- A or flux contours by the flow. However, while the behavior of $\langle B^2 \rangle$ is clear, we shall see that it is really $\langle B^2 \rangle_{\mathbf{k}}$ that enters the calculation of Γ_A , via a spectral sum.

To calculate Γ_A , standard closure methods (see, for example, Pouquet, 1978 or McComb, 1990) yield

$$\Gamma_A = \sum_{\mathbf{k}'} [v_x(-\mathbf{k}') \delta A(\mathbf{k}') - B_x(-\mathbf{k}') \delta \psi(\mathbf{k}')] = \sum_{\mathbf{k}'} \Gamma_A(\mathbf{k}'), \quad (7)$$

where $\delta A(\mathbf{k})$ and $\delta \psi(\mathbf{k})$ are, in turn, driven by the beat terms (in (3) and (4)) that contain the mean field $\langle A \rangle$. The calculational approach here treats fluid and magnetic fluctuations on an equal footing, and seeks to determine Γ_A by probing an evolved state of MHD turbulence, rather than a kinematically prescribed state of velocity fluctuations alone. The calculation follows those of Pouquet *et al.* (1976) and Pouquet (1978), and yields the result

$$\Gamma_A = - \sum_{\mathbf{k}'} [\tau_c^\psi(\mathbf{k}') \langle v^2 \rangle_{\mathbf{k}'} - \tau_c^A(\mathbf{k}') \langle B^2 \rangle_{\mathbf{k}'}] \frac{\partial \langle A \rangle}{\partial x} - \sum_{\mathbf{k}'} [\tau_c^A(\mathbf{k}') \langle A^2 \rangle_{\mathbf{k}'}] \frac{\partial}{\partial x} \langle J \rangle. \quad (8)$$

Here, consistent with the restriction to a weak mean field, isotropic turbulence is assumed. The quantities $\tau_c^\psi(\mathbf{k})$ and $\tau_c^A(\mathbf{k})$ are the self-correlation times (lifetimes), at \mathbf{k} , of the fluid and field perturbations, respectively. These are not at all necessarily equivalent to the coherence time of $v_x(-\mathbf{k}')$ with $A(\mathbf{k}')$, which determines Γ_A . For a weak mean field, both $\tau_c^\psi(\mathbf{k})$ and $\tau_c^A(\mathbf{k})$ are determined by nonlinear interaction processes, so that $1/\tau_c^{\psi,A}(\mathbf{k}') \geq k' \langle B \rangle$, i.e., fluctuation correlation times are

short in comparison to the Alfvén time of the mean field. In this case, the decorrelation process is controlled by the Alfvén time of the *r.m.s.* field (i.e., $[\mathbf{k}\langle B^2 \rangle^{1/2}]^{-1}$) and the fluid eddy turnover time, as discussed by Pouquet *et al.* (1976) and Pouquet (1978). Consistent with the assumption of unity magnetic Prandtl number, $\tau_c^\psi(\mathbf{k}) = \tau_c^A(\mathbf{k}) = \tau_c(\mathbf{k})$, hereafter.

The three terms on the right-hand side of (8) correspond respectively (Diamond *et al.*, 1984) to

- (a) a positive turbulent resistivity (i.e., Γ_A proportional to flux gradient) due to fluid advection of flux;
- (b) a negative turbulent resistivity symptomatic of the tendency of magnetic flux to accumulate on large scales;
- (c) a positive turbulent hyper-resistive diffusion, which gives Γ_A proportional to *current* gradient (Strauss, 1986). Such diffusion of current has been proposed as the mechanism whereby a magnetofluid undergoes Taylor relaxation (Taylor, 1986; Bhattacharjee and Hameiri, 1986; Bhattacharjee and Yuan, 1995).

Note that terms (b) and (c) both arise from $B_x(\mathbf{k})\delta\psi(\mathbf{k}')$, and show the trend in 2D MHD turbulence to pump large-scale H_A while damping small-scale H_A . For smooth, slowly varying mean potential profiles, the hyper-resistive term is negligible in comparison with the turbulent resistivity, (i.e., $\langle k'^2 \rangle > (1/\langle A \rangle)(\partial^2 \langle A \rangle / \partial x^2)$), so that the mean magnetic potential flux reduces to

$$\Gamma_A = -\eta_T \frac{\partial \langle A \rangle}{\partial x}, \quad (9)$$

where

$$\eta_T = \sum_{\mathbf{k}'} \tau_c(\mathbf{k}') \left(\langle v^2 \rangle_{\mathbf{k}'} - \langle B^2 \rangle_{\mathbf{k}'} \right). \quad (10)$$

As stated above, the critical element in determining Γ_A is to calculate $\langle B^2 \rangle_{\mathbf{k}'}$ in terms of $\langle v^2 \rangle_{\mathbf{k}'}$, Γ_A itself, etc. For this, mean-square magnetic potential balance is crucial! To see this, note that the flux equation may be written as

$$\frac{\partial A}{\partial t} + \mathbf{v} \cdot \nabla A = -v_x \frac{\partial \langle A \rangle}{\partial x} + \eta \nabla^2 A, \quad (11)$$

so multiplying by A and summing over modes gives

$$\frac{1}{2} \left[\frac{\partial}{\partial t} (\langle A^2 \rangle) + \langle \nabla \cdot (\mathbf{v} A^2) \rangle \right] = -\Gamma_A \frac{\partial \langle A \rangle}{\partial x} - \eta \langle B^2 \rangle, \quad (12)$$

assuming incompressibility of the flow. An equivalent, **k**-space version of (12) is

$$\frac{1}{2} \left[\frac{\partial}{\partial t} \langle A^2 \rangle_{\mathbf{k}} + T(\mathbf{k}) \right] = -\Gamma_A(\mathbf{k}) \frac{\partial \langle A \rangle}{\partial x} - \eta \langle B^2 \rangle_{\mathbf{k}}, \quad (13)$$

where $T(\mathbf{k})$ is the triple correlation

$$T(\mathbf{k}) = \langle \nabla \cdot (\mathbf{v} A^2) \rangle_{\mathbf{k}}, \quad (14)$$

which controls the nonlinear transfer of mean-square potential, and $\Gamma_A(\mathbf{k}) = \langle v_x A \rangle_{\mathbf{k}}$ is the **k**-component of the flux. Equations (12) and (13) thus allow the determination of $\langle B^2 \rangle$ and $\langle B^2 \rangle_{\mathbf{k}}$ in terms of Γ_A , $\Gamma_A(\mathbf{k})$, $T(\mathbf{k})$ and $\partial \langle A^2 \rangle_{\mathbf{k}} / \partial t$.

At the simplest, crudest level (the so-called τ -approximation), a single τ_c is assumed to characterize the response or correlation time in (10). In that case, we have

$$\Gamma_A = - \left[\sum_{\mathbf{k}} \tau_c (\langle v^2 \rangle_{\mathbf{k}} - \langle B^2 \rangle_{\mathbf{k}}) \right] \frac{\partial \langle A \rangle}{\partial x}. \quad (15)$$

For this, admittedly over-simplified case, (12) then allows the determination of $\langle B^2 \rangle$ in terms of Γ_A , the triplet and $\partial_t \langle A^2 \rangle$. With the additional restrictions of stationary turbulence and periodic boundary conditions (so that $\partial \langle A^2 \rangle / \partial t = 0$ and $\langle \nabla \cdot (\mathbf{v} A A) \rangle = 0$), it follows that

$$\langle B^2 \rangle = - \frac{\Gamma_A}{\eta} \frac{\partial \langle A \rangle}{\partial x}, \quad (16)$$

so that magnetic fluctuation energy is directly proportional to magnetic potential flux, via H_A balance. This corresponds to a balance between local dissipation and spatial flux in the mean-square potential budget (Gruzinov and Diamond, 1995, 1996). Inserting this into (10) then yields the following expression for the turbulent diffusivity:

$$\eta_T = \frac{\sum_{\mathbf{k}} \tau_c \langle v^2 \rangle_{\mathbf{k}}}{1 + \tau_c v_{A0}^2 / \eta} = \frac{\eta^k}{1 + R_m v_{A0}^2 / \langle v^2 \rangle}, \quad (17)$$

where η^k refers to the kinematic turbulent resistivity $\tau_c \langle v^2 \rangle$, v_{A0} is the Alfvén speed of the mean $\langle B \rangle$, and $R_m = \langle v^2 \rangle \tau_c / \eta$. It is instructive to note that (17) can be rewritten as

$$\eta_T = \frac{\eta \eta^k}{\eta + \tau_c v_{A0}^2}. \quad (18)$$

Thus, as indicated by mean-square potential balance, Γ_A ultimately scales directly with the collisional resistivity, a not unexpected result

for Alfvénized turbulence with dynamically interesting magnetic fluctuation intensities. This result supports the intuition discussed earlier. It is also interesting to note that for $R_m v_{A0}^2 / \langle v^2 \rangle > 1$ and $\langle v^2 \rangle \sim \langle B^2 \rangle$, $\eta_T \cong \eta(B^2) / \langle B \rangle^2$, consistent with the Zeldovich theorem prediction.

Equation (17) gives the well-known result for the quenched flux diffusivity. There, the kinematic diffusivity η_T^k is modified by the quenching or suppression factor $[1 + R_m v_{A0}^2 / \langle v^2 \rangle]^{-1}$, the salient dependencies of which are on R_m and $\langle B \rangle^2$. Equation (17) predicts a strong quenching of η_T with increasing $R_m \langle B \rangle^2$. Despite the crude approximations made in the derivation, numerical calculations indicate remarkably good agreement between the measured cross-field flux diffusivity (as determined by following marker particles tied to a flux element) and the predictions of (17). In particular, the scalings with both R_m and $\langle B \rangle^2$ have been verified, up to R_m values of a few hundred (Cattaneo, 1994).

Of course, the derivation of (17), as well as the conclusion of a quenched magnetic diffusivity, have provoked many questions, together with a vigorous debate in the community, though primarily in the context of directly analogous issues in the 3D alpha-quenching problem. The criticisms leveled at the treatment of α in the 3D problem (e.g., by Blackman and Field, 2000, 2002) must however also carry over to the treatment of β in the 2D case, and so we address this issue here. Criticism has focused primarily upon what is perceived as an inadequate treatment of the triplet term $\langle \nabla \cdot (\mathbf{v} \mathbf{A} \mathbf{A}) \rangle$ in (12). Note that $\langle \nabla \cdot (\mathbf{v} \mathbf{A} \mathbf{A}) \rangle$ makes no contribution to global \mathcal{H}_A balance in a periodic system. However, while $\langle \nabla \cdot (\mathbf{v} \mathbf{A} \mathbf{A}) \rangle = 0$ in this case, $\langle \mathbf{v} \cdot (\nabla \mathbf{A} \mathbf{A}) \rangle_{\mathbf{k}}$ does not. This contribution to the $\langle A^2 \rangle$ dynamics corresponds to

- (i) the divergence of the flux of mean-square potential, $\nabla \cdot \Gamma_{A^2}$, (here $\Gamma_{A^2} = \mathbf{v} \mathbf{A} \mathbf{A}$), when considered in a region of position space of scale $|\mathbf{k}|^{-1}$;
- (ii) the spectral transport of $\langle A^2 \rangle_{\mathbf{k}}$, when considered in \mathbf{k} -space.

In either case, a new timescale enters the mean-square magnetic potential budget which can, in principle, break the balance between $\Gamma_A \langle B \rangle$ and resistive dissipation. Physically, this timescale has been associated with

- (i) the net outflow of mean-square potential at the boundaries, in the case of a non-periodic configuration. In this regard, it has been conjectured that should the loss rate of $\langle A^2 \rangle$ exceed that of $\langle A \rangle$, the quench of η_T would be weaker.
- (ii) the *local* effective transport rate (on scales $\sim |\mathbf{k}|^{-1}$) of mean-square potential or, alternatively, the local spectral

transport rate of $\langle A^2 \rangle_{\mathbf{k}}$. Note that in this case, boundary conditions are irrelevant. Thus, *local* $\langle A^2 \rangle$ spectral transport effects should manifest themselves in numerical calculations with periodic boundaries, such as those by Cattaneo (1994).

To address these questions, one must calculate the triplet correlations. In this regard, it is instructive to consider them from the point of view of transport in position space (i.e., $\langle \nabla \cdot (\mathbf{v}AA) \rangle$), together with the equivalent spectral transfer in \mathbf{k} -space. The goal here is to assess the degree to which triplet correlations enter the relationship between resistive dissipation and magnetic flux transport, which is central to the notion of quenching.

Recall, on retaining the volume-averaged advective flux, that the equation for the mean-square potential fluctuation is

$$\frac{1}{2} \left(\frac{\partial}{\partial t} \langle AA \rangle + \langle \mathbf{v} \cdot \nabla AA \rangle \right) = -\langle v_x A \rangle \frac{\partial \langle A \rangle}{\partial x} - \eta \langle B^2 \rangle. \quad (19)$$

Observe that since $-\langle v_x A \rangle \partial \langle A \rangle / \partial x = \eta_T \langle B \rangle^2$, the right-hand side of (19) simply reduces to the Zeldovich theorem, $\langle B^2 \rangle / \langle B \rangle^2 = \eta_T / \eta$, in the absence of contributions from the triplet moment. For stationary turbulence, then, the proportionality between mean flux transport and resistive dissipation is broken by the triplet $\langle \mathbf{v} \cdot \nabla AA \rangle$, which may be rewritten as $\langle \mathbf{v} \cdot \nabla AA \rangle = \nabla \cdot (\mathbf{v}AA) = \int \Gamma_{A^2} \cdot d\mathbf{n}$, using Gauss's law. Here $\Gamma_{A^2} = \mathbf{v}AA$ is the flux of mean-square potential and the integration $\int d\mathbf{n}$ is normal to a contour enclosing the region of averaging denoted by the bracket. This scale must, of course, be smaller than the mean field scale ℓ_o for consistency of the averaging procedure. Mean-square potential evolution is thus given by

$$\frac{1}{2} \left(\frac{\partial}{\partial t} \langle AA \rangle + \int d\mathbf{n} \cdot \Gamma_{A^2} \right) = -\langle v_x A \rangle \frac{\partial \langle A \rangle}{\partial x} - \eta \langle B^2 \rangle, \quad (20)$$

so that the balance of mean flux transport and local dissipation is indeed broken by the *net* in/out flux of mean-square potential to the averaging region. Alternatively, the triplet correlation renders the mean-square potential balance *non-local*. Of course, $\int d\mathbf{n} \cdot \Gamma_{A^2}$ is determined by the values of the turbulent velocity and potential perturbation on the boundary of the averaging region. The non-local term in the \mathcal{H}_A budget is by no means “small” in any naive sense, either—indeed a straightforward estimate of the ratio of the second term (the A^2 flux) in (20) to the third term gives $(B / \langle B \rangle)(k_0 \ell)^{-1}$, where k_0 is a typical perturbation wave

vector and ℓ is the scale of the averaging region. As $B/\langle B \rangle \sim \sqrt{R_m} \gg 1$ and $(k_o\ell)^{-1} \leq 1$, this ratio can certainly be large, so the triplet term is by no means *a priori* negligible. However, two caveats are important. First, a *net* influx or outflux is required, these being more suggestive of an externally driven process, rather than one that is spontaneous (i.e., in 3D, of helicity injection rather than a dynamo). Second, the quantity Γ_{A^2} *may not* be calculated kinematically, for exactly the same reasons that the kinematic theory of Γ_A fails so miserably! This latter point is discussed at length, below.

Noting that a *net inflow or outflow* of mean-squared potential is required to break the local balance between resistive dissipation and mean potential transport (i.e., turbulent resistivity), critics (most prominently Blackman and Field, 2000) of the notion of quenching have advanced the suggestion that a net in/out flux Γ_{A^2} of mean-square potential at the system boundary may weaken the quench. Implicit in this suggestion is the idea that Γ_{A^2} will exceed Γ_A , or alternatively, that the in/out flow rate of mean-square potential exceeds that of the mean potential. We shall see below that when Γ_{A^2} and Γ_A are both calculated self-consistently, this is *not* the case. While a definitive numerical test of this hypothesis has yet to be performed, the results of recent numerical calculations that relax the periodic boundary conditions used in earlier studies by prescribing A or $\partial A/\partial y$ on the upper and lower boundaries indicate no significant departure from the predicted effective resistivity quench (Wilkinson and Hughes, 2005). We hasten to add, however, that while these calculations *do* suggest that the dynamics of turbulent transport are insensitive to boundary conditions, they do not actually examine the effects of external magnetic potential injection.

It is also instructive to examine the triplet correlations in \mathbf{k} -space, as well as in configuration space. Indeed, it is here that the tremendous departure of Γ_{A^2} from kinematic estimates is most apparent. In \mathbf{k} -space, \mathcal{H}_A evolution is described by

$$\frac{1}{2} \left(\frac{\partial}{\partial t} \langle AA \rangle_{\mathbf{k}} + T_{\mathbf{k}} \right) = -\langle v_x A \rangle_{\mathbf{k}} \frac{\partial \langle A \rangle}{\partial x} - \eta \langle B^2 \rangle_{\mathbf{k}}, \quad (21)$$

where the triplet $T_{\mathbf{k}}$ is just

$$T_{\mathbf{k}} = \langle \mathbf{v} \cdot \nabla AA \rangle_{\mathbf{k}}. \quad (22)$$

In \mathbf{k} -space, spectral transfer, rather than spatial transport and mixing, breaks the balance between resistive dissipation and turbulent transport. Thus, the key issue is the calculation of $T_{\mathbf{k}}$. This is easily accomplished by standard closure methods as discussed by Pouquet (1978).

Thus, applying EDQNM or DIA-type closures, $T_{\mathbf{k}}$ is straightforwardly approximated as

$$T_{\mathbf{k}} = \sum_{\mathbf{k}'} (\mathbf{k} \cdot \mathbf{k}' \times \hat{\mathbf{z}})^2 \theta_{\mathbf{k}, \mathbf{k}'} \left\{ \langle \psi^2 \rangle_{\mathbf{k}'} - \left[\frac{|\mathbf{k}'|^2 - |\mathbf{k}|^2}{|\mathbf{k} + \mathbf{k}'|^2} \right] \langle A^2 \rangle_{\mathbf{k}'} \right\} \langle A^2 \rangle_{\mathbf{k}} \\ - \sum_{\substack{\mathbf{p}, \mathbf{q} \\ \mathbf{p} + \mathbf{q} = \mathbf{k}}} (\mathbf{p} \cdot \mathbf{q} \times \hat{\mathbf{z}})^2 \theta_{\mathbf{k}, \mathbf{p}, \mathbf{q}} \langle \psi^2 \rangle_{\mathbf{p}} \langle A^2 \rangle_{\mathbf{q}}, \quad (23)$$

where $\theta_{\mathbf{k}, \mathbf{p}, \mathbf{q}}$ is the triad coherence time $\theta_{\mathbf{k}, \mathbf{p}, \mathbf{q}} = (1/\tau_{c\mathbf{k}} + 1/\tau_{c\mathbf{q}} + 1/\tau_{c\mathbf{p}})^{-1}$. In (23), the first and third terms represent advection of potential by the turbulent velocity, the first giving a turbulent resistivity, the third incoherent noise. Note that these two contributions conserve $\langle A^2 \rangle$ against each other when summed over \mathbf{k} . The second term in (23) corresponds to inverse transfer of mean-square potential via flux coalescence. Note that it is negative on large scales ($k^2 < k'^2$), yielding the negative turbulent resistivity, and positive on small scales ($k^2 > k'^2$), giving the positive hyper-resistivity. Observe that the second term is manifestly antisymmetric in \mathbf{k} and \mathbf{k}' , and so conserves $\langle A^2 \rangle$ individually, when summed over \mathbf{k} .

It is immediately clear that, just as in the case of $\langle A \rangle$, $\langle A^2 \rangle$ evolution is determined by the competition between advective straining and mixing of iso- A contours, together with the tendency of these flux structures to coalesce to progressively larger scales. This is hardly a surprise, since A and all its moments are frozen into the flow, up to resistive dissipation. Note also that a proper treatment of mean-square potential conservation (i.e., $\sum_{\mathbf{k}} T_{\mathbf{k}} = 0$) requires that nonlinear noise due to incoherent mode coupling also be accounted for.

Equation (21) can be re-written in the form

$$\frac{1}{2} \left(\frac{\partial}{\partial t} \langle AA \rangle_{\mathbf{k}} + \widehat{\eta}_T \mathbf{k} \langle AA \rangle_{\mathbf{k}} - N_{\mathbf{k}} \right) = -\langle v_x A \rangle_{\mathbf{k}} \frac{\partial \langle A \rangle}{\partial x} - \eta \langle B^2 \rangle_{\mathbf{k}}, \quad (24)$$

having written

$$T_{\mathbf{k}} = \widehat{\eta}_T \mathbf{k} \langle AA \rangle_{\mathbf{k}} - N_{\mathbf{k}} \quad (25)$$

with

$$\widehat{\eta}_T \mathbf{k} = \sum_{\mathbf{k}'} (\mathbf{k} \cdot \mathbf{k}' \times \hat{\mathbf{z}})^2 \theta_{\mathbf{k}, \mathbf{k}'} (\langle \psi^2 \rangle_{\mathbf{k}'} - \langle A^2 \rangle_{\mathbf{k}'}) \quad (26)$$

and

$$N_{\mathbf{k}} = \sum_{\substack{\mathbf{p}, \mathbf{q} \\ \mathbf{p} + \mathbf{q} = \mathbf{k}}} (\mathbf{p} \cdot \mathbf{q} \times \hat{\mathbf{z}})^2 \theta_{\mathbf{k}, \mathbf{p}, \mathbf{q}} \langle \psi^2 \rangle_{\mathbf{p}} \langle A^2 \rangle_{\mathbf{q}}. \quad (27)$$

Note that $\hat{\eta}_T \mathbf{k} \rightarrow \partial/\partial x (\eta_T \partial/\partial x)$ as $\mathbf{k} \rightarrow \mathbf{0}$. It is interesting to compare terms on the left- and right-hand side of (24). Nonlinear transfer terms $\sim \langle \nabla \cdot \Gamma_{A^2} \rangle_{\mathbf{k}}$ are $O(k A^2 |v|)$, while mean flux terms are $O(|vA| |\langle B \rangle|)$. Thus, the ratio $|T_{\mathbf{k}}| / |(vA)_{\mathbf{k}}| |\langle B \rangle| \sim O(B/\langle B \rangle)$. Here, $B/\langle B \rangle \gg 1$, as we are considering a strongly turbulent, weakly magnetized regime. Thus, to lowest order in $(B/\langle B \rangle)^{-1}$, (24) (at stationarity) must reduce to

$$T_{\mathbf{k}} = 0, \quad (28)$$

so that nonlinear transfer determines the magnetic potential spectrum. In physical terms, this means that $\langle AA \rangle_{\mathbf{k}}$ adjusts to balance nonlinear noise, which is the main source here. We formally refer to this spectrum as $\langle AA \rangle_{\mathbf{k}}^{(0)}$. Note that $\langle AA \rangle_{\mathbf{k}}^{(0)}$ is actually determined, as is usual for spectral transfer processes, by the balance between $N_{\mathbf{k}}$ (incoherent mode coupling) and $\hat{\eta}_T \mathbf{k} \langle AA \rangle_{\mathbf{k}}$ (turbulent dissipation). This guarantees that the net spectral flow rate is constant in \mathbf{k} , so \mathcal{H}_A is conserved. Nonlinear noise is critical here (to respect \mathcal{H}_A conservation) and, in fact, constitutes the dominant source for $\langle AA \rangle_{\mathbf{k}}^{(0)}$ when $B/\langle B \rangle \gg 1$. Note that a corresponding calculation for magnetic helicity by Blackman and Field (2002) neglects nonlinear noise. It is interesting to observe that, as a consequence, the classical “mean field electrodynamics” calculation of $\langle v_x A \rangle$ cannot be decoupled from the spectral transfer problem for $\langle AA \rangle_{\mathbf{k}}$. This of course follows from the constraint imposed upon the former by \mathcal{H}_A conservation. To next order in $(B/\langle B \rangle)^{-1}$ then, (24) gives

$$0 = -\langle v_x A \rangle_{\mathbf{k}} \frac{\partial \langle A \rangle}{\partial x} - \eta \langle B^2 \rangle_{\mathbf{k}}, \quad (29)$$

the solution of which trivially yields $\langle B^2 \rangle_{\mathbf{k}}$. Ultimately, this yields a quenched turbulent resistivity of the form

$$\eta_T = \sum_{\mathbf{k}} \frac{\tau_{c\mathbf{k}} \langle v^2 \rangle_{\mathbf{k}}}{1 + \tau_{c\mathbf{k}} v_{A0}^2 / \eta}. \quad (30)$$

Note that this is basically equivalent to the result in (17), with, however, the quench factor varying with \mathbf{k} .

Several comments are in order here. First, it cannot be over-emphasized that a self-consistent calculation of $\langle \nabla \cdot \Gamma_{A^2} \rangle_{\mathbf{k}}$ is crucial to this conclusion. Such a calculation necessarily must include both

nonlinear response and nonlinear noise. A kinematic calculation would leave $\hat{\eta}_T \mathbf{k} > \eta_T$, which is incorrect. Likewise, neglecting noise would violate \mathcal{H}_A conservation. It is also amusing to note that the question of the relation between $\langle v_x A \rangle_{\mathbf{k}} \langle A \rangle'$ and $\eta \langle B^2 \rangle_{\mathbf{k}}$ does *not* hinge upon boundary conditions or inflow/outflow, at all. Hence, the available numerical experiments, already published, constitute a successful initial test of the theory of flux diffusivity quenching in 2D, at least for modest values of R_m and for smooth $\langle A \rangle$ profiles.

It is instructive to return to configuration space now, in order to compare the rates of transport of $\langle A \rangle$ and $\langle AA \rangle$. The analysis given above may be summarized by writing the equations of evolution for $\langle A \rangle$, i.e.,

$$\frac{\partial}{\partial t} \langle A \rangle = \frac{\partial}{\partial x} \left(\eta_T \frac{\partial \langle A \rangle}{\partial x} \right), \quad (31)$$

where

$$\eta_T = \sum_{\mathbf{k}'} \tau_{c\mathbf{k}'} (\langle v^2 \rangle_{\mathbf{k}'} - \langle B^2 \rangle_{\mathbf{k'}}); \quad (32)$$

and for $\langle AA \rangle_{\mathbf{k}}$, i.e.,

$$\frac{1}{2} \left(\frac{\partial}{\partial t} \langle AA \rangle_{\mathbf{k}} + \hat{\eta}_T \mathbf{k} \langle AA \rangle_{\mathbf{k}} \right) = \frac{1}{2} N_{\mathbf{k}} - \langle v_x A \rangle_{\mathbf{k}} \frac{\partial \langle A \rangle}{\partial x} - \eta \langle B^2 \rangle_{\mathbf{k}}, \quad (33)$$

where

$$\hat{\eta}_T \mathbf{k} = \sum_{\mathbf{k}'} (\mathbf{k} \cdot \mathbf{k}' \times \hat{\mathbf{z}})^2 \theta_{\mathbf{k}, \mathbf{k}'} (\langle \psi^2 \rangle_{\mathbf{k}'} - \langle A^2 \rangle_{\mathbf{k'}}). \quad (34)$$

Not surprisingly, $\hat{\eta}_T \mathbf{k} \rightarrow \eta_T \nabla^2$ as $\mathbf{k} \rightarrow \mathbf{0}$. This is, of course, a straightforward consequence of the fact that the same physics governs the dynamics of $\langle A \rangle$ and $\langle A^2 \rangle$ since A is conserved along fluid trajectories, up to resistive dissipation. Hence, the total diffusive loss rates for $\langle A \rangle$ and $\langle A^2 \rangle$ are simply $1/\tau_A = \eta_T / L_A^2$ and $1/\tau_{A^2} = \eta_T / L_{A^2}^2$, where L_A and L_{A^2} are the gradient scale lengths for $\langle A \rangle$ and $\langle AA \rangle$, respectively. Here L_{A^2} is set either by the profile of forcing or injection, or by the profile of $\langle A \rangle$. For the latter, $\tau_A = \tau_{A^2}$ so that preferential loss of $\langle AA \rangle$ is impossible. For the former, inflow of flux at the boundary, say by plasmoid injection, could however decouple L_{A^2} from L_A . In this case, however, the magnetic dynamics are not spontaneous but, rather, strongly driven by external means.

2.4. Turbulent diffusion of flux and field in 3D

In this section, we discuss flux and field diffusion in three dimensions. In 3D, \mathbf{A} is *not* conserved along fluid element trajectories, so the flux diffusion problem becomes significantly more difficult. With this in mind, we divide the discussion of 3D diffusion into two sub-sections; one on turbulent diffusion in 3D reduced MHD (RMHD) (Strauss, 1976), the other on weakly magnetized, full MHD. This progression facilitates understanding, as 3D RMHD is quite similar in structure to 2D MHD, allowing us to draw on the experience and insight gained in the study of that problem.

Flux diffusion in 3D reduced MHD

The reduced MHD equations in 3D are:

$$\frac{\partial \mathbf{A}}{\partial t} + (\nabla \psi \times \hat{\mathbf{z}}) \cdot \nabla \mathbf{A} = B_0 \frac{\partial \psi}{\partial z} + \eta \nabla^2 \mathbf{A}, \quad (35)$$

$$\frac{\partial}{\partial t} \nabla^2 \psi + (\nabla \psi \times \hat{\mathbf{z}}) \cdot \nabla \nabla^2 \psi = v \nabla^2 \nabla^2 \psi + (\nabla \mathbf{A} \times \hat{\mathbf{z}}) \cdot \nabla \nabla^2 \mathbf{A} + B_0 \frac{\partial}{\partial z} \nabla^2 \mathbf{A}. \quad (36)$$

These equations describe incompressible MHD in the presence of a strong field $\mathbf{B}_0 = B_0 \hat{\mathbf{z}}$, which is externally prescribed and fixed. The “test field” undergoing turbulent diffusion is $\langle \mathbf{B} \rangle = \langle \mathbf{B}(x) \rangle \hat{\mathbf{y}}$, where $\langle \mathbf{B}(x) \rangle = -\partial \langle \mathbf{A} \rangle / \partial x$. Obviously, $\langle \mathbf{B} \rangle \ll B_0$ here.

Of course, the presence of a strong \mathbf{B}_0 renders 3D RMHD dynamics quite similar (but *not* identical!) to those in 2D. In particular, note that one can define a mean-square magnetic potential in 3D RMHD, i.e.,

$$H_A = \int d^2x \int A^2 dz, \quad (37)$$

and that H_A is *conserved up to resistive dissipation and Alfvénic coupling*, so that the fluctuation H_A balance becomes:

$$\frac{1}{2} \frac{\partial H_A}{\partial t} = -\langle v_x A \rangle \frac{\partial \langle A \rangle}{\partial x} - \eta \langle B^2 \rangle + B_0 \left\langle A \frac{\partial \psi}{\partial z} \right\rangle. \quad (38)$$

In contrast to its 2D counterpart (i.e., (5)), H_A balance is achieved by a competition between cross-field transport and resistive dissipation *together with Alfvénic propagation along B_0* (i.e., observe that the last term in (38) is explicitly proportional to B_0). It is interesting to note, however, that total H_A conservation is broken only by local dissipation

(as in 2D) and by a *linear* effect, which corresponds to wave propagation along B_0 . Thus, although H_A is not conserved (even as $\eta \rightarrow 0$), the potential equation nonlinearity (i.e., the nonlinearity in (38)) is *still* annihilated in 3D RMHD, as it is in 2D (i.e., $\langle \mathbf{v} \cdot \nabla A \rangle \rightarrow 0$, up to boundary flux terms)! Hence, the mean-square potential budget is still a powerful constraint on flux diffusion in 3D.

For simplicity and brevity, the discussion of flux diffusion in 3D is limited to the case of constant τ_c . Proceeding as in the previous section straightforwardly yields

$$\Gamma_A = -\tau_c(\langle v^2 \rangle - \langle B^2 \rangle) \frac{\partial \langle A \rangle}{\partial x}. \quad (39)$$

Here, the current diffusivity has been dropped, as for 2D. To relate $\langle B^2 \rangle$ to $\langle v^2 \rangle$ etc., mean-square potential balance and stationarity give

$$\langle B^2 \rangle = \frac{-\langle v_x A \rangle}{\eta} \frac{\partial \langle A \rangle}{\partial x} + \frac{B_0}{\eta} \left\langle A \frac{\partial \psi}{\partial z} \right\rangle. \quad (40)$$

Thus, the new element in 3D is the appearance of Alfvénic coupling (i.e., the last term on the right-hand side) in the H_A balance. This coupling is non-zero only if there is a net directivity in the radiated Alfvénic spectrum, or, equivalently, an imbalance in the two Elsasser populations, which account for the intensity of wave populations propagating in the $\pm \hat{\mathbf{z}}$ directions.

This contribution may be evaluated as before, i.e.,

$$\left\langle A \frac{\partial \psi}{\partial z} \right\rangle = \left\langle A \frac{\partial}{\partial z} \delta \psi \right\rangle + \left\langle \delta A \frac{\partial \psi}{\partial z} \right\rangle, \quad (41)$$

where $\delta \psi$ and δA are obtained via closure of (35, 36). A short calculation gives

$$\left\langle A \frac{\partial \psi}{\partial z} \right\rangle = \tau_c B_0 (\varepsilon_v \langle v^2 \rangle - \varepsilon_B \langle B^2 \rangle), \quad (42)$$

where

$$\varepsilon_v = \frac{\int k_z^2 \langle \psi^2 \rangle_{\mathbf{k}} d^3k}{\int (k_x^2 + k_y^2) \langle \psi^2 \rangle_{\mathbf{k}} d^3k}, \quad (43)$$

and ε_B similarly, with $\langle A^2 \rangle_{\mathbf{k}}$. Note that this approximation to $\langle A \partial \psi / \partial z \rangle$ clearly vanishes for equal Elsasser populations with identical spectral structure. This, of course, simply states that, in such a situation, there is no net imbalance or directivity in the Alfvénically radiated energy,

and thus no effect on the H_A budget. Taking $\varepsilon_v = \varepsilon_B$ and proceeding as in the 2D case finally yields

$$\Gamma_A = \frac{-\eta_T^k \partial \langle A \rangle / \partial x}{(1 + \tau_c/\eta) (\varepsilon_B B_0^2 + \langle B \rangle^2)}, \quad (44)$$

where

$$\eta_T^k = \tau_c \langle v^2 \rangle. \quad (45)$$

In 3D, τ_c is also a function of B_0^2 , i.e., $\tau_c = \tau_{NL}/(k_z^2 v_A^2 \tau_{NL}^2 + 1)$, where τ_{NL} is the amplitude-dependent correlation time.

The message of (44, 45) is that in 3D RMHD, the strong guide field \mathbf{B}_0 contributes to the quenching of η_T . The presence of the factor ε_B implies that this effect is sensitive to the parallel-perpendicular anisotropy of the turbulence, which is eminently reasonable. Thus, the degree of quenching in 3D RMHD is stronger than in 2D, as $B_0 \gg \langle B \rangle$. Finally, note however that the upshot of the quench is still that η_T scales with η , indicative of the effects of the freezing of magnetic potential into the fluid.

Given the attention paid to turbulence energy flux through the system boundary, it is worthwhile to comment here that the Alfvénic radiation contribution to the H_A budget ($\langle A \partial \psi / \partial z \rangle$) could be significantly different if there were a net imbalance in the two Elsasser populations. For example, this might occur in the solar corona, where Alfvén waves propagate away from the Sun, along “open” field-lines. In this case, a local balance between such Alfvénic leakage and cross-field transport could be established in the H_A budget. Such a balance would, of course, greatly change the scalings of η_T from those given here.

An interesting application of mean field electrodynamics within RMHD is to the problem of fast, turbulent reconnection in 3D (Lazarian and Vishniac, 1999; Kim and Diamond, 2001). As, in essence by definition, reconnection rates are measured globally (i.e., over some macroscopic region), they are necessarily constrained by conservation laws, such as that of H_A conservation. It is not surprising, then, that one upshot of the quenching of η_T (i.e., (18)) is that the associated magnetic reconnection velocity $V \leq (\langle v^2 \rangle / \langle v_A \rangle^2)^{1/2} v_{s-p}$, where v_{s-p} is the familiar Sweet-Parker velocity $v_{s-p} = \langle v_A \rangle / \sqrt{R_m}$, where $R_m = \langle v_A \rangle L / \eta$. Note that this result states that the reconnection rate is enhanced beyond the prediction of collisional theory, but still exhibits the Sweet-Parker type scaling with resistivity.

Magnetic field diffusion (β -effect) in full, 3D MHD

Moving now to consider the case of weakly magnetized, incompressible, 3D MHD, magnetic potential is no longer conserved, even approximately. Detailed calculations (Gruzinov and Diamond, 1994; Kim, 1999) predict that

$$\eta_T \cong \eta_T^k, \quad (46)$$

or, equivalently, that the kinematic turbulent resistivity is unchanged and unquenched, to leading order. The obvious question then naturally arises as to why α is quenched (see Section 3) but η_T (or, equivalently, β) is not. Here, we note that β being a scalar, and not a pseudo-scalar like α , plays no role in magnetic helicity balance. As magnetic helicity balance, which forces a balance between α and resistive dissipation of magnetic helicity ($\sim \eta(\mathbf{B} \cdot \mathbf{J})$), together with stationarity, is the origin of α -quenching, it is thus not at all surprising that η_T is not quenched in 3D, for weak fields. Of course the weak field result stated here must necessarily pass to the strong field RMHD case discussed earlier, as a strong guiding field is added. The analytical representation of β that smoothly connects these two limiting cases has yet to be derived, and remains an open question in the theory.

Computational studies have not yet really confronted the physics of magnetic flux diffusion in 3D. While two rather minimal studies exist (Thelen and Cattaneo, 2001; Brandenburg, 2001), neither presents systematic scans over R_m or a broad scan over $\langle B \rangle^2$. Although results indicate some tendency toward reduction of β as $\langle B \rangle^2$ increases, it is unclear whether or not the onset of this occurs in the “weak” or “strong” field limit. Further work is clearly needed.

2.5. Discussion and conclusion

In this section, we have reviewed the theory of turbulent transport of magnetic flux and field in 2D and 3D MHD. The 2D flux diffusion problem has been given special attention for its intrinsic interest and relative simplicity, as well as for its many similarities to the problem of the α -effect in 3D. Several issues that are of current interest have been addressed in detail. These include: boundary in-flow and out-flow effects on the mean-square potential budget, the role of nonlinear spectral transfer in the mean-square potential budget, and the dynamics

of magnetic flux in 3D reduced MHD. Several topics for further study have been identified, including, but not limited to

- (i) the derivation of an expression for diffusion in 3D that unifies the weak and strong field regimes;
- (ii) a numerical study of transport in 2D that allows a net flux of turbulence through the system boundary;
- (iii) both a theoretical and numerical study of flux diffusion in 3D with balanced and unbalanced Elsasser populations, for various along-field boundary conditions;
- (iv) a study of η_T quenching for $P_m \gg 1$ and a consideration of non-stationary states.

Constraints on space force us to omit several extensions and related topics. These include, for example, the turbulent transport of a passive scalar in 2D MHD (Diamond and Gruzinov, 1997), the self-consistent ambipolar diffusion problem in 2D (Kim, 1997; Leprovost and Kim, 2003) and the study of diffusion of magnetic fields in 2D electron MHD (Das and Diamond, 2000; Dastgeer *et al.*, 2000).

3. The generation of magnetic fields

In the previous section we discussed at length the turbulent diffusion of a magnetic field, concentrating primarily on the case of a two-dimensional, coplanar field and flow. The defining feature of this two-dimensional system is that there is no possibility of field *generation*; any large-scale field of zero mean is guaranteed to decay completely, the interest being in the nature and timescale of this decay. Decay in two dimensions is a consequence of the conservation of mean-square magnetic potential. In three dimensions this is no longer the case and, notwithstanding the fascinating similarities between the 2D and 3D cases, represents a major physical difference between them.

Historically, astrophysical interest has been in the behavior of large-scale magnetic fields, as manifested, for example, by the eleven-year solar cycle. Therefore, as explained for the 2D problem, it is natural to seek evolution equations for the large-scale magnetic field involving transport coefficients dependent on properties of the small-scale velocity field and small-scale magnetic field. The simplest approach is to assume a scale separation between the (small) scale of the velocity field, l , and the large scale, L , of the magnetic field. (Although within this framework it is straightforward also to include the effects of a large-scale velocity

field, we shall here, for simplicity, assume that the velocity field is only small-scale.) Thus we write

$$\mathbf{B} = \mathbf{B}_0 + \mathbf{b}, \quad (47)$$

where $\langle \mathbf{b} \rangle = \mathbf{0}$, angle brackets denoting an average over some intermediate length scale a satisfying $l \ll a \ll L$. By assumption, $\langle \mathbf{V} \rangle = \mathbf{0}$.

3.1. The linear regime

In order to highlight the crucial differences between the two- and three-dimensional problems it is instructive first to consider the kinematic problem, in which the velocity field is prescribed independently of the magnetic field. Averaging the induction equation

$$\frac{\partial \mathbf{B}}{\partial t} = \nabla \times (\mathbf{v} \times \mathbf{B}) + \eta \nabla^2 \mathbf{B}, \quad (48)$$

gives

$$\frac{\partial \mathbf{B}_0}{\partial t} = \nabla \times \mathcal{E} + \eta \nabla^2 \mathbf{B}_0, \quad (49)$$

where $\mathcal{E} = \langle \mathbf{v} \times \mathbf{b} \rangle$ is the mean electromotive force. Obviously, to make progress with (49) it is necessary to express \mathcal{E} in terms of the mean field \mathbf{B}_0 (and its derivatives). The standard approach, developed by Steenbeck, Krause and Rädler in the 1960s (see Krause and Rädler, 1980, for full references), comes from considering the equation for the fluctuating component of the magnetic field, \mathbf{b} , obtained by subtracting (49) from the induction equation (48). This gives

$$\left(\frac{\partial}{\partial t} - \eta \nabla^2 \right) \mathbf{b} = \nabla \times (\mathbf{v} \times \mathbf{B}_0) + \nabla \times \mathbf{G}, \quad (50)$$

where $\mathbf{G} = \mathbf{v} \times \mathbf{b} - \langle \mathbf{v} \times \mathbf{b} \rangle$. The right-hand side of (50) may be interpreted as expressing the two dynamical ingredients contributing to the evolution of the small-scale field \mathbf{b} . One is through the mean field \mathbf{B}_0 acting as a source for the small-scale field via the $\nabla \times (\mathbf{v} \times \mathbf{B}_0)$ term. The other reflects the evolution of the small-scale field *even in the absence of a mean field*. It is instructive to consider these contributions carefully since their understanding lies at the very heart of 3D mean field electrodynamics.

The traditional interpretation of (50) (see, for example, Krause and Rädler, 1980; Moffatt, 1978) has been to assume that the small-scale field \mathbf{b} is driven entirely by the source term $\nabla \times (\mathbf{v} \times \mathbf{B}_0)$ —in other

words, that in the absence of the mean field \mathbf{B}_0 the small-scale field would simply decay. From this assumption it becomes possible to construct the extremely elegant theory of mean field electrodynamics. If \mathbf{b} is linearly related to \mathbf{B}_0 , then, at least for a prescribed flow, $\mathcal{E} = \langle \mathbf{v} \times \mathbf{b} \rangle$ must also be linearly related to \mathbf{B}_0 . Since the mean field varies on a large length scale, it is natural to posit an expansion for \mathcal{E} of the form

$$\mathcal{E}_i = \alpha_{ij} B_{0j} + \beta_{ijk} \frac{\partial B_{0j}}{\partial x_k} + \dots \quad (51)$$

As we shall see, for consistency with (49) it is important that the first two terms (and only the first two terms) of this expansion are retained. Since \mathcal{E} is a polar vector whereas \mathbf{B} is an axial vector, it follows therefore that α_{ij} and β_{ijk} are *pseudo*-tensors. The physical interpretation of α_{ij} and β_{ijk} can be most clearly seen for the simple case of isotropic turbulence, for which α_{ij} and β_{ijk} , being dependent only on the properties of the flow, must be isotropic tensors; i.e., $\alpha_{ij} = \alpha \delta_{ij}$ and $\beta_{ijk} = \beta \epsilon_{ijk}$, where α is a *pseudo*-scalar and β is a true scalar. Substitution from (51) into (49) then gives

$$\frac{\partial \mathbf{B}_0}{\partial t} = \nabla \times \alpha \mathbf{B}_0 + (\eta + \beta) \nabla^2 \mathbf{B}_0, \quad (52)$$

where, for expository purposes, we have made the further simplifying assumption that β is a constant. Clearly β is an additional, turbulent, contribution to the diffusion, essentially as discussed at length in Section 2. The term involving α (the famous “ α -effect” of mean field electrodynamics) is of significance only for three-dimensional fields and flows, and represents the possibility of magnetic field generation—i.e., dynamo action. Over the last three decades, a substantial literature has developed through applying the ideas of mean field electrodynamics, as captured by (52) and more complicated versions thereof, to explaining observed astrophysical fields. Although this is a topic of considerable interest, it is beyond the scope of this review, in which we concentrate solely on the fundamental aspects of the transport coefficients.

Simply through parity considerations it is possible to deduce an immensely powerful result that gets to the very heart of the α -effect. Consider the idealized case of isotropic turbulence that is *reflectionally symmetric*; in other words, the fluid motions, on average, possess no *handedness*. For such turbulence, α , which is assumed to depend only on the statistical properties of the flow (and which, by assumption, possesses no handedness), must remain invariant under a change from a description in terms of a right-handed set of axes to one in terms of a left-handed set. Conversely, α , being a *pseudo*-scalar, must change sign

under such a parity transformation. Consequently, we are forced to the conclusion, purely through parity arguments, that α can be non-zero only for turbulence that lacks reflectional symmetry. Although handedness is imbued into practically all astrophysical systems via rotation, with the Coriolis force providing a natural breaking of reflectional symmetry, the significance of its influence will depend on the temporal and spatial scales of the fluid motions. For example, on the Sun, the fluid motions responsible for the large-scale field (i.e., that of the solar cycle) are rotationally influenced, whereas the influence of rotation on the smaller-scale granular and supergranular convection cells observed at the surface, which have a temporal scale that is very short compared to a solar rotation period, is negligible (see Cattaneo and Hughes, 2001).

Determination of the coefficients α and β

Even in the kinematic regime, in which the back-reaction of the Lorentz force is neglected, calculation of α_{ij} and β_{ijk} is not straightforward. Determination of \mathbf{b} from (50) is made difficult owing to the term $\nabla \times \mathbf{G}$, and it is therefore natural to look for circumstances under which this troublesome term can be neglected—an assumption sometimes referred to as the *quasi-linear approximation* or as *first order smoothing*. Let us define τ_c and l_c as the correlation time and correlation length of the turbulence with r.m.s. velocity v . A simple order of magnitude comparison of the terms in (50) shows that for conventional turbulent flows (for which $\tau_c \sim l_c/v$), $\nabla \times \mathbf{G}$ (and $\partial \mathbf{b} / \partial t$ also) can be neglected only provided that $Rm \ll 1$. An alternative scenario in which $\nabla \times \mathbf{G}$ (but not necessarily $\partial \mathbf{b} / \partial t$) can be neglected is the case of small Strouhal number, i.e., $S = v\tau_c/l_c \ll 1$. Under this latter premise two further distinctions can be made, depending on whether $S \ll Rm$ or $S \gg Rm$. For the first of these, putting aside for the moment any mathematical qualms we may have over dropping the diffusive term, α and β can be determined as

$$\alpha = -\frac{1}{3}\tau_c \langle \mathbf{v} \cdot \boldsymbol{\omega} \rangle, \quad \beta = \frac{1}{3}\tau_c \langle \mathbf{v}^2 \rangle. \quad (53)$$

As foreshadowed by the discussion above, α is dependent on the handedness of the flow, being directly proportional, and of opposite sign, to the helicity; β , on the other hand, depends not on any parity considerations but only on the kinetic energy of the flow.

It should be noted though that from an astrophysical standpoint, neither of the conditions $Rm \ll 1$ or $S \ll 1$ is ever satisfied; typically Rm is immense and, in what may be deemed as “conventional turbulence”, $S \sim O(1)$. Consequently, any results obtained under first order

smoothing should always be treated with some caution as to their validity when $Rm \gg 1$. Although an analytic theory for turbulence at high Rm remains elusive, it is however possible to make some headway for the case of perfectly conducting fluids (i.e., when η is zero and Rm is formally infinite), for which the magnetic field is frozen into the fluid. The field at any time t can be related to the field at some initial time ($t = 0$, say) through the Cauchy solution

$$B_i(\mathbf{x}, t) = B_j(\mathbf{a}, 0) \partial x_i / \partial a_j. \quad (54)$$

Formal substitution into the expression $\mathcal{E} = \langle \mathbf{v} \times \mathbf{b} \rangle$ then leads to the following expressions (Moffatt, 1978) for α and β (again assuming isotropy):

$$\alpha(t) = -\frac{1}{3} \int_0^t \langle \mathbf{v}^L(\mathbf{a}, t) \cdot \nabla_{\mathbf{a}} \times \mathbf{v}^L(\mathbf{a}, \tau) \rangle d\tau, \quad (55)$$

$$\begin{aligned} \beta(t) = & \frac{1}{3} \int_0^t \langle \mathbf{v}^L(\mathbf{a}, t) \cdot \mathbf{v}^L(\mathbf{a}, \tau) \rangle d\tau + \int_0^t \alpha(\tau) \alpha(\tau) d\tau \\ & + \frac{1}{6} \int_0^t \int_0^t \langle \mathbf{v}^L(t) \cdot \mathbf{v}^L(\tau_2) \nabla_{\mathbf{a}} \cdot \mathbf{v}^L(\tau_1) \\ & - (\mathbf{v}^L(t) \cdot \nabla_{\mathbf{a}} \mathbf{v}^L(\tau_1)) \cdot \mathbf{v}^L(\tau_2) \rangle d\tau_1 d\tau_2, \end{aligned} \quad (56)$$

where $\mathbf{v}^L(\mathbf{a}, t)$ is the Lagrangian representation of the velocity at time t of the fluid element located initially at $\mathbf{x} = \mathbf{a}$. As expected, the expression for α reflects the handedness of the flow (expressed now in a Lagrangian sense). Of more surprise is the expression for β ; the first term on the right-hand side of (56) is simply the effective turbulent diffusivity of a scalar field, whereas the second and third terms arise exclusively as a consequence of the vector character of \mathbf{B} . It is of interest to note that the expression for β contains products of α at different times, suggesting that the handedness of the flow may, at high Rm , be of significance in determining the diffusion of the magnetic field. A word of caution though is in order regarding expressions (55) and (56). First, there is no guarantee of convergence of the integrals contained in these expressions; furthermore, it is not clear if there is a unique interpretation of these results, owing to the fact that for a perfectly conducting fluid the initial state of the magnetic field is never forgotten. That said, they provide a potentially useful insight into the astrophysically relevant, and theoretically most challenging regime of $Rm \gg 1$.

3.2. Small-scale dynamo action

The exposition above has been based on the premise that the perturbed field \mathbf{b} owes its existence solely to the large-scale field \mathbf{B}_0 , and that, consequently, \mathcal{E} is a homogeneous linear functional of \mathbf{B}_0 and its derivatives. Recent studies have however revealed that turbulent flows (exhibiting exponential separation of particle trajectories) typically act as *small-scale* dynamos—with the magnetic field having scales comparable to or smaller than that of the driving flow—at sufficiently high values of Rm (see, for example, the monograph by Childress and Gilbert, 1995). Indeed, such flows (depending on their stretching and folding properties) can act as dynamos even in the limit of $Rm \rightarrow \infty$ (so-called *fast* dynamos). The case of astrophysical relevance is thus most likely to be that for which the growth of a large-scale magnetic field \mathbf{B}_0 is considered in the presence of a small-scale field that can exist independently of \mathbf{B}_0 . In this case, only part of the small-scale field \mathbf{b} in (50) owes its existence to \mathbf{B}_0 , and we expect \mathcal{E} to be an *inhomogeneous* linear functional of \mathbf{B}_0 . The possibility of small-scale dynamo action was clearly recognized by the pioneers of mean field dynamo theory (see, for example, Krause and Rädler, 1980) though it is only fairly recently that the pervasiveness of small-scale dynamo action at high Rm has been fully appreciated.

3.3. The nonlinear regime

Although obtaining an understanding of even the kinematic evolution of a 3D magnetic field is not straightforward—and remains far from complete—it is nonetheless important to address the problem of the nature of magnetic field transport in the nonlinear regime, i.e., when the back-reaction of the field on the flow cannot be neglected. The most important issue is to determine, for high values of Rm , the strength of the large-scale field at which α and β differ significantly from their kinematic values. A variety of approaches to this problem has been undertaken, based on the conservation laws of the ideal (diffusionless) system, on closure arguments, on numerical simulation of the governing equations or on some combination of these. The present state of play is that these differing approaches have not yet provided an agreed solution, making this one of the most controversial topics in astrophysical MHD. In this section we shall concentrate principally on the nature of the α -effect, the current understanding of turbulent diffusion in 3D being on even less solid ground. The physics of magnetic diffusion in 2D was extensively discussed in Section 2.

In order to calculate α it is sufficient to imagine an experiment in which homogeneous turbulence is permeated by a *uniform* magnetic field \mathbf{B}_0 . From simple physical considerations one expects that as the strength of the imposed field is increased, and the Lorentz forces become significant, the strength of the α -effect will be reduced—so-called α -*suppression*. Given the symmetry of the system under a change in sign of \mathbf{B}_0 , we may therefore expect, at high values of Rm , a dependence of the form

$$\alpha = \mathcal{F} \left(Rm^\gamma B_0^2 / B_E^2 \right), \quad (57)$$

where \mathcal{F} is a decreasing function of B_0^2 , and where B_E , denoting the equipartition strength of the large-scale field, is a useful reference measure of the field strength. The simplest such formula, which is often adopted, is

$$\alpha = \frac{\alpha_0}{1 + Rm^\gamma B_0^2 / B_E^2}, \quad (58)$$

where α_0 represents the kinematic value. The controversial nature of the subject resides in the value of the exponent γ . If γ is extremely small then large-scale fields close to equipartition strength can be generated before the α -effect loses its efficiency; conversely if γ is $O(1)$ then the generation term shuts down when the large-scale field is still extremely weak. Whereas certain theories of MHD turbulence may suggest the former alternative, numerical simulations—backed up by theoretical interpretation—point most decidedly to the latter.

The dependence of α on Rm and B_0 can be determined unambiguously—at least for a range of values of Rm and B_0 —by numerical solution of the nonlinear MHD equations for a turbulent flow permeated by a uniform magnetic field. Cattaneo and Hughes (1996) and Cattaneo *et al.* (2002) have considered this problem for a flow driven by helical forcing. It is worth reiterating that α is a statistical quantity and hence has a meaningful value only when averaged correctly; this point is illustrated in Fig. 3, which shows that although α clearly has a well-defined long-term mean, averaging over too short an interval could lead to quite erroneous results. The results of the simulations, illustrated in Fig. 4, are compatible only with an $O(1)$ value of the exponent γ , i.e., they show a dramatic α -suppression.

In a series of complementary calculations, Cattaneo *et al.* (2002) considered the evolution of a magnetic field of zero mean (i.e., no imposed field) in an extended spatial domain. Reassuringly, from the point of view of mean field theory, the nature of the α -effect driving the growth

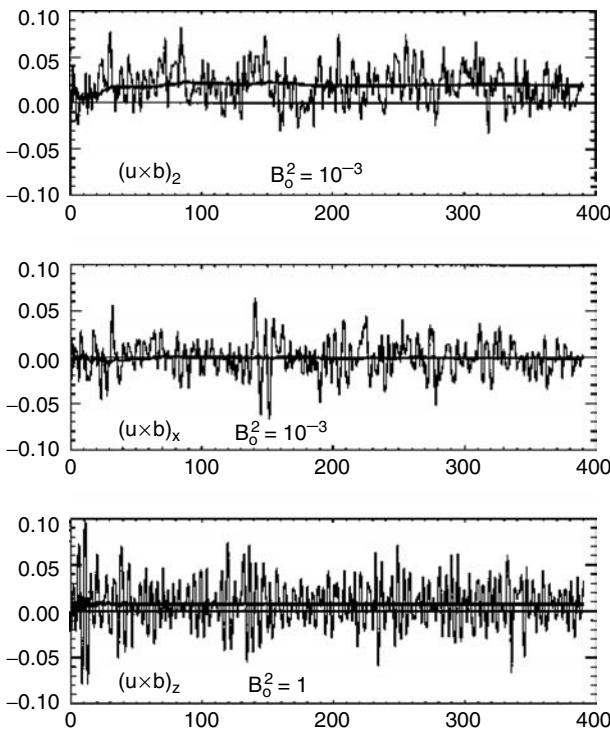


Figure 3. Time histories and time averages (thick lines) of components of the e.m.f. for a helically forced turbulent flow with an imposed uniform field (from Cattaneo and Hughes, 1996). In units of the equipartition strength, $B_0^2 = 10^{-3}$ in the uppermost panels, and $B_0^2 = 1$ in the lower panel. The e.m.f. (and hence α) has sizeable temporal variations, but a well-defined time average.

of the largest-scale field possible is entirely consistent with that derived from the calculations of α from an imposed uniform field; i.e., the growth of the large-scale field is halted at a very low value ($O(B_E/\sqrt{Rm})$).

It is of course important to address the physical cause of the drastic α -suppression found in these numerical simulations. As discussed earlier, for two-dimensional turbulence the suppression of β can be traced to the fact that the strong small-scale field imbues the fluid particles with a “memory” (see Cattaneo, 1994); this inhibits their tendency to disperse via random walking, and consequently reduces the diffusion of the magnetic field. For three-dimensional magnetic fields we envisage a similar physical picture. For high values of Rm , it is indubitable that strong small-scale fields ($O(Rm^{1/2}B_0)$) are generated, even while

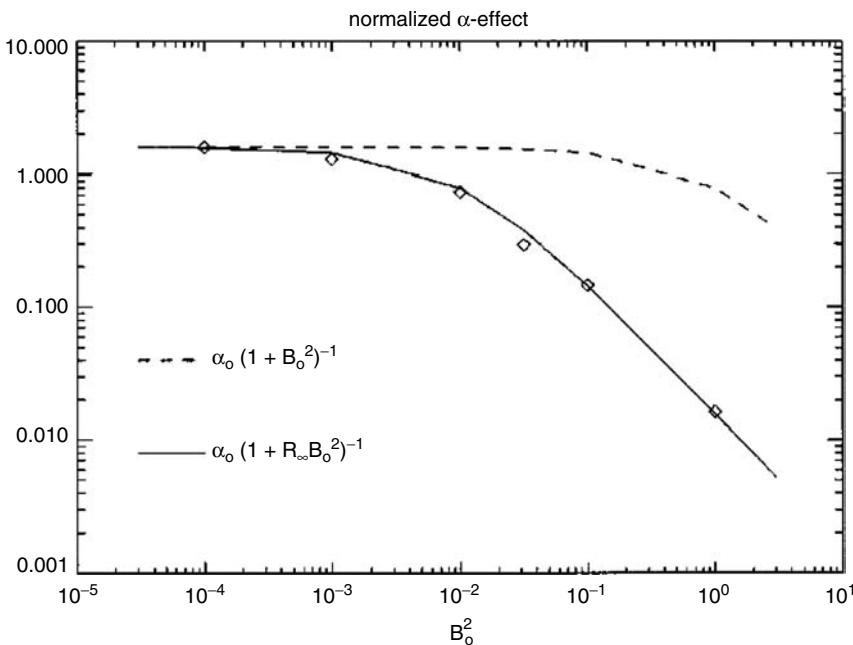


Figure 4. Normalized α -effect as a function of imposed field energy B_0^2 for a helically forced flow with $Rm = 100$ (from Cattaneo and Hughes, 1996). The dashed and solid lines are fits to the data for the two forms of α -quenching indicated.

the large-scale field is weak. Consequently, we may expect a marked reduction in Lagrangian transport properties and hence in α . It is though fair to say that a detailed numerical calculation of α via its Lagrangian properties has yet to be performed.

Whereas high-resolution numerical simulations point most definitely toward a dramatic (or even “catastrophic” in certain eyes) suppression of α , it is of course important to consider alternative approaches to the problem. One such approach, already discussed in Section 2, is through the use of turbulence closure models, following their success at reproducing many of the features of hydrodynamic (non-magnetic) turbulence. The most widely used scheme is the EDQNM model of Pouquet *et al.* (1976), the magnetic version of the scheme proposed by Orszag (1970). The key result of their analysis is the derivation of an expression for α in the following form

$$\alpha = -\frac{1}{3}\tau_c(\langle \mathbf{v} \cdot \boldsymbol{\omega} \rangle - \langle \mathbf{j} \cdot \mathbf{b} \rangle), \quad (59)$$

which subsequently has been widely used. It is however not only worth bearing in mind that this is a result borne of a number of approximations and assumptions—such as the assumption that the correlation times for the velocity and magnetic fields are the same—but it is also worth discussing how the result fits in with the classical α -effect picture described above. As discussed by Proctor (2003), the fact that the induction equation remains linear in the magnetic field—even though in the dynamic regime the flow is of course affected by the field—simply leads to the usual quasi-linear result (53). Any nonlinearity is simply manifested in a change to the kinetic helicity distribution. So what is the origin of the second term in (59)? If, instead of the classical picture of \mathbf{b} being dependent on \mathbf{B}_0 , we consider the introduction of a large-scale field \mathbf{B}_0 into a *pre-existing* state of MHD turbulence with a small-scale velocity \mathbf{v} and a small-scale field \mathbf{b} —leading to further perturbations \mathbf{v}' and \mathbf{b}' —then, under the quasi-linear approximation,

$$\mathcal{E} = \langle \mathbf{v} \times \mathbf{b}' \rangle + \langle \mathbf{v}' \times \mathbf{b} \rangle, \quad (60)$$

which, in combination with the momentum equation, leads to the result (59) (Pouquet *et al.*, 1976; Kleeorin and Ruzmaikin, 1982; Gruzinov and Diamond, 1994, 1996; Kleeorin and Rogachevskii, 1999; Proctor, 2003). It is though vitally important to be clear about the exact meanings of \mathbf{v} and \mathbf{b} in this formula. To obtain a further insight into the α -effect it is instructive to consider the ideal topological invariant (Gruzinov and Diamond, 1994, 1996). Whereas the physics of the diffusion of a magnetic field in two dimensions is underpinned by the conservation (in the absence of diffusion) of the mean-square potential, in three dimensions the conserved quantity is not $\langle \mathbf{A}^2 \rangle$, but the magnetic helicity $\langle \mathbf{A} \cdot \mathbf{B} \rangle$. The equations for \mathbf{a} and \mathbf{b} , the perturbations of the vector potential and the magnetic field, are

$$\frac{\partial \mathbf{a}}{\partial t} = (\mathbf{v} \times \mathbf{B}_0) + (\mathbf{v} \times \mathbf{b}) - \nabla \phi - \eta \nabla \times \mathbf{b}, \quad (61)$$

$$\frac{\partial \mathbf{b}}{\partial t} = \nabla \times (\mathbf{v} \times \mathbf{B}_0) + \nabla \times (\mathbf{v} \times \mathbf{b}) + \eta \nabla^2 \mathbf{b}. \quad (62)$$

If it is assumed that the small scales are stationary, such that $\partial \langle \mathbf{a} \cdot \mathbf{b} \rangle / \partial t = 0$, then (61) and (62), subject to reasonable boundary conditions, imply that

$$\mathbf{B}_0 \cdot \langle \mathbf{v} \times \mathbf{b} \rangle = \mathbf{B}_0 \cdot \mathcal{E} = -\eta \langle \mathbf{b} \cdot \nabla \times \mathbf{b} \rangle = -\eta \langle \mathbf{j} \cdot \mathbf{b} \rangle, \quad (63)$$

and consequently we have the *exact* result, dependent only on stationarity, that

$$\alpha = -\frac{\eta \langle \mathbf{j} \cdot \mathbf{b} \rangle}{B_0^2}, \quad (64)$$

where \mathbf{b} is the entire small-scale magnetic field. If we equate the two expressions for $\langle \mathbf{j} \cdot \mathbf{b} \rangle$ from (59) and (64) then we obtain the strong suppression result (58), though it is worth reiterating that in (64) \mathbf{b} refers to the total small-scale field, whereas in (59) it refers to a pre-existing small-scale field. It is interesting to note that (64) is the analogue for α , in 3D, of the Zeldovich theorem in 2D. In particular, it establishes a *direct* proportionality between α and the collisional resistivity, just as the Zeldovich theorem demonstrates the direct proportionality of η_T to η . Also, note that taking $\gamma = 1$ and $R_m B_0^2 / B_E^2 \gg 1$ in (58), along with the assumption of Alfvénized turbulence (so that $\mathbf{v} \sim \mathbf{b}$, $\omega \sim \mathbf{j}$), recovers (64).

3.4. The role of boundary conditions

In a series of papers, Blackman, Field and co-workers have raised a number of different—though sometimes self-contradictory—objections to the idea that the nonlinear dependence of the α -effect on the strength of the large-scale field should involve the magnetic Reynolds number in a critical manner. Field *et al.* (1999) claimed that the strong suppression result found by Cattaneo and Hughes (1996) was simply incorrect, though gave no explanation as to why they thought this might be the case. However it should be noted that a number of the assumptions of Field *et al.* (1999) are highly questionable and that, accordingly, the validity of their analysis is in doubt. A year later Blackman and Field (2000) changed the nature of their objection, arguing—in a sharp contradiction of Field *et al.* (1999)—that the strong (Rm -dependent) suppression was, after all, correct, but was however not applicable to astrophysical situations. Instead they claim that the suppression of α found in calculations such as those of Cattaneo and Hughes (1996) “is not a dynamic suppression” and that the suppression occurs entirely through the choice of periodic boundary conditions. The first of these assertions is patently false; the suppression results entirely from the action of the Lorentz force and is thus as dynamical as it can be! The second point, at least couched in the form that the choice of boundary conditions may be important, is however at least worthy of exploration. In Section 2 we considered this very issue for the simpler 2D case and showed how

the issue of strong suppression of turbulent diffusion could not easily be dismissed simply through changing the boundary conditions. Here we look at what is involved for the 3D case.

From Ohm's law we can readily derive (for uniform \mathbf{B}_0) the *exact* result

$$\alpha B_0^2 = \mathcal{E} \cdot \mathbf{B}_0 = -\frac{1}{\sigma}(\mathbf{j} \cdot \mathbf{b}) + \langle \mathbf{e} \cdot \mathbf{b} \rangle, \quad (65)$$

where σ is the electrical conductivity. As discussed above, under certain assumptions it is possible, simply from the $\langle \mathbf{j} \cdot \mathbf{b} \rangle$ term, to derive the strong suppression result. So the interesting question is whether this is the dominant term or if it can be eclipsed by $\langle \mathbf{e} \cdot \mathbf{b} \rangle$. From (61) and (62) it follows that the small-scale magnetic helicity satisfies

$$\frac{\partial \langle \mathbf{a} \cdot \mathbf{b} \rangle}{\partial t} = -2\langle \mathbf{e} \cdot \mathbf{b} \rangle + \langle \nabla \cdot (\mathbf{b}\phi) \rangle - \langle \nabla \cdot (\mathbf{a} \times \mathbf{e}) \rangle, \quad (66)$$

where ϕ is the electrostatic potential. The divergence terms can of course be readily transformed into surface integrals. For periodic boundary conditions the surface terms will vanish; it then follows that, for stationary turbulence, $\langle \mathbf{e} \cdot \mathbf{b} \rangle$ must also vanish. Thus, for stationary turbulence and periodic boundary conditions, α depends only on $\langle \mathbf{j} \cdot \mathbf{b} \rangle$. The thrust of Blackman and Field's argument appears to be that this is a rather special case and that, for other boundary conditions, $\langle \mathbf{e} \cdot \mathbf{b} \rangle$ will dominate and that the whole picture of α -suppression will be changed. Although this is an interesting suggestion it remains, at the moment, nothing more. For the 2D diffusion problem, discussed in Section 2, we saw that periodic boundary conditions are not particularly special and that the surface integral in question vanishes for most "reasonable" boundary conditions. Similarly, for the 3D problem the surface terms may still vanish for reasonable non-periodic conditions. Even more importantly, the mere fact that the surface integrals may not vanish does not, of itself, invalidate the strong-suppression results. Indeed it is by no means obvious, *a priori*, what the magnitude or even the sign of the surface terms will be. Furthermore, as discussed earlier, there is a good, *local* physical argument to explain the strong suppression of α . Any argument that claims the strong suppression result is an artefact of the choice of boundary conditions must also explain away the physical explanation of strong suppression. That said, the precise role of the boundary conditions remains an interesting issue that needs to be properly explored.

4. Momentum and flux transport in 2D MHD

4.1. Overview

In the previous sections, we have reviewed the status of mean field theory for flux diffusion in 2D MHD and for the α -effect in 3D MHD. Here we discuss the transport of momentum and magnetic potential in 2D MHD, incorporating a mean background shear flow. In view of the widespread occurrence of large-scale magnetic fields and shear flows in astrophysical objects—such as in the solar tachocline, accretion discs and galaxies—an outstanding problem in astrophysical MHD is to determine how these two structures influence one another; i.e., how magnetic fields alter the evolution of a mean shear flow via momentum transport, and how a shear flow affects the diffusion of magnetic fields via magnetic flux transport. The introduction of a mean shear flow into an MHD system presents us with rich and complex dynamics. Thus, for the sake of simplicity, the discussion here is limited to 2D MHD, with a mean shear flow parallel to the mean magnetic field; furthermore, a magnetic Prandtl number of unity is assumed. The related problem of momentum transport in 3D RMHD is treated by Kim *et al.* (2001). Given the constraints on space, we do not address the many works on the problem of momentum or angular momentum transport *per se*, as captured by the anisotropic kinetic alpha (AKA) effect (Frisch *et al.*, 1987), or the Λ -effect (Rüdiger, 1989; Kichatinov and Rüdiger, 1993).

4.2. Mean field theory

We consider forced 2D MHD turbulence in the (x, y) plane, in which energy is injected on small scales by an external forcing F . The evolution of the magnetic field \mathbf{B} and the fluid velocity \mathbf{v} , in terms of the vector potential A ($\mathbf{B} = \nabla \times A\hat{\mathbf{z}}$) and the vorticity ω ($\nabla \times \mathbf{v} = \omega\hat{\mathbf{z}}$), is described by the equations:

$$\left(\frac{\partial}{\partial t} + \mathbf{v} \cdot \nabla \right) \omega = -(\mathbf{B} \cdot \nabla) \nabla^2 A + v \nabla^2 \omega + F, \quad (67)$$

$$\left(\frac{\partial}{\partial t} + \mathbf{v} \cdot \nabla \right) A = \eta \nabla^2 A, \quad (68)$$

where F is the external small-scale forcing.

We assume that both the mean shear flow \mathbf{V} and mean magnetic field \mathbf{B} are in the y -direction, with $\mathbf{V} = V(x)\hat{\mathbf{y}}$ and $\mathbf{B} = B\hat{\mathbf{y}}$ (or $A = A(x)$).

Adopting a two-scale analysis, we decompose fields into mean and fluctuating components as $\mathbf{v} = \langle \mathbf{v} \rangle + \mathbf{v}' = \mathbf{V} + \mathbf{v}'$, $\omega = \langle \omega \rangle + \omega' = \partial V / \partial x + \omega'$, $\mathbf{b} = \langle \mathbf{b} \rangle + \mathbf{b}' = \mathbf{B} + \mathbf{b}'$ and $a = \langle a \rangle + a' = A + a'$, where angular brackets denote an average over the statistics of the forcing. Note that in this section, the mean fields and fluctuations are denoted by capital letters and primes, respectively. Employing a quasi-linear closure then gives the equations for fluctuations as

$$\left(\frac{\partial}{\partial t} + V \frac{\partial}{\partial y} - v \nabla^2 \right) \omega' = - \left(B \frac{\partial}{\partial x} \right) \nabla^2 a' + F, \quad (69)$$

$$\left(\frac{\partial}{\partial t} + V \frac{\partial}{\partial y} - \eta \nabla^2 \right) a' = - \left(v'_x \frac{\partial}{\partial x} \right) A. \quad (70)$$

The fluctuating fields influence the evolution of mean fields via the fluxes of magnetic potential $\langle v'_x a' \rangle$ and momentum $\langle v'_x v'_y - b'_x b'_y \rangle$, which appear in the mean field equations as follows:

$$\left(\frac{\partial}{\partial t} - v \frac{\partial^2}{\partial x^2} \right) V = - \frac{\partial}{\partial y} \langle \Pi \rangle - \frac{\partial}{\partial x} \langle v'_x v'_y - b'_x b'_y \rangle, \quad (71)$$

$$\left(\frac{\partial}{\partial t} - v \frac{\partial^2}{\partial x^2} \right) A = - \frac{\partial}{\partial x} \langle u'_x a' \rangle, \quad (72)$$

where Π is the total pressure. The flux of magnetic potential $\langle v'_x a' \rangle$ in (72), which was discussed in Section 2, represents the effective dissipation rate of a mean magnetic field due to small-scale fluctuations, and can be expressed as $\langle v'_x a' \rangle = -\eta_T \partial A / \partial x$, on invoking a turbulent diffusivity η_T . As noted earlier, it consists of two competing processes, which transport magnetic potential to small and large scales via fluid advection and the Lorentz force, respectively. With energy being injected into the fluid, as assumed here, the former process wins, resulting in the overall dissipation of a mean magnetic field through a positive η_T . The dissipation of a mean magnetic field is in fact a natural consequence of the absence of dynamo action in 2D MHD. However, if there is a magnetic forcing in a system, for instance via winds from stars carrying magnetic fields, negative η_T (i.e., an inverse transfer of magnetic potential) is possible (Kim and Dubrulle, 2002).

The momentum flux or total stress $\langle v'_x v'_y - b'_x b'_y \rangle$ in (71) represents the average y -component of flux of the x -component of momentum and captures the overall effect of small-scale fluctuations on the evolution of the mean shear flow V . Acting as an effective force on the mean shear flow, it modifies the mean profile of background shear; it can be put in

the form of a turbulent viscous force by introducing a transport coefficient, the so-called turbulent viscosity ν_T , as $\langle v'_x v'_y - b'_x b'_y \rangle = -\nu_T \partial V / \partial x$. It can, in general, take either sign, being negative in the case of an inverse cascade and positive for a direct cascade. Note that momentum transport involves the Maxwell stress $\langle b'_x b'_y \rangle$ as well as the usual fluid Reynolds stress $\langle v'_x v'_y \rangle$.

In order for the two turbulent transport coefficients (η_T and ν_T) not to vanish in 2D MHD, non-ideal effects (i.e., irreversibility) in the system are absolutely crucial. One obvious example of such an effect is dissipation, which makes the system deviate from the Alfvénic state into which \mathbf{B} naturally forces it. The flux of magnetic potential, based on this effect, was reviewed earlier. The presence of a shear flow brings in another non-ideal effect through resonance between the flow and fluctuations (i.e., critical layers). As we shall see, transport of magnetic potential and momentum is reduced by a shear flow as well as by magnetic fields, although the presence of a shear flow itself is critical to a non-vanishing flux of momentum. As the net effect of either a shear flow or a magnetic field on momentum and flux transport is difficult to ascertain, we first discuss the effect of a shear flow on transport in general, and how it may be incorporated non-perturbatively, before discussing transport in 2D MHD with shear (e.g., the effect of shear flow on flux diffusion and the effect of magnetic fields on momentum transport). The results are summarized in [Table 2](#).

4.3. The effect of shear flows on transport

A shear flow acts to tilt and elongate eddies, resulting in finer scales as time progresses (see [Fig. 5](#)); i.e., for a flow $V(x)\hat{y}$, the wavenumber k_x grows linearly in time as $k_x(t) = k_x(0) - tk_y \partial V / \partial x$ with constant $k_y(t) = k_y(0)$ (Goldreich and Lynden-Bell, 1965). For an incompressible fluid, as assumed here, fluid velocities perpendicular to the shear are

Table 2. Summary of η_T and ν_T for turbulence with background shear

$\nu = \eta$	Strong shear ($\xi \gg 1$)	Weak shear ($\xi \ll 1$)
2D HD	$\nu_T \propto -1/\nu^2$	$\nu_T \propto -1/\Omega^2$
2D MHD ($\gamma \rightarrow 0$)		
Kinematic limit	$\eta_T \propto 1/\nu^2$	$\eta_T \propto 1/\Omega^2$
2D MHD ($\gamma \gg 1$)	$\eta_T \propto 1/\nu^2$	$\eta_T \propto (\nu/\Omega)^{2/3}/B^2$
Strong B	$\eta_T \propto 1/B^2$	$\eta_T \propto 1/B^2$

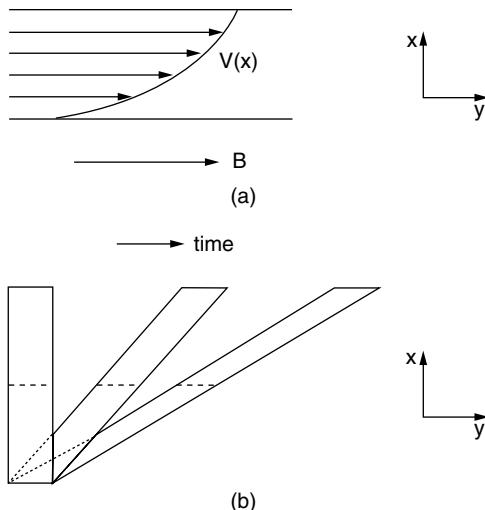


Figure 5. Panel (a) depicts the configuration of a mean shear flow $V(x)\hat{y}$ and a mean magnetic field $B\hat{y}$. Panel (b) illustrates the tilting of an eddy by a shear $V(x)\hat{y}$. The solid and dashed lines represent an eddy in the absence and presence of the field $B\hat{y}$. The x -extent of the eddy is smaller in the presence of $B\hat{y}$ since the Lorentz force prevents the eddy motion in the x direction. Tilting elongates an eddy in the y direction, generating small scales in the x direction as time progresses.

smaller than those along the flow, effectively reducing the perpendicular transport of scalar fields. Furthermore, as the perpendicular scale (i.e., that in the x -direction) decreases, the eddies will eventually be torn apart by dissipation, again inhibiting the transport in the x -direction. The reduction of transport by shearing is a common phenomenon, occurring in various physical systems, such as heat transport in geophysical convection (Or and Busse, 1987), and particle and heat transport in magnetically confined plasmas (Biglari *et al.*, 1990; Diamond *et al.*, 1998; Kim and Diamond, 2003).

Owing to the generation of fine scales, a careful, non-perturbative analysis is desirable to capture the effects of shearing. For instance, the effect of shearing on dissipation can be critical to determining the transport, since the overall dissipation, due to this shearing, increases in time. Thus, even if the dissipation may be negligible at some instant, this may not be the case for all subsequent times. The linear increase in wavenumber can be best incorporated non-perturbatively by following a particle trajectory in an extended phase space $(\mathbf{x}, \mathbf{k}, t)$ along which $k_x(t)$ evolves as $k_x(t) = k_x(0) + t k_y \partial V / \partial x$. As shall be shown shortly, this can be achieved via the Gabor transform (a kind of wavelet transform),

defined as:

$$\text{GT}[A(\mathbf{x}, t)] \equiv \hat{A}(\mathbf{k}, \mathbf{x}, t) \equiv \int d^2x' f(|\mathbf{x} - \mathbf{x}'|) e^{i\mathbf{k} \cdot (\mathbf{x} - \mathbf{x}')} A(\mathbf{x}', t), \quad (73)$$

where $f(x)$ is a filter function with a characteristic scale λ lying between the characteristic scales of fluctuating and mean fields, thus screening out information outside a domain of size λ . For simplicity, we adopt a Gaussian filter with $f(x) = \exp(-x^2/\lambda^2)$.

In terms of Gabor transforms, (69) and (70) for the fluctuations can be written as

$$\left[\frac{D}{Dt} + v(k^2 + p^2) \right] \hat{\omega} = iBk(k^2 + p^2)\hat{a} + \hat{F}, \quad (74)$$

$$\left[\frac{D}{Dt} + \eta(k^2 + p^2) \right] \hat{a} = \frac{ik}{k^2 + p^2} \hat{\omega} B = \hat{u}_y B. \quad (75)$$

Here, $\mathbf{k} = (p, k, 0)$ and $\hat{\mathbf{v}}(\mathbf{x}, \mathbf{k}, t) = (\hat{v}_x, \hat{v}_y, 0)$; $D/Dt \equiv \partial/\partial t + V \partial/\partial y - k \partial/\partial x (V \partial/\partial p) = \partial/\partial t + V \partial/\partial y + k \Omega \partial/\partial p$ is the total time derivative, which includes the linear increase of k_x in time, i.e., $Dy/Dt = V$, $Dx/Dt = 0$, $Dk/Dt = 0$, $Dp/Dt = k\Omega$; without loss of generality, $\Omega = -\partial V / \partial x$ is assumed to be positive. Thus, ray equations along particle trajectories are simply $x = x_0$, $y = y_0 + V(t - t_0)$, $p = p_0 + k\Omega(t - t_0)$, and $k = k_0$, where a subscript “0” denotes the initial value at $t = t_0$. Equations (74) and (75), together with the ray equations, describe the motions of the wave packets of vorticity and magnetic potential in phase space, under the action of large-scale fields. The center of the wave packet is advected at the mean velocity, while its wavenumber is varied according to the local shear, with its amplitude changed by mutual interactions and forcing.

The turbulent transport coefficients follow from the solutions to (74) and (75), which can be found along a particle trajectory for a given forcing F . Therefore, the solutions for ω' and a' , and consequently the values of the transport coefficients η_T and v_T , depend on the properties of the forcing. While it is interesting to study these dependencies, the following discussion focuses on only one special, but non-trivial case, for which the statistics of the forcing are homogeneous in space with an infinitesimally short temporal correlation time (δ -correlated forcing). One interesting property of δ -correlated forcing containing the entire range of frequencies in the spectrum, is that it allows resonance between a shear and fluctuations over a wide range of frequencies. Before discussing the effect of shear and magnetic fields on η_T and v_T it is illuminating to study how shearing affects transport in general, through the examples of the transport of passive scalar fields and momentum transport in 2D hydrodynamic turbulence (HD).

Transport of passive scalar fields: Kinematic limit of 2D MHD

As an instructive example of the reduction of the transport of passive scalar fields, let us consider the transport of magnetic potential in 2D MHD in the kinematic limit, in which the field is passively advected; i.e., we consider the coupled equations (67) and (68) (or (74) and (75)) but neglect the Lorentz force term. Assuming that the system settles into a stationary state in the long-time limit, we can investigate how the flux of magnetic potential $\langle v'_x a \rangle$ depends on the shear Ω in this stationary state.

Depending on timescales, there are two interesting cases to consider. The first is that of a weak shear, with $\xi = v k^2 / \Omega \gg 1$, for which the effect of dissipation dominates that of shearing, with the dissipation timescale $\tau_v = 1/vk^2$ being much smaller than the characteristic shear timescale, $\tau_s = 1/\Omega$; here k refers to the wavenumber of the prescribed forcing. The second is the case of strong shear, with $\xi = v k^2 / \Omega \ll 1$, in which the effect of shearing dominates that of dissipation. Note that $\xi \ll 1$ can be satisfied even in the long-time limit, despite the generation of fine scales for a' and ω' .

For δ -correlated forcing, turbulent diffusivity in the weak shear case $\xi \gg 1$ is given by

$$\eta_T = \frac{1}{4(2\pi)^2} \int d^2k \frac{\hat{\psi}(\mathbf{k})}{v^2 k^6} \left[1 - \frac{4\Omega^2}{(vk^2)^2} \right], \quad (76)$$

while in the strong shear case $\xi \ll 1$, it has the form

$$\eta_T = \frac{1}{32\Omega^2} \int d^2k \frac{\hat{\psi}(\mathbf{k})}{k^2}, \quad (77)$$

where $\hat{\psi}(\mathbf{k})$ is the power spectrum in Fourier space of the forcing F . In both limits, the turbulent diffusivity, being positive, is a decreasing function of Ω , confirming the reduction of flux transport by shear. Note that in the strong shear case, the scaling of $\eta_T \propto \Omega^{-2}$ in (77) is a consequence of the δ -correlated forcing, and may thus change for a forcing with different statistical properties.

Momentum transport in 2D HD

Shear also leads to a reduction in the transport of a vector quantity such as momentum. For example, in 2D HD, the turbulent viscosity in the strong shear limit $\xi \gg 1$ is given by

$$\nu_T \sim -\frac{1}{2\Omega^2(2\pi)^2} \int d^2k \frac{\hat{\phi}(\mathbf{k})}{k^2}. \quad (78)$$

The negative sign of v_T is the manifestation of an inverse cascade of energy in 2D HD due to the conservation of enstrophy; the amplitude of v_T , decreasing with Ω , indicates the suppression of the momentum flux by a strong shear. Observe that when the Ω^{-2} dependence of v_T is substituted into the mean field equation (71), one recovers the logarithmic equilibrium profile for a mean shear flow (Kim and Dubrulle, 2001). Note that the foregoing calculations assumed a (local) periodic box for fluctuations, while proper wall boundary conditions for a mean flow in (71) are critical to obtaining this equilibrium profile.

4.4. The effect of shear flows on flux diffusion in 2D MHD

As discussed in Section 2, flux diffusion is reduced owing to the back-reaction of the Lorentz force, with an enormous suppression factor for large magnetic Reynolds R_m . The presence of a shear flow introduces an additional non-ideal effect (i.e., irreversibility) via the stochasticity of fluid elements in the presence of resonance between a shear flow and fluctuations—*critical layers*—and the overlap of these layers. Note that this stochasticity may justify our quasi-linear analysis. In the strong shear limit, this resonant absorption leads to flux transport that decreases with increasing shear. Nevertheless, being independent of R_m , this additional effect offers the possibility of weakening the strong dependence of the flux diffusion on R_m .

The exact form of flux diffusion, which depends on magnetic fields and shear, can be obtained from (74) and (75). While it is a formidable task to obtain general solutions to these equations, simple analytical solutions are available in the limit of a strong magnetic field, for which the Alfvén wave timescale of the k mode (associated with a mean magnetic field) is smaller than the timescale for shearing, i.e., $\gamma = |Bk/\Omega| \gg 1$. Once such solutions are found, they can be used to obtain the turbulent diffusivity, yielding the results:

$$\eta_T = \frac{1}{4(2\pi)^2} \int d^2 k \frac{\hat{\phi}(\mathbf{k})}{k^2 + p^2} \frac{1}{(\nu k^2)^2} \frac{1}{1 + \Lambda}, \quad (79)$$

for the weak shear case with $\xi \gg 1$, and

$$\eta_T = \frac{\nu}{4B^2(2\pi)^2} \int d^2 k \frac{\hat{\phi}(\mathbf{k})}{k^2 + p^2} \frac{2}{3} \Gamma\left(\frac{1}{3}\right) \left(\frac{3}{2\nu k^2}\right)^{1/3} \Omega^{-2/3}, \quad (80)$$

for the strong shear case $\xi \ll 1$. Here, $\Lambda = (B/\nu k)^2$, $\Gamma(x)$ is the Gamma function, and $\hat{\phi}(\mathbf{k})$ is the power spectrum of the forcing, which is

assumed to be homogeneous, but not necessarily isotropic. In both cases, the turbulent diffusivity is positive to leading order, becoming small as the magnetic field B becomes strong. As discussed previously, this is due to the Alfvénization of turbulence by a strong field. It may also be viewed as the inhibition of the eddy motion in the y direction by magnetic tension, effectively reducing the effect of tilting by the shear, as demonstrated by the dashed line in Fig. 5b. In the weak shear limit $\xi \gg 1$, the diffusion of magnetic field is dominated by the strong field, with (79) coinciding with the kinematic result (76) to leading order for $\Lambda \ll 1$.

In the more interesting case of a strong shear, the scaling of $\eta_T \propto \Omega^{-2/3}B^{-2}$ in (80) reveals the interesting, combined effect of shear and magnetic field on the turbulent diffusion of magnetic field. Note that this particular scaling is a consequence of the assumption $\gamma \gg 1$ (which implies the dominance of the effect of magnetic field compared to the shearing), possibly as well as the assumption of δ -correlated forcing. As η_T is primarily suppressed by a magnetic field, the dependence of η_T on Ω is much weaker than that in the kinematic case (see (77)) where $\eta_T \sim \Omega^{-2}$. Nevertheless, in this limit, the amplitude of η_T is smaller than the kinematic value, roughly by a factor of $\xi^{2/3}/\gamma^2$. Note that this factor is proportional to $V^2/B^2R_m^{2/3}$, where V is the characteristic velocity. The R_m dependence of this suppression factor is weak compared to the case without shear, as shown below.

In the absence of a shear flow, the usual quasi-linear analysis of (69) and (70) via Fourier analysis yields the kinematic turbulent diffusivity

$$\eta_k \sim \frac{\tau_f \langle F^2 \rangle}{8\nu^2 k_0^6} \sim \frac{\langle v^2 \rangle}{2\nu k_0^2}. \quad (81)$$

Here, τ_f and k_0 are the correlation time and characteristic scale of the forcing F , respectively, and the relation $\langle v^2 \rangle \sim \tau_f \langle F^2 \rangle / 4\nu k_0^4$ has been used. On the other hand, in the limit of a strong magnetic field, the turbulent magnetic diffusivity is given by

$$\eta_T \sim \frac{\tau_f \langle F^2 \rangle}{4k_0^4 B^2} \sim \frac{\eta \langle v^2 \rangle}{B^2}. \quad (82)$$

Thus, without shear, η_T is reduced by a factor of $2(\eta k_0/B)^2 = (\xi/\gamma)^2 \sim (U/BR_m)^2$ by a magnetic field (U is a typical velocity). Here, note that compared to the suppression factor of v^2/B^2R_m in Kim (1999), the extra factor of $R_m (= R_e)$ comes from the R_e dependence of the kinematic value of turbulent diffusivity (recall, $\eta = \nu$ is assumed), and has nothing to do with a dynamical effect of the Lorentz force. Therefore, $\eta_T \propto R_m^{-3/2}$

with a shear flow and $\eta_T \propto R_m^{-2}$ without it. Therefore, the presence of a shear flow weakens the R_m dependence of η_T through resonance.

We may check that the Zeldovich theorem $\eta_T = \eta \langle b^2 \rangle / B^2$ is valid in the strong shear case (recall $v = \eta$) (Zeldovich, 1957)—as indeed it must be. To see this, we note that, to leading order, kinetic $\langle v'^2 \rangle$ and magnetic energies $\langle b^2 \rangle$ are given by

$$\langle v'^2 \rangle = \langle b^2 \rangle = \frac{1}{6(2\pi)^2 \Omega} \int d^2 k \frac{\hat{\phi}(\mathbf{k})}{k^2} \Gamma\left(\frac{1}{3}\right) \left(\frac{3\Omega}{2v k^2}\right)^{1/3}. \quad (83)$$

The divergence of $\langle v'^2 \rangle$ and $\langle b^2 \rangle$ as $v \rightarrow 0$ is caused by the accumulation of energy on small scales due to the direct cascade of energy with the (small-scale) forcing. Therefore, $\eta_T = \eta \langle b^2 \rangle / B^2$, with the Zeldovich theorem remaining valid for a strong shear. Note that the weaker dependence of η_T on R_m results from the fact that $\langle b^2 \rangle$ itself depends on η , diverging as $\eta \rightarrow 0$.

4.5. Effect of magnetic fields and shear on momentum transport

We now discuss the effect of both magnetic fields and shear flows on momentum transport, starting with the effect of magnetic fields. In 2D HD, enstrophy is an ideal invariant, causing an inverse cascade of energy. Introducing magnetic fields into the 2D HD system breaks this conservation law, since vorticity can be generated by the Lorentz force. Therefore, one of the important effects of magnetic field on momentum transport is to alter the direction of the energy cascade. The other interesting consequence of including magnetic fields is the change in the dependence of the momentum flux with shear. In 2D HD the momentum flux is reduced solely by the shear, while in 2D MHD it is suppressed by both shear and magnetic field. The reduction in momentum transport by a mean magnetic field is due to the Alfvénization of turbulence, and is indicated by the appearance, with opposite signs, of both the Reynolds and Maxwell stresses in the momentum flux. It is reminiscent of the flux of magnetic potential $\Gamma_A \propto \langle v'^2 - b^2 \rangle$, for which perfect Alfvénization leads to $\Gamma_A = 0$. Thus, a significant reduction in momentum transport is possible via a cancelation between Reynolds and Maxwell stresses. In fact, for Alfvén waves, a perfect cancelation between the two stresses is expected. Obviously, the presence of a shear flow (which is necessary for momentum transport in the first place) breaks this perfect Alfvénic state, and leads to a finite momentum transport. As a matter of fact,

in the strong shear limit, each of the Reynolds and Maxwell stresses diverges in the ideal limit, but the total stress remains finite owing to the cancelation between the two stresses. The deviation from a pure Alfvénic state can also be achieved by the incorporation of dissipation in the system.

Expressions for the turbulent viscosity follow from the solutions to the coupled equations (74) and (75) with the following results:

- (i) In the weak shear case $\xi \gg 1$:

$$\nu_T = \frac{1}{4B^2(2\pi)^2} \int d^2k \frac{\hat{\phi}(\mathbf{k})}{k^2(k^2 + p^2)} \frac{\Lambda(\Lambda - 1)}{(1 + \Lambda)^2}, \quad (84)$$

where $\Lambda \equiv (B/vk)^2$. The direct cascade of energy is already indicated in this weak shear case by a positive ν_T when $\Lambda \gg 1$, in contrast to a negative ν_T in 2D HD. As noted previously, this is a consequence of the Lorentz force, which relaxes the conservation of vorticity constraint in 2D HD, thereby reversing the direction of the energy cascade from inverse to direct. The cancelation between Reynolds and Maxwell stresses, as a result of Alfvénization, is suggested in the amplitude of the turbulent viscosity, which becomes small for a strong magnetic field.

- (ii) In the strong shear case $\xi \ll 1$:

$$\nu_T = \frac{1}{4B^2(2\pi)^2} \int d^2k \frac{\hat{\phi}(\mathbf{k})}{k^2(k^2 + p^2)}, \quad (85)$$

which is now always positive, indicating the direct cascade of energy for a strong shear. Furthermore, it is independent of the shear. Note that while $\xi < 1$ for shearing to be of interest (Alfvén frequency less than shear rate), γ can be either greater or less than unity. For $\gamma \gg 1$, the regime of strong magnetization, it is thus not surprising that suppression of ν_T occurs primarily via Alfvénic coupling. When the expression for ν_T is substituted into (71), an equilibrium mean shear flow is found to have either parabolic or linear profiles, depending on the boundary condition for the average total pressure $\langle \Pi \rangle$. Therefore, another important effect of a magnetic field is to change the profile of a mean equilibrium shear flow from logarithmic to parabolic or linear. This is suggestive of the so-called *buffer layer*, which has been studied in the context of turbulent drag reduction

in various laboratory experiments (Tsinober, 1989). Note that turbulent viscosity here is the same as that in the weak shear case for $\Lambda \gg 1$, simply because the magnetic field is the main source for the suppression of momentum transport in both cases.

4.6. Concluding remarks

Mean magnetic fields and shear flow are ubiquitous structures in many astrophysical objects, and, as such, the problems of the evolution of these mean magnetic fields (dynamos or diffusion) and of the transport of (angular) momentum are by far two of the most important issues in astrophysical MHD. While it is necessary to consider fully 3D MHD, the primary focus of this section has been to elucidate the role of mean shear flows and magnetic fields on momentum transport and magnetic diffusion, by considering 2D MHD with a mean flow parallel to the magnetic field. The main conclusions are:

- (i) magnetic fields have a significant effect on momentum transport—via Alfvénization—leading to the suppression of momentum transport and to laminarization of an equilibrium shear flow;
- (ii) a shear flow can weaken (slightly) a problematic strong R_m dependence of magnetic flux diffusion by introducing a route to collisionless irreversibility via resonance overlap.

These results may have significant implications for the solar tachocline where a mean shear flow (provided by the radial differential rotation) is aligned with a strong toroidal magnetic field (see the review by Tobias, 2004, Chapter 7 in this volume). In the solar tachocline, $\gamma = Bk/\Omega \gtrsim 1$ since $B \sim 10^4 - 10^5 \text{ G}$, $\Omega \sim V/L \sim 10^{-6} \text{ s}^{-1}$ ($L \sim 10^{10} \text{ cm}$ and $V \sim 10^4 \text{ cm/s}$), and $k > 10^{-10} \text{ cm}^{-1}$. Thus, if we take the molecular values for ν and η as $\nu \sim 10^4 \text{ cm}^2/\text{s}$ and $\eta \sim 10^2 \text{ cm}^2/\text{s}$, our results obtained in the strong shear limit ($\xi = \nu k^2/\Omega \ll 1$), together with the assumption $\gamma \gg 1$, are applicable in this region, although the assumption of unit magnetic Prandtl number ($\nu = \eta$) is not rigorously justified. Indeed, the solar tachocline is a very interesting site for both dynamo action and angular momentum transport. In order to combat the problems of diffusivity quenching by a strong field, discussed earlier, Parker (1993) proposed the idea of an *interface dynamo*, in which the sites of generation of toroidal field (via a velocity shear) and poloidal

field (via the α -effect) are spatially separated, but coupled by turbulent diffusion. Obviously, in the light of the results outlined in this section, it would be of interest to investigate this coupling further, by quantifying the effect of the velocity shear on the diffusion of magnetic field. Furthermore, a possible implication of the result (ii) for 3D MHD is that the incorporation of a shear flow may weaken the notorious α quench. This issue should be investigated by extending the analysis to three dimensions.

Acknowledgments

The authors wish to thank F. Cattaneo, A. Das, B. Dubrulle, A.V. Gruzinov, M. Malkov, A. Pouquet, M.R.E. Proctor and S.M. Tobias for many stimulating conversations. PHD and EK are supported by the U.S. Department of Energy under Grant No. FG03-88ER 53275.

Appendix A: Derivation of the Zeldovich theorem

The Zeldovich theorem is, in essence, a statement of the balance between transport of magnetic flux and resistive dissipation for a stationary, 2D magnetofluid. The Zeldovich theorem may be derived from the evolution equations for the magnetic potential fluctuation A , namely

$$\frac{\partial A}{\partial t} + \mathbf{v} \cdot \nabla A = -v_x \frac{\partial \langle A \rangle}{\partial x} + \eta \nabla^2 A. \quad (\text{A.1})$$

Multiplying by A and summing over space for an incompressible flow yields:

$$\frac{1}{2} \left(\frac{\partial}{\partial t} \langle A^2 \rangle + \langle \nabla \cdot (\mathbf{v} A^2) \rangle \right) = -\langle v_x A \rangle \frac{\partial \langle A \rangle}{\partial x} - \eta \langle B^2 \rangle. \quad (\text{A.2})$$

Here, we assume that the spatial variation of the fluctuations is faster than that of the mean potential $\langle A \rangle$, so that $\partial \langle A \rangle / \partial x$ falls outside the brackets in the first term on the RHS of (A.2). For a periodic domain, or one for which $v_n = 0$ on the boundaries, and a stationary state, (A.2) may be re-written as

$$\langle B^2 \rangle = \frac{-\langle v_x A \rangle}{\eta} \frac{\partial \langle A \rangle}{\partial x} = -\frac{\Gamma_A}{\eta} \frac{\partial \langle A \rangle}{\partial x}. \quad (\text{A.3})$$

Finally, writing Γ_A in Fick's law form yields

$$\langle B^2 \rangle = \frac{\eta_T}{\eta} \left(\frac{\partial \langle A \rangle}{\partial x} \right)^2 = \frac{\eta_T}{\eta} \langle B \rangle^2. \quad (\text{A.4})$$

Equation (A.4) is effectively the statement of the Zeldovich theorem.

Equation (A.4) has several interpretations and implications. First, it indicates that the effective turbulent resistivity η_T must scale directly with the collisional resistivity η , in proportion to $\langle B^2 \rangle / \langle B \rangle^2$. This, of course, is a straightforward consequence of the freezing-in law, to which the magnetic potential evolution equation is equivalent. Second, (A.4) states that the mean-square magnetic fluctuation $\langle B^2 \rangle$ level can be large even if the mean magnetic field $\langle B \rangle$ is weak. Note that $\eta_T / \eta \sim (R_m)_{\text{eff}} \gg 1$, so that $\langle B^2 \rangle / \langle B \rangle^2 \gg 1$. Third, (A.4) may be viewed as a statement of Prandtl mixing-length theory for magnetic potential. This follows from the fact that it states an equality between the decay rate of the mean potential ($\sim \eta_T (\partial \langle A \rangle / \partial x)^2$ — i.e., the dissipation rate on large scales) and the decay rate of the potential fluctuations ($\sim \eta \langle \nabla A^2 \rangle = \eta \langle B^2 \rangle$ — i.e., the dissipation rate on small scales). Such a relation constitutes an important constraint on the mean magnetic flux transport, Γ_A .

It is useful to mention here that the Zeldovich theorem is a very robust result, which persists in the presence of a mean shear flow, etc. This is, of course, a consequence of the fact that it is basically a straightforward consequence of magnetic flux conservation, or, equivalently, the freezing-in law.

References

- Bhattacharjee, A. and Hameiri, E., "Self-consistent dynamo-like activity in turbulent plasmas," *Phys. Rev. Lett.* **57**, 206–209 (1986).
- Bhattacharjee, A. and Yuan, Y., "Self-consistent constraints on the dynamo mechanism," *Astrophys. J.* **449**, 739–744 (1995).
- Biglari, H., Diamond, P. H. and Terry, P. W., "Influence of sheared poloidal rotation on edge turbulence," *Phys. Fluids B* **2**, 1–4 (1990).
- Biskamp, D. and Welter, H., "Dynamics of decaying two-dimensional magnetohydrodynamic turbulence," *Phys. Fluids B* **1**, 1964–1979 (1989).
- Blackman, E. G. and Field, G. F., "Constraints on the magnitude of alpha in dynamo theory," *Astrophys. J.* **534**, 984–988 (2000).
- Blackman, E. G. and Field, G. F., "New dynamical mean field dynamo theory and closure approach," *Phys. Rev. Lett.* **89**, 265007-1–265007-4 (2002).
- Brandenburg, A., "The inverse cascade and nonlinear alpha-effect in simulations of isotropic helical hydromagnetic turbulence," *Astrophys. J.* **550**, 824–840 (2001).

- Cattaneo, F., “On the effect of a weak magnetic field on turbulent transport,” *Astrophys. J.* **434**, 200–205 (1994).
- Cattaneo, F. and Hughes, D. W., “Nonlinear saturation of the turbulent α -effect,” *Phys. Rev. E* **54**, R4532–R4535 (1996).
- Cattaneo, F. and Hughes, D. W., “Solar dynamo theory: a new look at the origin of small-scale magnetic fields,” *Astron. Geophys.* **42**, 3.18–3.22 (2001).
- Cattaneo, F. and Vainshtein, S. I., “Suppression of turbulent transport by a weak magnetic field,” *Astrophys. J.* **376**, L21–L24 (1991).
- Cattaneo, F., Hughes, D. W. and Kim, E., “Suppression of chaos in a simplified nonlinear dynamo model,” *Phys. Rev. Lett.* **76**, 2057–2060 (1996).
- Cattaneo, F., Hughes, D. W. and Thelen, J.-C., “The nonlinear properties of a large-scale dynamo driven by helical forcing,” *J. Fluid Mech.* **456**, 219–237 (2002).
- Childress, S. and Gilbert, A. D., *Stretch, Twist, Fold: The Fast Dynamo*, Springer-Verlag, Berlin (1995).
- Das, A. and Diamond, P. H., “Theory of two-dimensional mean field electron magnetohydrodynamics,” *Phys. Plasmas* **7**, 170–177 (2000).
- Dastgeer, S., Das, A., Kaw, P. and Diamond, P. H., “Whistlerization and anisotropy in two-dimensional electron magnetohydrodynamics,” *Phys. Plasmas* **7**, 571–579 (2000).
- Diamond, P. H. and Gruzinov, A. V., “Suppression of cross-field turbulent transport of passive scalar concentration in two dimensional magnetohydrodynamics,” *Phys. Rev. Lett.* **78**, 3306–3309 (1997).
- Diamond, P. H., Hazeltine, R. D., An, Z. G., Carreras, B. A. and Hicks, H. R., “Theory of anomalous tearing mode growth and the major tokamak disruption,” *Phys. Fluids* **27**, 1449–1462 (1984).
- Diamond, P. H. et al., “Dynamics of zonal flows and self-regulating drift-wave turbulence,” In: *Plasma Phys. and Controlled Nuclear Fusion Research* (IAEA, Vienna, 1998) IAEA-CN-69/TH3/1.
- Dubrulle, B., Laval, J.-P. and Sullivan, P., “A new subgrid model for the atmospheric surface layer. II. Prediction of mean flow and velocity and temperature variances,” *J. Atmos. Sci.* **59**, 873–887 (2002).
- Field, G. F., Blackman, E. G. and Chou, H. S., “Nonlinear alpha-effect in dynamo theory,” *Astrophys. J.* **513**, 638–651 (1999).
- Frisch, U., She, Z. S. and Sulem, P.-L., “Large-scale flow driven by the anisotropic kinetic alpha effect,” *Physica D* **28**, 382–392 (1987).
- Fyfe, D. and Montgomery, D., “High-beta turbulence in two-dimensional magnetohydrodynamics,” *J. Plasmas Phys.* **16**, 181–191 (1976).
- Goldreich, P. and Lynden-Bell, D., “II. Spiral arms as sheared gravitational instabilities,” *Mon. Not. R. Astron. Soc.* **130**, 125–158 (1965).
- Goldreich, P. and Sridhar, S., “Toward a theory of interstellar turbulence. II. Strong Alfvénic turbulence,” *Astrophys. J.* **438**, 763–775 (1995).
- Goldreich, P. and Sridhar, S., “Magnetohydrodynamic turbulence revisited,” *Astrophys. J.* **485**, 680–688 (1997).

- Grappin, R., Pouquet, A. and Léorat, J., “Dependence of MHD turbulence spectra on the velocity field-magnetic field correlation,” *Astron. Astrophys.* **126**, 51–58 (1983).
- Gruzinov, A. V. and Diamond, P. H., “Self-consistent theory of mean field electrodynamics,” *Phys. Rev. Lett.* **72**, 1651–1654 (1994).
- Gruzinov, A. V. and Diamond, P. H., “Self-consistent mean field electrodynamics of turbulent dynamos,” *Phys. Plasmas* **2**, 1941–1946 (1995).
- Gruzinov, A. V. and Diamond, P. H., “Nonlinear mean field electrodynamics of turbulent dynamos,” *Phys. Plasmas* **3**, 1853–1857 (1996).
- Iroshnikov, T. S., “Turbulence of a conducting fluid in a strong magnetic field,” *Sov. Astron.* **7**, 566–571 (1964).
- Kichatinov, L. L. and Rüdiger, G., “ Λ -effect and differential rotation in stellar convection zones,” *Astron. Astrophys.* **276**, 96–102 (1993).
- Kim, E., “Turbulent diffusion of large-scale magnetic fields in the presence of ambipolar drift,” *Astrophys. J.* **477**, 183–195 (1997).
- Kim, E., “Nonlinear dynamo in a simplified statistical model,” *Phys. Lett. A* **259**, 232–239 (1999).
- Kim, E. and Diamond, P. H., “On turbulent reconnection,” *Astrophys. J.* **556**, 1052–1065 (2001).
- Kim, E. and Diamond, P. H., “Effect of mean flow shear on cross-phase and transport, reconsidered,” *Phys. Rev. Lett.* **91**, 075001-1–075001-4 (2003).
- Kim, E. and Dubrulle, B., “Turbulent transport and equilibrium profile in 2D MHD with background shear,” *Phys. Plasmas* **8**, 813–824 (2001).
- Kim, E. and Dubrulle, B., “Are the energy and magnetic potential cascades direct or inverse in 2D MHD turbulence?,” *Physica D* **165**, 213–227 (2002).
- Kim, E., Hahm, T. S. and Diamond, P. H., “Eddy viscosity and laminarization of sheared flow in 3D reduced magnetohydrodynamic turbulence,” *Phys. Plasmas* **88**, 3576–3582 (2001).
- Kleeorin, N. and Rogachevskii, I., “Magnetic helicity tensor for an anisotropic turbulence,” *Phys. Rev. E* **59**, 6724–6729 (1999).
- Kleeorin, N. and Ruzmaikin, A. A., “Dynamics of the average turbulent helicity in a magnetic field,” *Magnetohydrodynamics* **18**, 116–122 (1982).
- Kraichnan, R. H., “Inertial-range spectrum of hydromagnetic turbulence,” *Phys. Fluids* **8**, 1385–1389 (1965).
- Krause, F. and Rädler, K.-H., *Mean Field Electrodynamics and Dynamo Theory*, Pergamon Press (1980).
- Lazarian, A. and Vishniac, E., “Reconnection in a weakly stochastic field,” *Astrophys. J.* **517**, 700–718 (1999).
- Leprovost, N. and Kim, E., “Self-consistent mean field theory in weakly ionized media,” *Astrophys. J. Lett.* **598**, L99–102 (2003).
- McComb, W. D., *The Physics of Fluid Turbulence*, Oxford University Press, Oxford (1990).
- Moffatt, H. K., *Magnetic Field Generation in Electrically Conducting Fluids*, Cambridge University Press, Cambridge (1978).

- Or, C. R. and Busse, F. H., "Convection in a rotating cylindrical annulus. Part 2. Transitions to asymmetric and vacillating flow," *J. Fluid Mech.* **174**, 313–326 (1987).
- Orszag, S. A., "Analytical theories of turbulence," *J. Fluid Mech.* **41**, 363–386 (1970).
- Parker, E. N., "A solar dynamo surface wave at the interface between convection and uniform rotation," *Astrophys. J.* **408**, 707–719 (1993).
- Pouquet, A., "On two-dimensional magnetohydrodynamic turbulence," *J. Fluid Mech.* **88**, 1–16 (1978).
- Pouquet, A., Frisch, U. and Léorat, J., "Strong MHD helical turbulence and the nonlinear dynamo effect," *J. Fluid Mech.* **77**, 321–354 (1976).
- Proctor, M. R. E., "Dynamo processes: the interaction of turbulence and magnetic fields," in: *Stellar Astrophysical Fluid Dynamics* (Eds. M. J. Thompson and J. Christensen-Dalsgaard), Cambridge University Press, (2003).
- Riyopoulos, S., Bondeson, A. and Montgomery, D., "Relaxation toward states of minimum energy in a compact torus," *Phys. Fluids* **25**, 107–115 (1982).
- Rüdiger, G., *Differential Rotation and Stellar Convection: Sun and Solar-type Stars*, Gordon and Breach, New York (1989).
- Strauss, H. R., "Nonlinear, three-dimensional magnetohydrodynamics of non-circular tokamaks," *Phys. Fluids* **19**, 134–140 (1976).
- Strauss, H. R., "Hyper-resistivity produced by tearing mode turbulence," *Phys. Fluids* **29**, 3668–3671 (1986).
- Taylor, J. B., "Relaxation and magnetic reconnection in plasmas," *Rev. Mod. Phys.* **58**, 741–763 (1986).
- Thelen, J.-C. and Cattaneo, F., "Preliminary measurement of the turbulent magnetic diffusivity in three dimensions," in: *Dynamo and Dynamics, a Mathematical Challenge* (Eds. P. Chossat, D. Armbruster and I. Oprea), pp. 101–108, Kluwer (2001).
- Tobias, S. M., "The Solar Tachocline: formation, stability and its role in the solar dynamo," this volume, Chapter 7.
- Tsinober, A. B., in: *Viscous Drag Reduction in Boundary Layers, Progress in Astronautics and Aeronautics* (Eds. D. M. Bushnell and J. N. Hefner), Vol. 123, p. 327, American Institute of Aeronautics and Astronautics, Washington D. C. (1989).
- Wilkinson, L. and Hughes, D. W., "The role of boundary conditions in turbulent magnetic diffusion," in preparation (2005).
- Zeldovich, Ya. B., "The magnetic field in the two dimensional motion of a conducting turbulent liquid," *Sov. Phys. JETP* **4**, 460–462 (1957).

7 The solar tachocline: Formation, stability and its role in the solar dynamo

S. M. Tobias

*Department of Applied Mathematics, University of Leeds,
Woodhouse Lane, Leeds LS2 9JT, UK*

In this paper I shall review recent progress in our theoretical understanding of the solar tachocline. This layer of strong radial and latitudinal differential rotation at the base of the solar convection zone is believed to play a vital role in the generation of the solar magnetic field. I shall review the observations of the location, size and temporal variability of the tachocline, before discussing competing theories for its formation. I shall then describe possible instability mechanisms of the tachocline and its role in generating the Sun's magnetic field by dynamo action.

1. Introduction

The origin of magnetic activity in the Sun is a fundamental problem of solar magnetohydrodynamics. The magnetic field, as well as being responsible for the solar cycle and the formation of active regions, underlies all solar magnetic phenomena such as solar flares, coronal mass ejections and the solar wind, and is responsible for heating the solar corona to such high temperatures (Tobias, 2002b). Solar magnetic activity is believed to be the result of a hydromagnetic dynamo where inductive motions within the ionized solar plasma are able to generate a magnetic field, overcoming the action of Ohmic dissipation (see, e.g., Weiss, 1994).

The nature of the inductive motions of the plasma within the Sun are therefore of paramount importance in determining the generation of the large-scale magnetic field. It has been known since the time of Carrington (1858) that the Sun rotates differentially at the solar surface with the equator rotating significantly faster than the poles. However, it is only with the advent of helioseismology that new important

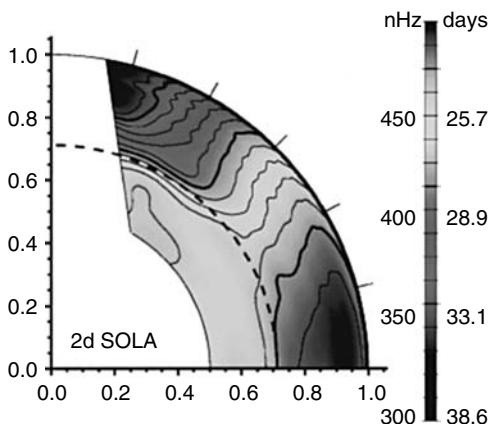


Figure 1. Inversion for the interior solar rotation rate $\Omega/2\pi$ with radius and latitude using the two-dimensional SOLA inversion technique (from Schou *et al.*, 1998). (See color insert.) The dashed line represents the base of the convection zone.

information about the interior dynamics of the Sun has been obtained. Helioseismology involves the analysis of the properties of acoustic waves (pressure fluctuations) measured at the solar surface by either ground-based telescopes or satellites; the frequency of these waves is modified not only by the sound speed of the medium in which they propagate, but also by any large-scale fluid flows. The techniques involved in helioseismological inversions are many and varied and certainly beyond the scope of this review (the reader is directed to Thompson, 2003; Toomre, 2003, for further information). It is sufficient for our purposes to note that helioseismology provides maps not only of the sound speed in the solar interior but also a crucial piece of information for dynamo theory—the differential rotation of the interior of the Sun, an example of which, taken from Schou *et al.* (1998) is shown in Fig. 1 (see color insert following page ____).

This figure shows a color-coded map of the (azimuthally averaged) solar rotation profile as a function of latitude and radius, using the two-dimensional-SOLA inversion technique. The differential rotation pattern appears to be closely related to the underlying thermal stability properties of the plasma as expected, with different rotational profiles in the convection zone and radiative interior. The differential rotation observed at the solar surface appears to be largely imprinted throughout the convection zone. Contours of constant angular frequency are therefore nearly radial at mid and high latitudes, but at lower latitudes they are more aligned with the rotation axis. A thin shear layer close to

the solar surface can also be clearly seen in the inversions. These features in the convection zone must be related to the turbulent transport of angular momentum by the convective motions, but the precise mechanism leading to this rotational profile in the convection zone remains unclear (Brun and Toomre, 2002). Observations of the rotation in the radiative interior of the Sun appear to be consistent with solid body rotation there, with the interior rotating at a rate intermediate between that of the pole and equator (Charbonneau *et al.*, 1998). These two profiles are matched at (or just below) the base of the convection zone by a layer of strong radial shear at low and high latitudes (at mid latitudes the differential rotation depends only weakly on radius). This layer of strong radial shear in the stably stratified radiative interior has been termed the *tachocline*. It is well established that such strong layers of differential rotation are key players in generating magnetic fields via dynamo action — they facilitate the generation of strong toroidal fields from poloidal — and for this reason there has been much interest in the tachocline from dynamo theorists, as discussed in Section 5.

The discovery of the strong layer of differential rotation at the base of the solar convection zone, now known as the tachocline, has proved a pivotal point in our understanding of the solar interior. Since the inversions by Libbrecht (1988) and Brown *et al.* (1989) led to the discovery of this intriguing shear layer, many subsequent studies have attempted to pinpoint both the spatial structure and temporal evolution of the tachocline. The next section summarizes the current bounds on the physical properties of the tachocline from helioseismology and solar modeling.

2. Properties of the solar tachocline

[Figure 2](#), taken from the SOHO/MDI web-site (see also Kosovichev *et al.*, 1997), shows the rotation rate as a function of radius at three different solar latitudes 0° (the equator), 30° and 60° . The radial shear is clearly strong at the equator (where the rotation rate decreases rapidly with depth from the solar surface) and at high latitudes (where the rotation rate increases with depth). At mid latitudes the rotation rate in the convection zone is comparable to that in the solar interior and so no matching shear layer is necessary; the rotation profile for 30° shows no significant variation with radius except for the thin shear layer just below the solar surface. The change in angular frequency at the equator is roughly 30 nHz corresponding to a change in velocity there of approximately $\Delta U \approx 50\text{ ms}^{-1}$ across the tachocline. At high latitudes the

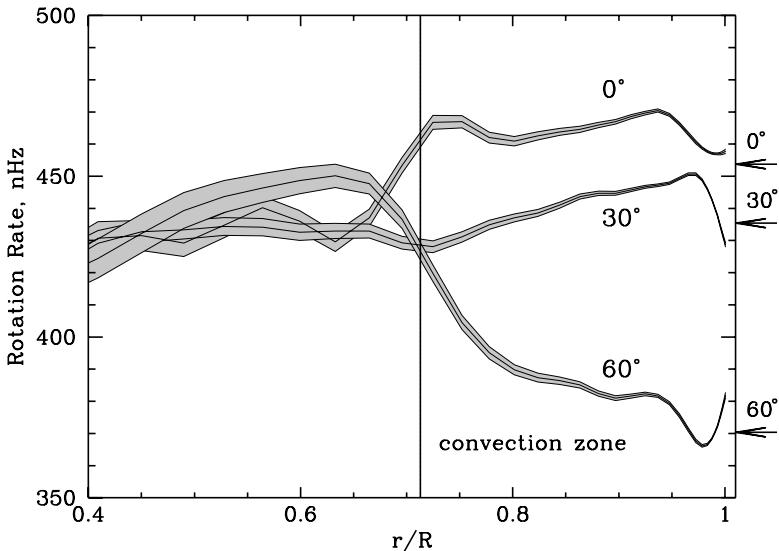


Figure 2. Inferred rotation rate as a function of depth and latitude. Here the tachocline is clearly visible in the inversions for the equator and high latitudes. Courtesy of SOHO/MDI consortium. SOHO is a project of international cooperation between ESA and NASA.

change in angular frequency is somewhat larger, although the reliability of the helioseismic inversions becomes less certain at high latitudes.

Charbonneau *et al.* (1999a) utilized two different inversion methods (Regularized Least-Squares (RLS) and Subtractive Optimally Localized Averages (SOLA)) to analyse the helioseismological data (LOWL 2 year frequency-splitting data) in order to determine the properties of the tachocline. Their analysis indicated that the position of the tachocline is given by $R_c/R_\odot = 0.69$ at the equator, where R_c is the radial position of the center of the tachocline and R_\odot is the solar radius. The tachocline was also determined to be prolate, with a difference in central radius $R_c/R_\odot = 0.024$ between latitude 60° and the equator. The important measurement of the thickness of the tachocline cannot be determined as accurately from the rotation inversions owing to the thickness of the inversion kernels at that solar depth. Charbonneau *et al.* (1999a) estimated an equatorial thickness of $R_{\text{tach}}/R_\odot = 0.04$, although it is likely that this is an *upper bound* for the thickness; the data appears to be consistent with an arbitrarily thin tachocline and no lower bound is possible from differential rotation inversions (Gough; private communication). The thickness of the tachocline does not appear to show any statistically significant variation with latitude.

The thickness of the tachocline can also be established by examining sound speed anomalies from a standard solar model (Elliott and Gough, 1999). These anomalies, again determined by helioseismology, arise owing to a helium deficit that decreases the mean molecular weight of the plasma and hence increases its sound speed. This helium deficit may arise owing to mixing by flows (either turbulence or meridional circulations as discussed later) in the tachocline. This method yields an estimate for the mean tachocline thickness of $0.02 R_{\odot}$.

A very interesting recent result indicates the existence of oscillations in the rotational frequency in the tachocline. The oscillations are measured to have a mean period of 1.3 years (Howe *et al.*, 2000). The largest temporal changes in the angular velocity are approximately 6 nHz, or 20% of the 30 nHz difference in rotation rate of the tachocline at the equator. Spatiotemporal analysis would seem to indicate that these oscillations take the form of a standing wave. These results have however been contested (see Basu and Schou, 2000; Basu and Antia, 2001). These authors claim not to find any significant temporal variation in the depth of the convection zone, the position of the tachocline or the extent of overshoot below the convection zone. No systematic variation in any other properties of the tachocline is found either. It is clear that continued observations and inversions are needed to establish the presence or absence of such oscillations, although preliminary further studies (Thompson; private communication) indicate that the oscillations appear to remain as the observations are continued.

It is also important that theoretical models of the tachocline take account of the physical conditions there. The shear layer that constitutes the tachocline lies within a high density ($\rho \approx 0.2 \text{ g cm}^{-3}$), high temperature ($T \approx 2 \times 10^6 \text{ K}$) region with a gas pressure of $p \approx 6 \times 10^{12} \text{ Pa}$ (Cally, 2003). As the site of solar dynamo action, the tachocline is believed to generate a strong large-scale toroidal field. A toroidal field strength of 10^4 – 10^5 G yields a magnetic pressure of $5 \times 10^7 \text{ Pa}$ and so the plasma $\beta \approx 10^5$ in the tachocline. As noted above, it is believed that the tachocline lies mainly in a region of strong stable stratification, although the upper portion of the tachocline may lie in a region of convective overshoot; strong downward sinking plumes from the convection zone above are able to penetrate into the nominally stable stratification. As described below, the strong stratification and magnetic field are believed to play a dominant role in determining the properties of the tachocline.

It is also important to note that the various physical processes that contribute to the formation and dynamics of the tachocline take place on a wide range of temporal and spatial scales. The Sun has been spun

down to its current angular velocity profile through magnetic braking and angular momentum redistribution—processes that have been steadily occurring over the lifetime of the Sun (10^9 years). Meridional flows within the tachocline itself are thought to lead mixing and redistribution of angular momentum on a turnover time of approximately 10^6 years. These slow processes should be contrasted with the timescales for magnetic field evolution (the 22-year magnetic cycle may be modulated on timescales of a few hundred years) while convective timescales for the turbulence in the solar convection zone are of the order of 1 month. There is less disparity in the spatial scales pertaining to the physics of the tachocline. Whilst the thickness of the tachocline can be bounded above using helioseismology, as discussed above, other length scales can only be determined by theoretical means. Standard solar models demonstrate that the pressure scale height at the base of the solar convection zone is given by $H_p \approx 50$ Mm. Stix (2002) and numerical models of overshooting convection estimate that the extent of convective overshoot could be up to half the radius of the tachocline or approximately 7000 km. It is also estimated that the region in which the solar dynamo magnetic flux is stored has a width of 10000–20000 km (Galloway and Weiss, 1981), although it is possible that some of this flux is stored in the lower reaches of the solar convection zone. The interaction of these processes on different spatial and temporal scales must invariably play an important role in the formation and dynamics of the tachocline as discussed at length in the next three sections.

3. Why is the solar tachocline so thin?

The fundamental question of tachocline dynamics is simply put—why is there a tachocline at all? As the Sun has spun down owing to the action of magnetic braking (Mestel, 1999), angular momentum has been lost via the solar wind. Magnetic and hydrodynamic torques must have acted to spin down both the outer convective envelope and the radiative interior to their current rotation rates and this suggests that there exists some coupling between them. Given this coupling between the two regions, why are their rotation profiles so different? What mechanisms prevent the differential rotation profile of the convection zone propagating into the radiative interior and destroying the thin layer of shear?

This interesting problem of the origin of the tachocline was first posed by Spiegel and Zahn (1992a). (Incidentally the tachocline was first termed the *tachycline* by Spiegel and Zahn (1992b)). In their seminal paper they considered a purely hydrodynamic axisymmetric spherical shell in hydrostatic and geostrophic balance. This thin shell

was subjected to a viscous stress on the outer radial boundary to mimic that imposed on the stable layer from the latitudinal differential rotation in the convection zone. This stress, in conjunction with radiative effects, was shown to drive meridional flows within the shell. These flows transported angular momentum from the upper boundary and were responsible for spreading the shear layer into the radiative interior, therefore significantly thickening, and eventually destroying, the tachocline. Spiegel and Zahn's solution to stopping this spread of the tachocline was argued as follows. The radiative interior is very stably stratified causing three-dimensional (in particular radial) motions to be energetically unfavorable. (Note that this is a *hydrodynamic* statement — magnetohydrodynamic effects can and do act so as to allow three-dimensional motions in parts of the tachocline as discussed later.) Any turbulence generated within the tachocline will therefore be largely two-dimensional and highly anisotropic. Spiegel and Zahn (1992a) argued that this turbulence acts on the global motions as an anisotropic eddy viscosity, preferentially mixing angular momentum on two-dimensional surfaces toward a state of uniform rotation. They went on to demonstrate that such an anisotropic viscosity included in a global model could prevent the radiative spreading of the tachocline into the interior. Numerical solutions of the large-scale dynamics with imposed isotropic and anisotropic eddy viscosities (Elliott, 1997) support these arguments. However, a number of important questions are still left open by such a model. What is the source of instability that leads to the formation of the anisotropic turbulence and hence the anisotropic viscosity? Even if an instability mechanism can be identified, is the nature of the turbulence guaranteed to act like an anisotropic viscosity driving the system toward uniform rotation? What is the role of magnetic fields in modifying the picture — both the assumption of two-dimensionality and the nature of the turbulence in transporting angular momentum?

Another purely hydrodynamic mechanism for suppressing the spread of the tachocline relies on the action of gravity waves. Here, gravity waves generated at the interface between the radiative interior and convection zone by penetrative convection play an important role (Zahn *et al.*, 1997; Talon *et al.*, 1997). In this picture prograde and retrograde gravity waves are dissipated differentially by the shear flow and act in such a way as to suppress the shear in the radiative interior. However, as argued by many authors (see, e.g., Gough, 1977), torques due to gravity waves have little tendency to drive the large-scale dynamics toward a state of uniform rotation and so it is unclear whether such a mechanism can be supported (see also Kim and MacGregor, 2001, for a discussion of the role of gravity waves in the solar tachocline).

An alternative approach to the purely hydrodynamic mechanisms suggested above is to invoke the presence of a magnetic field in the interior of the Sun. It has long been understood (Cowling, 1957) that such a magnetic field in the radiative interior of the Sun would cause it to rotate as a solid body. Mestel and Weiss (1987) estimate that a weak fossil poloidal magnetic field of 10^{-3} – 10^{-2} G is sufficient to suppress any differential rotation there (see also Rüdiger and Kitchatinov, 1997). Any sufficiently large relic magnetic field in the interior could therefore act to prevent the radiative spreading of the tachocline into the interior. This is not the end of the story, however, and some important questions remain. What is the nature of this magnetic field and how does the magnetic field couple the interior to the convection zone? If the coupling is magnetic, what mechanism prevents angular momentum from the convection zone being transported *along magnetic field lines* into the radiative interior? MacGregor and Charbonneau (1999) showed that a large-scale fossil magnetic field could induce solid body rotation in the interior *and* lead to a thin tachocline provided none of the field lines were embedded in the convection zone, i.e., if the magnetic field were completely confined to the radiative zone. If, however, the poloidal field were to possess a nonzero component normal to the core-envelope interface then the transmission of significant differential rotation into the radiative interior would be likely. How can the two statements be reconciled? Clearly some form of magnetic coupling between the convection zone and radiative interior is desirable to help spin down the star. This coupling must act, however, in such a way so as not to transmit the differential rotation into the interior along the field lines.

This interesting problem was addressed as part of an elegant discussion by Gough and McIntyre (1998). The first part of the paper deals with one of the hydrodynamic questions posed above—that of the nature of two-dimensional, strongly stratified turbulence and its role in the transportation of angular momentum. They argue (see McIntyre, 2003, for a more detailed exposition) that the tachocline is similar in nature to the Earth's atmosphere (in particular to the stratosphere), being a stably stratified rotating layer. Studies of atmospheric models indicate that two-dimensional turbulence does *not* act as an anisotropic eddy viscosity driving large-scale flows toward uniform rotation. Instead the turbulence acts so as to mix *potential vorticity* (not angular momentum) until a state of uniform potential vorticity is reached. This state of uniform potential vorticity in a thin spherical layer is very different from one of uniform rotation. In this way the turbulence acts to drive the system *away* from solid body rotation and is often described as an anti-friction. It should be noted, however, that the Prandtl number

(v/κ) is extremely small in the tachocline and $O(1)$ in the atmosphere, and this may have a significant effect on the dynamics and transport.

Further numerical studies of the transport effects of tachocline turbulence have yielded interesting conclusions. A recent study initiated by Miesch has investigated the nonlinear couplings in both freely evolving turbulence (Miesch, 2001) and forced hydrodynamic turbulence (Miesch, 2003) in a thin layer. In the second of these papers, the turbulence is driven from above by convective overshoot and the role of an imposed differential rotation in modifying the nature of the turbulence is investigated. In this model, which is fully three-dimensional and stably stratified, the nonlinear Reynolds stresses are found to transport momentum poleward and outward. This implies diffusive (down-gradient) latitudinal transport and antidiffusive (counter-gradient) vertical transport. Overall, the latitudinal transport dominates and acts to dissipate the imposed shear.

A cautionary note for these purely hydrodynamic models of turbulent transport in the tachocline comes from accretion disc studies. There it has been shown that the presence of even a weak magnetic field can have serious implications for the nature of the turbulent transport. In models of differentially rotating hydromagnetic discs, Maxwell stresses may play a key role in angular momentum transport—significantly larger than the associated Reynolds stresses. The Maxwell stresses can even act so as to reverse the direction of angular momentum transport from purely hydrodynamic models (see the review by Balbus and Hawley, 1998). Another important consequence of including a magnetic field, as mentioned earlier, may be to increase the three-dimensionality of the turbulence. Here magnetic buoyancy may be important in allowing radial motions to be energetically favorable and have a significant impact on the nature of the turbulence.

The second line of argument in the Gough and McIntyre model attempts to address the dynamics within the tachocline in a magnetically confined model. Here a relic magnetic field in the radiative interior ensures solid body rotation there while the differential rotation in the convection zone (maintained presumably by an anisotropic transfer of angular momentum by the Reynolds stresses) there supplies an upper boundary condition for the tachocline. Gough and McIntyre (1998) argue that angular momentum transport within the tachocline is driven by meridional circulations there. These meridional flows are driven by gyroscopic pumping from the convection zone above and take the form of two cellular flows with downwellings near the pole and equator and an upwelling where the shear is small at mid latitudes, as shown in Fig. 3 (see color insert following page 234). The upwelling could in principle

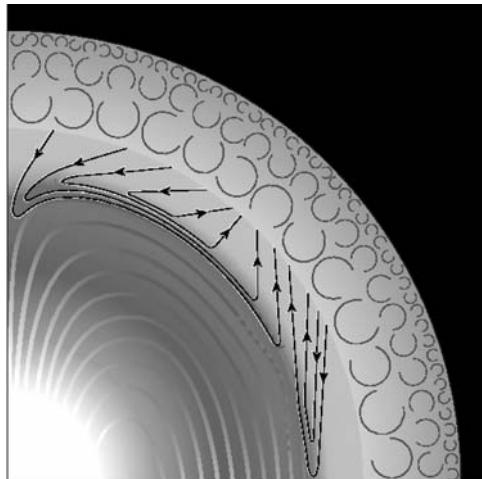


Figure 3. Schematic representation of a meridional quadrant of the Sun (after Gough and McIntyre, 1998) (See color insert.) The arrows represent the tachocline circulation, which follows surfaces S of constant specific angular momentum in the (green) body of the tachocline (whose thickness has been exaggerated by a factor 5), and is deflected by the magnetic field in the (blue) diffusive boundary layer (whose thickness has been exaggerated by a factor 50).

advect poloidal magnetic flux from the radiative interior and produce a coupling between the radiative interior and the convection zone, while magnetic flux at low and high latitudes is held down by the downwellings. A delicate balance then ensues determining the tachocline dynamics. A key feature of the model is that the magnetic field that emerges at mid latitudes *does* provide coupling between the radiative core and the convection zone *without* propagating differential rotation into the interior (recall that the rotation rate at mid latitudes is similar to that in the core). A more detailed understanding of the balance can be gained from numerical solution of the equations. Garraud (2002) undertook a numerical study of tachocline dynamics in a magnetohydrodynamic incompressible model of a rotating spherical shell. An initially poloidal magnetic field is imposed in the shell and the inner core is assumed to be a rigid rotator. In this simplified model meridional flows are driven by Ekman pumping due to stresses at the outer boundary of the spherical shell. Here the crucial non-dimensional parameter is the Elsasser number $\Lambda = B_0^2 / \rho_0 \Omega_c^2 r_c^2$, which measures the strength of the imposed magnetic field relative to rotation. For small Λ , the magnetic field is passive and the shear imposed at the outer boundary propagates into the radiative interior parallel

to the rotation axis driven by meridional flows. Here the meridional flows penetrate deep into the radiative interior and the steady states are very close to those for pure hydrodynamics. At the other extreme with strong field (large Λ) the field is not easily advected by the meridional flows and remains close to its initial configuration. This field strongly couples the radiative interior with the convection zone and the latitudinal differential rotation propagates inward along the magnetic field. Constant angular velocity contours in this case therefore closely follow the magnetic field lines. The interesting case is the delicate $\Lambda \sim O(1)$ regime where the first signs of “tachocline-like” dynamics are observed. Near the equator the meridional flow is able to hold down magnetic field in the interior; near-uniform rotation is enforced there with a shear layer being maintained. More structure is found in the interior, however, at higher latitudes where field lines *do* protrude into the convection zone and rotation propagates inward along these field lines. So tachocline-like behavior can only be found at low latitudes and then only if the field strength is in the intermediate regime.

From the arguments given above, it is clear that magnetic fields should play an active role in the formation of the tachocline. Not only will a large-scale poloidal field ensure that the radiative interior rotates as a solid body, but magnetic stresses may be important in driving flows in the tachocline. Furthermore, correlations in any tachocline turbulence engendered by the presence of a magnetic field *could have* significant consequences for the properties of angular momentum transport. Finally, it is important to note that the magnetic field, via the effects of magnetic buoyancy, *will* have the effect of allowing some three-dimensionality in any turbulence. This may have a significant effect on the nature of the transport and have important implications for models of the formation of the tachocline.

4. Instabilities of the solar tachocline

The last section described two competing models for the formation of the tachocline. In each a physically plausible scenario has been developed in order to explain the presence, and confinement, of the shear layer in the stable region at the base of the solar convection zone. It is important to understand whether the current tachocline configuration is stable to perturbations and, if the configuration is found to be unstable, how the instability can saturate and lead to modification of the basic state. In particular the Spiegel and Zahn model for the formation of the tachocline relies on the presence of an anisotropic turbulent viscosity to prevent the spreading of the shear layer into the radiative interior.

It is argued that the nonlinear development of an instability into quasi two-dimensional turbulence may provide the basis for this anisotropic viscosity. It is therefore vital to understand the possible dynamics of tachocline instabilities. This section will review the progress that has been made in this direction.

The tachocline is a stratified region of strong differential rotation and strong magnetic field and a number of instability mechanisms may be envisaged. In particular, instabilities of the radial and latitudinal differential rotation, instabilities of the magnetic field (including magnetic buoyancy) and joint instabilities of the magnetic field and flow are available. These are described below.

4.1. Differential rotation and joint instabilities

As discussed above, the helioseismic results indicate that the tachocline exhibits a very strong radial and weaker latitudinal differential rotation. It is well established from laboratory experiments and atmospheric studies that such shear flows may be susceptible to instability. It is natural, therefore, to determine whether instability in the tachocline could arise as a result of a purely hydrodynamic differential rotation instability.

At this point it is worth inserting a note of caution. The instability of the solar tachocline is a classic example of a “chicken and egg” scenario. Many of the stability analyses described below, particularly those dealing with differential rotation instabilities, investigate the available instability mechanisms for a differential rotation state derived from the helioseismic data. However, the current differential rotation profile is presumably a nonlinear state (which could be turbulent) that has resulted from the saturation of any instabilities that have occurred. Nonlinear systems typically saturate so as to suppress the initial instability mechanism and the system then exists in a state of statistical marginal stability. Ideally one would like to investigate the instability of a differential rotation profile that undergoes instability to a nonlinear statistically saturated state similar to the current differential rotation profile. This, in practice, is difficult to achieve, but it is perhaps not surprising then that some of the stability analyses described below indicate that the stability of the tachocline is close to being marginal.

Hydrodynamic instabilities

A number of possible purely hydrodynamic instabilities associated with differential rotation and leading to angular momentum redistribution

are available in stellar interiors. Among these are shear instabilities, baroclinic instabilities and Goldreich-Schubert-Fricke instabilities as discussed in Knobloch and Spruit (1982) and the references therein. The most striking facet of the solar tachocline is the extremely sharp radial gradient in differential rotation. It is well known that vortex sheets associated with strong shear flows can undergo “wrap-up” (Kelvin-Helmholtz type) instabilities leading to the onset of shear turbulence. However, in a layer of strong stratification any shear instability must compete with the stabilizing effect of the stratification owing to the presence of strong radial gradients in density. These make it energetically unfavorable for any radial motions required by the shear flow instability to occur. These competing effects can be quantified by the Richardson Criterion (see, e.g., Drazin and Reid, 1981) for stratified shear flows, which states that instability is not possible in a *hydrodynamic* stratified shear flow if

$$Ri = \frac{N^2}{(dU/dz)^2} > \frac{1}{4}, \quad (4.1)$$

where N is the Brunt-Väisälä frequency given by $N^2 = (g/\vartheta)(\partial\vartheta/\partial z)$ (g being the gravitational acceleration, ϑ the potential temperature and z the local radial coordinate). The crucial non-dimensional parameter, the Richardson Number (Ri), is therefore the ratio of the stabilizing effects of buoyancy to the strength of the radial shear flow. In the tachocline the Richardson Number has the slightly modified definition $Ri = N^2(r \sin \theta \partial \Omega / \partial r)^{-2}$, and is a function of depth ranging from 10^4 near the base of the tachocline to 10^2 near the top. It is therefore commonly assumed that the radial shear profile is very stable. It is, though, important to stress again that this criterion for the stability of shear flows is a purely hydrodynamic result. It is often argued that, because the tachocline is in a region of very strong stratification, the effects of buoyancy will successfully suppress vertical motions and that motion in the tachocline will largely be two-dimensional. However, this ignores possible effects of the magnetic field, which, as we shall see, can have a non-trivial effect both in modifying shear instabilities *and* altering the stabilizing effect of the stratification via magnetic buoyancy. This is an important caveat that should be recalled when reading the discussion in the following sections — one to which we shall return later.

Within the hydrodynamic approximation, the possibility of a radial shear instability was investigated by Schatzman *et al.* (2000). They consider the radial differential rotation profile inferred from helioseismology and the stratification derived from a standard solar model to

calculate a local Richardson number for the flow at a number of latitudes. Here they argue correctly that the local buoyancy in the tachocline is reduced owing to the loss of heat from turbulent eddies. This process is determined by the Péclet number of the eddies (ul/K) where u is a typical eddy velocity of length-scale l and K is the thermal diffusivity; for $Pe \ll 1$ a modified Richardson criterion is applicable and shear turbulence may still arise in a narrow region at the equator and in even thinner layers at high latitude. Note that this argument is similar in spirit to that for the destabilizing effect of magnetic buoyancy which may also lead to a modified Richardson criterion. Schatzman *et al.* (2000) conclude, however, that the turbulence generated by this vertical hydrodynamic shear is unlikely to play a significant role in the transport of matter and angular momentum, and that other mechanisms must be invoked to achieve this.

Because of the apparent difficulty of releasing energy via radial motions due to the strong solar stratification, much attention has focused on instabilities that do not require work to be done against gravity—i.e., purely two-dimensional instabilities of the *latitudinal* differential rotation profile (where two-dimensional means instabilities that only depend on solar latitude θ and longitude ϕ , and where radial motions are suppressed).

The stability of shear flows to motions restricted to isobaric surfaces (so that no work is done against gravity) has been extensively studied. For two-dimensional motions on a spherical surface an analogous result to Rayleigh's famous inflexion point criterion guarantees stability of a flow with angular velocity $\omega(\mu)$ if

$$\frac{d^2}{d\mu^2} [\omega(\mu)(1 - \mu^2)] \neq 0$$

anywhere in the domain (Fjørtoft, 1950), where $\mu = \cos \theta$. Watson (1981) investigated the two-dimensional instabilities of a parameterized solar latitudinal differential rotation profile. He examined the inviscid, linear stability of an imposed rotation law where $\omega = s_0 - s_2 \cos^2 \theta$, and demonstrated that, for an equator-to-pole angular velocity difference of less than about 29%, such a differential rotation profile is stable. The current estimate is that this places the latitudinal profile well within this regime of stability. It may be that the differential rotation profile in the tachocline is better modeled by a quartic dependence of rotational frequency on latitude, i.e., $\omega = s_0 - s_2 \cos^2 \theta + s_4 \cos^4 \theta$. The addition of this quartic term may have the effect of reducing the angular velocity difference between pole and equator required for instability. Hence for large enough s_4 the differential rotation profile may be hydrodynamically

unstable (Dziembowski and Kosovichev, 1987; Charbonneau *et al.*, 1999b). However, as noted by Charbonneau *et al.* (1999b) the current helioseismic inversions show no significant deviations from a rotation profile with $s_4 = 0$ and so it is unclear whether a hydrodynamic latitudinal instability can play a large part in driving the tachocline dynamics.

As noted earlier, owing to the strong stratification in the tachocline, two-dimensional modeling of tachocline instabilities has been extensively employed. A natural progression from two-dimensional purely hydrodynamic models (again in analogy with progress in geophysics) is to include limited three-dimensional effects via the shallow water approximation. The shallow water equations (Dikpati and Gilman, 2001a, and see the references therein) support three-dimensionality by allowing the radial deformation of the boundary of a thin shell of fluid. Buoyancy effects are parameterized by the inclusion of an “effective gravity” term—high effective gravity terms correspond to very stable stratification, while low effective gravity corresponds to a weakly stratified layer. For strong stratification, the results of a stability analysis unsurprisingly agree with the fully two-dimensional models—instability is only possible for an extremely large equator-to-pole angular velocity difference. As the effective gravity is lowered, a new class of unstable disturbances is found, for smaller values of the equator-to-pole differential rotation. These disturbances are global non-axisymmetric (with azimuthal wavenumbers $m = 1, 2$) modes that differ in character from the purely two-dimensional modes. Thus, as the assumption of quasi two-dimensionality is relaxed the instability mechanism appears to change in nature. These unstable modes are demonstrated to possess a nonzero kinetic helicity $\langle \mathbf{u} \cdot \boldsymbol{\omega} \rangle$ (as might be expected from an instability in a rotating system). Dikpati and Gilman argue that this flow may have significant consequences for the solar dynamo by producing a localized α -effect (see Section 5 for a brief discussion of the importance of the α -effect in the solar dynamo and the role of the tachocline in generating the large-scale solar field). However, this line of argument is uncertain; in mean-field theory the α -effect arises from interactions between small-scale flows and small-scale magnetic fields. It can be shown under certain approximations (first-order smoothing and short-correlation time approximations) that this effect can be related to interactions between small-scale velocity and vorticity (via the net helicity). However, there is no mechanism for relating the α -effect to the kinetic helicity of a *global* hydrodynamic flow (although such flows may play a role in the generation of large-scale fields). Indeed many large-scale flows on the Sun have a nonzero net helicity (for example, the combination of the differential rotation and postulated meridional flows have a

nonzero helicity) but are not thought to be responsible for providing an α -effect. Further clarification of this mechanism is therefore necessary.

The important question of the saturation of any hydrodynamic instability has received less attention, but has been investigated within the two-dimensional framework by Garaud (2001). She found that for a latitudinal differential rotation profile relevant to the tachocline, any hydrodynamic instability saturated at a very low level in the nonlinear regime. Hence, owing to the fact that the angular momentum transport by the associated Reynolds stresses (which act so as to transport angular momentum from equator to pole) is small, the latitudinal differential rotation profile in the nonlinear saturated state is very similar to the imposed shear profile.

MHD instabilities

As noted earlier, the tachocline is believed to possess a strong dynamo-generated toroidal field (in addition to a weaker poloidal field), which may play a part in its formation and stability. It is therefore important to determine the role of magnetic fields in modifying any differential rotation instabilities. If one maintains the analogy with geophysical flows and ignores any role of magnetic field in modifying the buoyancy properties of the tachocline, then it is possible to perform stability calculations analogous to the hydrodynamic calculations described above. The simplest reference state configuration is a latitudinal differential rotation profile together with a toroidal magnetic field that depends strictly on latitude. Gilman and Fox (1997) consider the stability to two-dimensional disturbances of a latitudinal differential rotation of the form $\omega = s_0 - s_2 \cos^2 \theta$ in the presence of a toroidal magnetic field with latitudinal profile $b_\phi = b_0 \cos \theta \sin \theta$. This toroidal field is the simplest profile that is antisymmetric about the equator and zero at the poles. They show that for a wide range of magnetic and kinetic energies of the reference state, from well below equipartition to well above it, instability sets in to a non-axisymmetric ($m = 1$) mode which allows field-lines and flows in the perturbation to “tip” across the poles. The instability (often termed a tipping instability) is very robust, occurring for a wide variety of parameter choices. Interestingly, the linear problem indicates that Maxwell stresses in the nonlinear regime would act so as to transport angular momentum toward the pole while Reynolds stresses would be of the opposite sign.

This joint instability clearly has much in common with differential rotation instabilities of azimuthal flows in accretion discs (Balbus

and Hawley, 1998). Differentially rotating discs with Keplerian rotation laws are known to be hydrodynamically stable to small perturbations (although they may be unstable to finite amplitude perturbations). However, the addition of a magnetic field can provide a powerful local instability mechanism, first noted by Velikhov (1959) and Chandrasekhar (1961), that is now termed a magnetorotational instability (MRI). These instabilities may occur if the magnetic field is poloidal (Balbus and Hawley, 1991), where the instability is axisymmetric ($m = 0$) or, more relevantly to the solar case, toroidal, where the instability has preferred wavenumber $m = 1$ (Ogilvie and Pringle, 1996). The analogy continues in the nonlinear regime where Maxwell stresses within the disc transport angular momentum radially outward and saturate the instability. These magnetic stresses are more significant than the Reynolds stresses that are of the opposite sign. Clearly this analogy between these local instabilities and the global instability of the tachocline requires further investigation.

A series of papers investigating the properties of two-dimensional MHD latitudinal differential rotation instabilities followed. In each of these the mechanism for instability (and method of solution of the equations) remains the same as in Gilman and Fox (1997) and only the choice of magnetic field profile differs. These are summarized below. Gilman and Fox (1999) extended the earlier results of Gilman and Fox (1997) to include the case where the unperturbed toroidal field has a node between the pole and the equator mimicking the solar dynamo field that is believed to have a zero toroidal component between the bands that migrate toward the equator. (Note that there is an error in Gilman and Fox (1999) that is corrected in Gilman and Fox (2000)). The role of the latitudinal extent of the toroidal field was investigated next. Dikpati and Gilman (1999) imposed a narrow band of toroidal field at mid latitudes (near 45°) and demonstrated that a combination of concentrated toroidal fields and solar-type latitudinal differential rotation is again unstable, not only to longitudinal wavenumber $m = 1$, but also to $m > 1$ modes for sufficiently narrow toroidal-field profiles. Unsurprisingly, shortening the length-scale of the basic state for the toroidal field in latitude can lead to a shortening of the perturbation length-scale in longitude. The analysis was extended even further by Gilman and Dikpati (2000) who demonstrated that the instability persists if the thin band of toroidal field is imposed at a range of different latitudinal positions (representing different phases of the solar cycle). Here modes with longitudinal wavenumber between $m = 1$ and $m = 7$ may be destabilized. To summarize, the joint instability of differential rotation and toroidal field to two-dimensional disturbances appears to remain

robust for a large range of spatial profiles of toroidal fields at a large selection of latitudinal positions.

A simplified form of three-dimensionality in MHD can be included in a similar manner to that for the hydrodynamic instabilities discussed earlier, by allowing the radial deformation of the boundary of a thin spherical shell. This form of shallow water MHD was developed by Gilman (2000). This approach includes the effect of magnetic tension and curvature forces in modifying the flows in the thin layer and the role of inductive processes in modifying the magnetic field, but does not include any contribution of the magnetic field to the buoyancy force and the subsequent enhancement of three-dimensional motions. Joint instabilities of differential rotation and magnetic fields have been studied in this shallow water approximation by Gilman and Dikpati (2002) for broad toroidal field profiles. This calculation is the analogue of Gilman and Fox (1997) and further investigations for thin bands of toroidal fields, at different latitudinal positions, are promised in the future. In shallow water MHD, the instability is found to be not only a function of the reduced gravity parameter (as in the hydrodynamic case) but also the strength of the imposed magnetic field. A large effective gravity leads to the instability taking a similar form to its two-dimensional analogue, while reducing the effective gravity changes the character of the instability until, for sufficiently weak effective gravities, the shallow water approximations are no longer valid. As for the hydrodynamic case, the kinetic helicity of the global modes is calculated and related to an α -effect. It seems strange to attempt to describe the α -effect in MHD using solely the kinetic helicity of the perturbed flows (i.e., using only hydrodynamic variables). For this instability problem the α -effect could be calculated directly (although one has to allow for the misgivings about the separation of spatial and temporal scales for the global mode perturbations described earlier) by calculating the EMF driven by correlations of the perturbation flow \mathbf{u}' and field \mathbf{B}' ($\mathcal{E} = \langle \mathbf{u}' \times \mathbf{B}' \rangle$) and relating it to the mean toroidal field $\langle \mathbf{B} \rangle$ via the formula $\mathcal{E}_i = \alpha_{ij} \langle B_j \rangle$, where here $\langle \cdot \rangle$ denotes a longitudinal average.

A promising approach to the inclusion of three-dimensionality for these instabilities has been formulated by Cally (2003). He extends the analysis of the instability of a latitudinal differential rotation profile to include the possibility of three-dimensional disturbances. He utilizes a Boussinesq thin layer, where the equilibrium field and flow are independent of depth (i.e., ignoring radial shear and magnetic field gradients that may lead to magnetic buoyancy instabilities—see later) and locates a new class of instabilities for Alfvén angular velocities larger than rotational angular velocities. These instabilities, which are

concentrated in the polar regions, take the form of an $m = 1$ tipping mode, where the magnetic field twists in a manner reminiscent of Parker's cyclonic events. Here the inclusion of (even a small) vertical variation and vertical velocity again alters the character of the instability from the purely two-dimensional case.

Work has begun on the nonlinear development of the two-dimensional MHD instabilities described above. Cally (2001) has investigated the time evolution of a subset of the profiles considered by Gilman and collaborators. He considers combinations of strong and weak latitudinal shears and magnetic fields. As it is not possible to perform nonlinear calculations in the ideal regime, a small amount of magnetic diffusivity is added. The most interesting calculations focus on the case of a weak (solar-like) shear and a strong toroidal field. This configuration leads to a "clam instability" where the magnetic field in the two hemispheres opens and closes, eventually reconnecting and changing topology before disappearing. A common characteristic of all the simulations is that the nonlinear instabilities produce strong angular momentum mixing that pushes the rotation toward a solid body form. Cally *et al.* (2003) extend these calculations to yet more profiles for the magnetic field, including single bands, double bands and mixed profiles (where a band is added to a broad profile).

Finally, it is interesting to note that the discovery of the layer of shear and strong aligned toroidal field at the base of the solar convection zone has also renewed interest in the classical problem of the role of magnetic fields in modifying shear instabilities. Recently, both Cally (2000) and Hughes and Tobias (2001) have improved the bounds on stability of shear flows in the presence of an aligned magnetic field (in the absence of stratification and buoyancy) found by Adam (1978) in different ways, using a modified version of the famous semi-circle theorem of Howard (1961). This simple analysis can give important information on the role of magnetic fields in modifying and suppressing shear flow instabilities.

4.2. Magnetic buoyancy instabilities

The previous section described the many varied descriptions of differential rotation and joint instabilities in two-dimensional hydrodynamics and magnetohydrodynamics. These models assume (by analogy with atmospheric flows) that radial motions in the tachocline are not energetically favorable because of the strong stratification there. However, there is one form of instability in which the magnetic field drives the

instability by modifying the negative buoyancy of the stably stratified fluid and the instability *proceeds* by releasing energy through radial motions. This instability, termed *magnetic buoyancy*, is known to be important in the solar context as it is responsible for the formation of active regions at the solar surface from the toroidal field generated in the solar tachocline. As the magnetic buoyancy instability is the only one whose consequences we can unambiguously observe at the solar surface, it is important to understand its role in the solar tachocline.

The underlying instability mechanism of magnetic buoyancy is well understood. Here, an atmosphere with a horizontal magnetic field that increases sufficiently with depth is to some extent top-heavy, owing to the contribution of the magnetic pressure, and can become unstable through the release of gravitational potential energy *via vertical motions* (Newcomb, 1961; Parker, 1966). There is considerable literature on both the linear and nonlinear aspects of magnetic buoyancy instabilities, focusing, predominantly, on either static or uniformly rotating equilibria (see the reviews of Hughes and Proctor, 1988; Hughes, 1991). The main results for static, stratified atmospheres with a horizontal magnetic field can be obtained via either a parcel argument, a linear stability analysis or most elegantly through the energy principle (Newcomb, 1961). Here magnetic configurations will be linearly stable to the magnetic buoyancy instability if δW (representing the change of internal energy caused by the linear perturbation) is greater than zero at all depths of the layer. For static atmospheres the self-adjointness of the operator also guarantees instability if $\delta W < 0$ somewhere in the domain of interest. These conditions reduce to an atmosphere being unstable to three-dimensional disturbances if somewhere

$$\frac{d\rho}{dz} < \frac{\rho g}{c^2}, \quad (4.2)$$

where z is in the local radial direction pointing downward and c is the local sound speed, and unstable to two-dimensional instabilities that do not bend field lines (interchange instabilities) if, somewhere,

$$\frac{d\rho}{dz} < \frac{\rho g}{c^2 + a^2}, \quad (4.3)$$

where a is the local Alfvén speed.

At first sight these results do not appear to show any dependence on the field structure. However, because the magnetic field acts so as to modify the mean density profile of the equilibrium state (via the magnetic pressure contribution to the magnetohydrostatic equation) the field enters naturally into the stability criteria, and instability is

guaranteed if

$$a^2 \frac{d}{dz}(\ln B) > -\frac{d}{dz}[\ln(p\rho^{-\gamma})], \quad (4.4)$$

for three-dimensional modes and

$$a^2 \frac{d}{dz} \left[\ln \left(\frac{B}{\rho} \right) \right] > \frac{d}{dz}[\ln(p\rho^{-\gamma})], \quad (4.5)$$

for interchange modes. Hence, if gradients in the magnetic field are sufficiently large compared to the stratification then the atmosphere becomes neutrally buoyant and even unstable to three-dimensional modes. Note that the condition on the instability for pure interchange modes is more stringent, requiring a sufficient increase in B/ρ for instability.

It is now clear that for a full understanding of magnetic buoyancy instability in the solar tachocline it is necessary to incorporate the effects of a velocity shear. It should be obvious that there are two important questions for the tachocline. First, how does the inclusion of a shear flow (either radial or latitudinal) modify the magnetic buoyancy instability and second, what is the role of the magnetic contribution to the buoyancy force in modifying shear flow instabilities? These problems have been considered by Acheson (1978), who performed a local analysis of magnetic buoyancy instability incorporating the effects of differential rotation, and by Adam (1978), who analyzed the stability of a magnetic field with an aligned flow and, through the energy principle, obtained certain sufficient stability criteria for a restricted class of (undulatory) modes. Tobias and Hughes (2003) have addressed these questions through a combination of analytical and numerical techniques.

Similar bounds to those derived by Newcomb (1961) may be derived in the presence of an aligned shear flow $U(z)$ (i.e., the flow and the field are both functions of z and \mathbf{U} is parallel to \mathbf{B}). In a similar manner an energy principle may be derived and the internal energy of the system calculated. The addition of a radial shear flow, however, renders the relevant linear operator non-self-adjoint; thus while it is possible to derive sufficient conditions for *stability*, it is no longer possible to obtain criteria for instability. This is similar in nature to the Richardson criterion discussed earlier which guarantees stability but cannot determine instability. Nonetheless two stability criteria (to all disturbances) for the flow can be derived. Details of the derivation are in Tobias and Hughes (2003) and are omitted here. The first condition ensures stability of the model if, in some frame of reference, both $U^2 < c_T^2$ (where c_T is the cusp

or tube speed given by $c_T^2 = a^2c^2/(a^2 + c^2)$) and

$$(a^2 + c^2)(c_T^2 - U^2)g \frac{d\rho}{dz} \geq \rho g^2(a^2 - U^2), \quad (4.6)$$

everywhere in the flow. For $U = 0$ this condition clearly reduces to (4.2). The second condition is more complicated and involves the wavenumbers of the disturbance along the field lines (k) and perpendicular to the field (l). Here stability is guaranteed if $U^2 < c_T^2$ everywhere and

$$\begin{aligned} & -\frac{d}{dz} \left(\frac{\rho g k^2 U^2 (a^2 - U^2)}{(k^2 + l^2)(a^2 + c^2)(c_T^2 - U^2) + k^2 U^4} \right) \\ & < g \frac{d\rho}{dz} + \rho k^2 (a^2 - U^2) - \frac{\rho g^2 (k^2 + l^2)(a^2 - U^2)}{(k^2 + l^2)(a^2 + c^2)(c_T^2 - U^2) + k^2 U^4}. \end{aligned} \quad (4.7)$$

Neither of these stability criteria (and the second one especially) is immediately tractable and it is hard to visualize the role of the shear flow in modifying the magnetic buoyancy instability, or indeed the magnetic field contribution to buoyancy effects in modifying radial shear instabilities. For a given atmosphere and velocity shear it is very simple to use the derived conditions to determine stability. The results of Adam (1978), Cally (2000) and Hughes and Tobias (2001) indicate that magnetic fields tend to stabilize radial shear instabilities if buoyancy effects are not included. However, the contribution of the magnetic pressure to reducing the stabilizing effects of stratification and favoring radial motions has a tendency to move the system away from the Richardson stability criterion. The balance between these effects is delicate—it may not be sufficient to model the effects of magnetic buoyancy by a simple reduction of the effective gravity or Brunt-Väisälä frequency—and more investigation of their interaction is important.

To determine how the addition of an aligned shear flow modifies the behavior of a buoyantly *unstable* atmosphere, Tobias and Hughes (2003) solved the relevant linear stability problem numerically. The unstable modes in the presence of shear were calculated for an ideal, isothermal atmosphere that is known to be unstable to three-dimensional magnetic buoyancy modes in the absence of shear (i.e., the atmosphere considered by Hughes and Cattaneo, 1987). This was repeated for a number of radial shear profiles and generic results common to all flows were extracted. The shear was found to have a different effect for modes with different wavenumbers perpendicular to the field (l) as shown in Fig. 4. For fully three-dimensional modes with l small, the shear initially destabilizes further the buoyantly unstable modes before stabilizing the mode completely for large enough shear. For modes with large

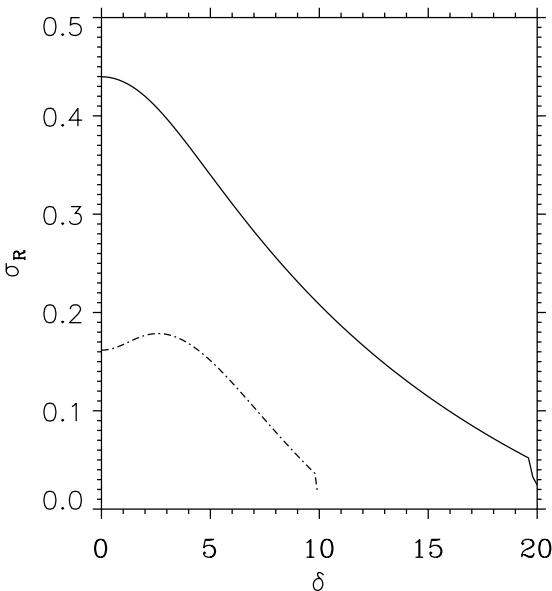


Figure 4. Comparison of role of shear for moderate and large l . Growth rate (σ_R) versus shear strength (δ) for fixed $k = 0.32$, $l = 8$ (solid line) and $k = 0.32$, $l = 3$ (dot-dashed line). Note that for $l = 3$ the flow initially destabilized the instability while for $l = 8$ the flow always has a stabilizing role (after Tobias and Hughes, 2003).

l values, the shear flow appeared always to have a stabilizing effect for the buoyancy mode. An aligned shear flow can have no effect on the growth rate of an interchange mode and so the shear flow can be seen to have two effects; the first is stabilizing on the buoyancy instability in three dimensions and the second is axisymmetrizing—the shear pushes the instability toward the interchange limit.

Furthermore the shear has a significant role in modifying the form of the eigenfunctions of the magnetic buoyancy instability, as shown in Fig. 5. Comparison of the eigenfunctions for the vertical velocity w (corresponding to the radial velocity) and the horizontal velocity u (corresponding to the azimuthal velocity) for the cases of no shear and moderate shear demonstrates that the addition of shear induces more radial structure in the perturbations. By suppressing the instability, the shear acts so as to reduce the relative importance of vertical motions to azimuthal motions. Perhaps the addition of a strong enough radial shear pushes the model into a regime where the vertical motions of the buoyancy instability are reduced and two-dimensional dynamics are again dominant. This interaction between the shear flow

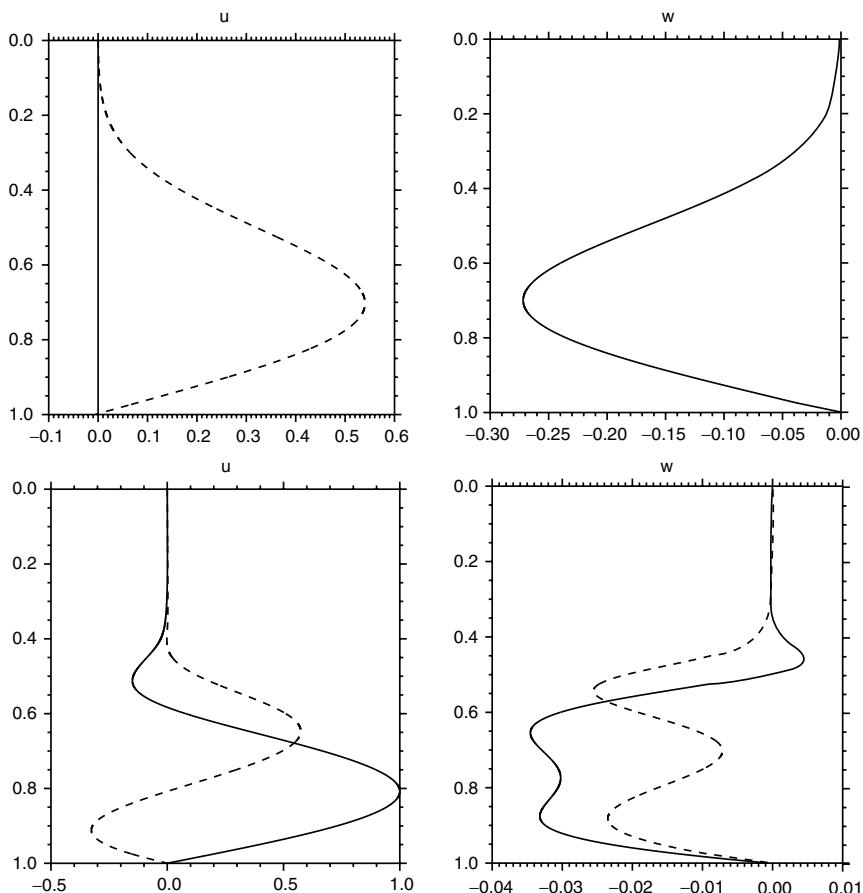


Figure 5. Comparison of the vertical dependence of the eigenfunctions for u and w for a magnetic buoyancy instability with no shear (top panels) and moderate shear (bottom panels). Here $k = 0.32$, $l = 8.0$ and $\delta = 10$. The addition of shear adds structure and suppresses the vertical motions (after Tobias and Hughes, 2003).

and magnetic buoyancy instabilities is very complicated, not least because they can act in a mutually suppressing manner. Strong magnetic fields may inhibit shear flow instabilities, while strong flows appear to inhibit magnetic buoyancy instabilities. The two instability mechanisms are very different and the preferred modes inhabit different regimes in wavenumber space; more specifically inviscid shear flow instabilities prefer to be two-dimensional while buoyancy instabilities are fully three-dimensional. The interplay between these two

instabilities will definitely be of interest in future linear and nonlinear studies.

The interaction of shear flows and magnetic buoyancy instabilities has also been examined numerically in the fully nonlinear regime by Brummell *et al.* (2002a) and Cline *et al.* (2003b). In these studies a depth-dependent, latitudinal toroidal shear acts by induction on an initially weak poloidal field to generate a strong toroidal field. This toroidal field is therefore not prescribed *a priori* but emerges in response to the shear flow (although of course the form of the toroidal field is very closely linked to the choice of shear profile). This strong magnetic field takes the form of thick oppositely signed tube-like structures that are unstable to magnetic buoyancy instabilities. The subsequent evolution of this instability is extremely complicated with many interactions between non-uniform flow and magnetic field taking place. Many subsequent configurations are possible depending on the magnetic Reynolds number (Rm) and other parameters in the flow, yet in all calculations highly twisted buoyant coherent magnetic structures are formed. Intriguingly the combination of magnetic buoyancy and twisted fields leads to the possibility of self-sustained nonlinear dynamo action (Cline *et al.*, 2003a). Once a magnetic field leaves the stable layer as a coherent structure, then its dynamics as it rises through the convection zone to reach the solar surface is of particular interest. However, this problem is beyond the scope of this review and the reader is directed elsewhere (see, e.g., Hughes and Falte, 1998; Wissink *et al.*, 2000; Emonet *et al.*, 2001; Fan *et al.*, 2003) for an in-depth discussion.

In summary, the role of magnetic buoyancy in the solar tachocline is vital to the dynamics there (particularly for the magnetic field). Magnetic buoyancy makes radial motions energetically possible and drives the flows away from two-dimensionality. Clearly a greater understanding of the interaction between the various instability mechanisms in the tachocline is necessary. Progress may be made by adopting a model that includes both simplified magnetic buoyancy effects (such as that studied by Spiegel and Weiss, 1982; Corfield, 1984) and latitudinal shear.

5. The tachocline and the solar dynamo

The formation and stability of the tachocline discussed in the previous two sections is of fundamental importance because the tachocline is believed to play a key role in the generation of the large-scale solar magnetic field and thus the solar activity cycle. The field of solar dynamo theory is a vast and complicated one and a complete review of the

key observational results and all theoretical contributions to the theory is well beyond the scope of this paper (the reader is directed to Weiss (1994), Weiss and Tobias (2000), and Ossendrijver (2003), for a full discussion of the models and observations and to Tobias (2002b) for a basic introduction to the fundamentals of the solar dynamo). Here I shall not discuss the observational evidence for dynamo action in the Sun as detailed reviews are available, but I shall concentrate on the postulated role of the tachocline in the solar dynamo, via interface dynamo models. Moreover, on the theoretical side, many fundamental questions of dynamo theory applicable to the Sun, such as the apparent catastrophic suppression of the transport coefficients in mean field theory (e.g., Diamond *et al.*, 2004) or the important role of magnetic helicity in generating large-scale magnetic fields will either not be discussed or touched upon briefly only when the tachocline may play a key role; again the reader is directed elsewhere for elucidation (Diamond *et al.*, 2004)

Briefly, the Sun has a large-scale magnetic field that is mainly dipolar, with a toroidal field, thought to be responsible for the formation of active regions, that is usually antisymmetric about the equator and a weaker poloidal field. Both the toroidal and poloidal components reverse, out-of-phase, with a mean period of eleven years, and the location for the formation of active regions migrates from mid latitudes (at the start of a solar cycle) to the equatorial regions (at the end of the cycle). Before the end of each cycle, sunspots for the new cycle, with reversed polarity, appear at mid latitudes so there is some small overlap between the magnetic field for neighboring cycles. Interestingly, the amplitude of the solar cycle is modulated on a timescale longer than eleven years (periods of roughly eighty and two-hundred years are visible in the sunspot and proxy records) and recurrent periods of reduced activity, known as Grand Minima, have interrupted solar activity (see, e.g., Tobias, 2002a, for an in-depth discussion of the modulation of cyclic activity). The systematic behavior of sunspots (their polarity and dependence of tilt angle on emergence latitude) suggests that they are the manifestation of a strong magnetic field generated deep within the Sun and are not surface generated. One may think of the toroidal field that leads to active regions as a wave of magnetic activity propagating from mid latitudes to the equator. Sunspot pairs are formed when the field goes unstable via a magnetic buoyancy instability (discussed earlier) to form loops that rise and emerge at the solar photosphere.

The fundamental question here is what maintains the solar magnetic field against the action of Ohmic dissipation and what processes can cause the magnetic field to reverse on a timescale much shorter than the diffusive timescale for the field? This question, though easily

posed, has proved difficult to answer, although a coherent picture of solar dynamo action is slowly emerging. To maintain the solar magnetic field two physical processes are often invoked. The first process is straightforward: differential rotation in an azimuthal flow is known to generate the linear growth of a toroidal field from a constant poloidal field (essentially from the $(\mathbf{B} \cdot \nabla) \mathbf{u}$ term in the induction equation). The reverse process, the regeneration of poloidal field from toroidal, is more problematic and subtle. In an early paper, Parker (1955) argued that the interaction between the turbulent eddies influenced by rotation in the solar convection zone and the toroidal field would yield a net EMF leading to a regeneration of poloidal flux. This idea of the net effect of the correlation of small-scale flows and small-scale magnetic fields leading to the generation of a large-scale field led to the mean-field formalism employed by Steenbeck *et al.* (1966) that has proved a subject of much intense research. Here the regeneration of large-scale poloidal field is governed by the α -effect, while the small-scale turbulence also acts so as to enhance the turbulent diffusivity of the magnetic field (parameterized by the β -effect). I repeat here that I will not discuss the many interesting issues associated with applying mean-field theory to the Sun, except where they have a direct consequence for the role of the tachocline in mean-field solar dynamos. It is important to stress though that these issues—in particular the behavior of the mean-field transport coefficients in the nonlinear regime—are fundamental. The inclusion of these extra effects leads to the mean-field dynamo equations that have proved so successful in modeling the large-scale solar magnetic field. Early dynamo models combined the latitudinal differential rotation visible at the solar surface with an α -effect throughout the convection zone (these models were termed distributed dynamos) to produce kinematic and nonlinear dynamo wave solutions that matched the observed fields with varying degrees of success. However, the discovery of the tachocline using helioseismic inversions combined with the theoretical realization that a weak large-scale mean field could suppress the α -effect (see later and Vainshtein and Cattaneo, 1992; Diamond *et al.*, 2004) changed the paradigm for dynamo action and a new class of dynamo models were needed—the interface dynamo.

5.1. The interface dynamo

The discovery of the tachocline as an extremely thin, strong layer of radial differential rotation at the base of the solar convection zone obviously had important consequences for solar dynamo theory. As noted

earlier, differential rotation is the key ingredient in the generation of a strong toroidal magnetic field. It had earlier been suggested that the base of the solar convection zone formed a suitable place for storage of the toroidal field because the stable stratification there would allow the generation of magnetic field on a long enough timescale before flux was lost due to the actions of magnetic buoyancy (Galloway and Weiss, 1981). Solar dynamo modelers were therefore starting to construct solar dynamo models with a strong shear at the base of the solar convection zone (Deluca and Gilman, 1986). Fortunately the discovery of the tachocline also allowed solar dynamo theory to circumvent the objection of catastrophic quenching of the mean-field α -effect. Vainshtein and Cattaneo (1992) argued that the α -effect of mean-field theory was suppressed in the presence of even a very weak large-scale magnetic field; weak in this context implies one whose magnetic energy is up to a factor of the magnetic Reynolds number smaller than equipartition with the kinetic energy. The argument about the precise nature of this quenching continues today, not only for the α -effect but also for the turbulent diffusivity. If the large-scale toroidal field were generated throughout the region of cyclonic turbulence in the solar convection zone then the α -effect (and hence the dynamo) would switch off for mean field strengths far below those believed to be necessary to explain the formation of active regions. This constraint therefore appeared to signal the end of distributed dynamos.

Parker (1993) devised a solution to this problem that made use of the recent discovery of the tachocline. He envisaged a dynamo that was primarily generated at the interface between the solar convection zone and the radial shear of the tachocline. His model, although kinematic, contains all the fundamental physical ideas for a dynamo of this type to operate. In his Cartesian model (schematically shown in Fig. 6), the convection zone is parameterized as possessing a strong α -effect owing to the cyclonic motions there. The presence of turbulence in the convection zone also leads to a significant contribution to the turbulent transport of magnetic flux (which in the mean-field framework can be most simply parameterized by the addition of a turbulent diffusivity β).

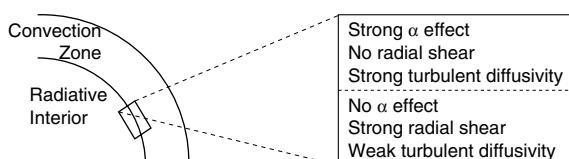


Figure 6. Schematic representation of the interface dynamo (after Parker, 1993).

In the convection zone the radial shear is small (recall that the differential rotation is largely constant on radial lines throughout the convection zone) and Parker's original local model ignored any effects of latitudinal differential rotation. Below the interface at the base of the convection zone the tachocline was modeled as a relatively quiescent layer of strong radial shear. In this case both the α -effect and turbulent diffusivity were switched off in the lower layer and the dynamics was controlled by the Ω -effect of the large-scale shear. Parker argued that this configuration *may* be able to circumvent the catastrophic α -quenching proposed by Vainshtein and Cattaneo (1992). The important result is that the precise combination of reduced diffusivity in the lower layer and enhanced transport of flux in the upper layer, leads to the formation of a strong mean toroidal field located just below the interface. This strong mean field, because it is stored away from the region of cyclonic convection, does not suppress the α -effect. (An interesting observation, bearing in mind the increasing importance placed on magnetic helicity conservation for dynamo action, is that the interface scenario allows the storage of large-scale magnetic helicity of one sign in the tachocline away from the small-scale magnetic helicity in the convection zone.) The much weaker poloidal field generated in the lower reaches of the convection zone has little effect on α -suppression and is transported into the tachocline by the turbulence. Parker could not, however, explore the dynamic role of the magnetic field (if any) in bringing about this scenario given the kinematic nature of his model. He did note that the balance for interface dynamos is a delicate one — recall that the α -effect and turbulent diffusivity may be quenched by a factor $Rm(B)^2$ and Rm depends not only on the strength of the turbulence but also the length-scale of the turbulence and the molecular diffusivity. All of these factors may play a role in the balance of the interface dynamo.

The key ingredients for the dynamo to operate can be summarized as follows:

- Strong radial shear at the base of the convection zone — provided by the tachocline.
- A mechanism for regenerating poloidal field operating in the lower reaches of the convection zone — Parker envisaged this as helical turbulence, but other possible mechanisms have also been suggested (see later).
- An efficient downward transport of magnetic flux from the convection zone into the stable layer below — this was modeled by Parker as a turbulent diffusion, but some other physical processes could be equally important. Indeed a robust

efficient mechanism for flux transport from the convection zone into the tachocline, known as magnetic pumping, has recently been identified and is discussed below.

- A reduced transport of magnetic field in the stably stratified lower layer — this could be a purely hydrodynamic effect (e.g., the suppression of turbulence by strong stratification or shear flows) or a magnetohydrodynamic effect where the turbulent diffusivity is suppressed by the back-reaction of the strong toroidal field.

Further mean-field interface dynamos in the kinematic and nonlinear regime have proved very successful at reproducing many facets of solar magnetic activity (described at length in Tobias, 2002a, but summarized briefly here). The consequences of including a primordial field in the solar interior, believed to be responsible for the solid-body rotation of the radiative core, in a kinematic interface dynamo model was investigated by Boruta (1996) who utilized the interface model to place an upper bound of around 30G on the magnetic field in the Sun's core. Interface dynamos have also been shown in a series of papers (Charbonneau and MacGregor, 1996; MacGregor and Charbonneau, 1997; Charbonneau and MacGregor, 1997) successfully to avoid the problem of catastrophic α -quenching. In their nonlinear mean-field model, the α -effect is catastrophically quenched by a factor $Rm(B)^2$, but solutions can be found that successfully store the magnetic field in the tachocline. Moreover, weak diffusivity quenching has been shown to lead naturally to layering and confinement of magnetic fields to thin shells at the interface (Tobias, 1996a). Further modifications and generalizations of the interface scenario in the nonlinear regime have been studied by a number of authors (Tobias, 1996b, 1997; Roald, 1998; Markiel and Thomas, 1999) who have demonstrated, inter alia, that the precise form of the differential rotation prescribed in the mean-field model is not crucial to its operation, that the addition of a meridional flow from pole to equator near the tachocline can encourage the equatorward migration of dynamo waves, and that modulation of the basic cycle in interface dynamos arises naturally as a result of the dynamic back-reaction of the mean toroidal field on the mean differential rotation. More recently, Mason *et al.* (2002) have demonstrated that an interface dynamo operating at the base of the solar convection zone in conjunction with the radial shear layer of the tachocline is more efficient than a (Babcock-Leighton) dynamo where the α -effect operates solely at the solar surface (Durney, 2000).

In addition, there are many variants of the interface model, all of which place the dynamo at the base of the solar convection in or near to the tachocline. A number of authors have derived a “dynamic α -effect” generated by the interaction of buoyantly unstable magnetic fields with the Coriolis force. These small-scale magnetic fields emerge owing to either the instability of isolated magnetic flux tubes (Schmitt *et al.*, 1996) or of a magnetic layer (Thelen, 2000*a,b*). Here the tachocline will play an important role not only in generating the field, but also in modifying the dynamic α -effect via the injection of helicity into the field. As noted earlier, Dikpati and Gilman (2001*b*) consider an α -effect due to the instability of a shallow water model to global modes.

The interface model as proposed is very appealing, and clearly the tachocline plays a key role not only in generating the magnetic field and storing it away from the turbulence but also potentially in modifying turbulent transport properties. It is important, however, to examine systematically the small-scale physics that is inevitably parameterized in such global mean-field interface models. There are many unanswered questions at a fundamental level. What is the nature of the transport of magnetic flux from the lower reaches of the convection zone into the tachocline? Recent simulations of the interaction of turbulent penetrative convection (see, e.g., Brummell *et al.*, 2002*b*) with magnetic fields have revealed a robust transport mechanism for magnetic flux termed magnetic pumping (Tobias *et al.*, 1998, 2001; Dorch and Nordlund, 2001, and the references therein). This mechanism may play an important role in interface dynamos even in the absence of conventional turbulent diffusion. Another important question is the role of the turbulence in setting the length-scale of the generated magnetic fields. Mean-field models where the turbulence and differential rotation are confined to thin layers, say the width of the tachocline (Rüdiger and Brandenburg, 1995), tend to produce magnetic fields with latitudinal length-scales too small to be compatible with the observations of active regions at the solar surface. Is the length-scale therefore determined by convective turbulence in the convection zone, or by anisotropic turbulence in the tachocline leading to latitudinal spreading of magnetic field belts? The role of anisotropic turbulence in thin layer interface dynamos is currently being investigated. Finally what is the role of the tachocline itself in modifying transport of magnetic flux? We have seen earlier that shear layers can inhibit magnetic buoyancy instabilities (a form of buoyant transport) and it is also known that shear flows can inhibit turbulent diffusion of magnetic fields in two-dimensional turbulence (Kim and Dubrulle, 2001; Kim *et al.*, 2001). It is possible that the

role of the shear could be crucial in maintaining the delicate balance of interface dynamos. Further investigation of all these questions at the level of small-scale interactions is under way and is clearly vital to our understanding of the solar dynamo.

I will conclude this section with a brief discussion of the possibility of interface-type dynamos operating in other stars. Stellar magnetic fields can be observed both directly and using proxy data and the dynamical magnetic behavior of these stars can be used to calibrate dynamo theories (Rosner, 2000). Stars with transitions between convective envelopes and radiative interiors are likely to possess thin transition regions similar in nature to the solar tachocline. The possibility of detecting stellar tachoclines using astroseismology is discussed in Mazumdar and Antia (2001). These tachoclines in combination with the cyclonic convection zone may lead naturally to interface scenarios in both moderately and rapidly rotating stars. The properties of dynamos in moderately rotating stars have also been investigated within the interface dynamo scenario by Tobias (1998) and Montesinos *et al.* (2001) who describe the relationship between the period of the magnetic cycle and the rotation rate of the star, and the dangers of relating such dynamo calculations to the observations. Lanza *et al.* (2000) use an interface dynamo argument to explain the gap in the distribution of the orbital periods of cataclysmic variables. Here the dynamo is believed to be operating at the base of the convective envelope of the secondary component, and such a dynamo can lead naturally to a bimodal distribution in periods. Blackman *et al.* (2001) consider interface dynamos in asymptotic-giant-branch stars where the dynamo operating at the interface between the rapidly rotating core and the more slowly rotating envelope of the star may lead to the formation of planetary nebulae. Finally, Bushby (2003) considers the dynamo properties of rapidly rotating late-type stars. In this case the rotational frequency is constant on cylindrical surfaces in the convection zone and the tangent cylinder plays an important role in determining the differential rotation profile. The inner core is again assumed to rotate as a solid body and the form of the tachocline is therefore somewhat different than for solar-type stars, being concentrated at high latitudes. The dynamo then has a tendency to produce polar magnetic fields. An interesting question is whether the role of the tachocline is changed in stars with very deep convection zones. In fully convective stars, with no tachocline, the dynamo must have a different character and the structure of the generated magnetic fields is likely to be very different. In all of the examples cited above, a tachocline-like region plays a significant role in the generation of the large-scale mag-

netic field. Clearly further investigation of the properties of tachoclines is necessary to facilitate a greater understanding of dynamo processes in both the Sun and other stars.

6. Discussion and conclusion

In this review I have summarized our understanding of the solar tachocline. Considerable progress has been made in addressing the important questions of the formation and stability of the tachocline, although much remains unclear. Significant advances have been made by drawing on analogies with geophysical flows, both for the stability properties of the tachocline and for the role of turbulence in transporting angular momentum. However, it should be stressed that in astrophysical fluid dynamics, magnetic fields often play an important role (even if at first sight their effects appear to be small). I believe that in (at least some parts of) the tachocline magnetic fields are important. The inclusion of magnetic buoyancy in tachocline models will drive the system away from two-dimensional dynamics by promoting radial motions and will clearly have important consequences for the stability of the tachocline. This process requires further investigation, especially as the radial shear in the tachocline may act so as to modify the nature of this instability. Moreover, the role of magnetic fields in modifying the form of any shear turbulence and the resulting angular momentum transport may prove crucial in determining the nature of the differential rotation in the tachocline.

The tachocline is also a vital component of the solar dynamo that generates the large-scale magnetic field. The tachocline not only provides the mechanism for the generation of strong large-scale toroidal fields through the action of large-scale shear, but also provides the perfect location for the storage of the mean fields that lead to large-scale solar activity. Interface dynamos appear successfully to circumvent catastrophic quenching of the turbulent transport and generation mechanisms and to allow the generation of significant (near-equipartition) mean fields. Moreover, mean-field interface dynamo models *have* proved very successful in explaining generic features of the solar magnetic cycle—though now a deeper understanding of the mechanism of such models is necessary. There is certainly a limit to the depth of understanding that can be gained from analyzing the results of these global models. What is required is a greater understanding of the nature of interface dynamo models at a micro-physics level. What is the role

of the small-scale interactions in modifying transport phenomena, for example, and how are these modified by the presence of shear and magnetic fields in the tachocline? It is only by determining systematically the properties of magnetohydrodynamic turbulence, whether in stably stratified or convectively unstable domains, that we may gain an insight into tachocline and convection zone dynamics and hence into the generation of magnetic field in the Sun.

Acknowledgments

I am very grateful to Pat Diamond, Douglas Gough, David Hughes, Michael McIntyre, Michael Thompson and Nigel Weiss for helpful discussions.

References

- Acheson, D. J., "On the instability of toroidal magnetic fields and differential rotation in stars," *Phil. Trans. R. Soc. Lond. A* **289**, 459–500 (1978).
- Adam, J. A., "Stability of aligned magnetoatmospheric flow," *J. Plasma Phys.* **19**, 77–86 (1978).
- Balbus, S. A. and Hawley, J. F., "A powerful local shear instability in weakly magnetized disks. I-Linear analysis. II-Nonlinear evolution," *Astrophys. J.* **376**, 214–233 (1991).
- Balbus, S. A. and Hawley, J. F., "Instability, turbulence, and enhanced transport in accretion disks," *Rev. Modern Phys.* **70**, 1–53 (1998).
- Basu, S. and Antia, H. M., "A study of possible temporal and latitudinal variations in the properties of the solar tachocline," *Mon. Not. R. Astron. Soc.* **324**, 498–508 (2001).
- Basu, S. and Schou, J., "Does the tachocline show solar cycle related changes?," *Solar Phys.* **192**, 481–486 (2000).
- Blackman, E. G., Frank, A., Markiel, J. A., Thomas, J. H. and Van Horn, H. M., "Dynamos in asymptotic-giant-branch stars as the origin of magnetic fields shaping planetary nebulae," *Nature* **409**, 485–487 (2001).
- Boruta, N., "Solar dynamo surface waves in the presence of a primordial magnetic field: A 30 gauss upper limit in the solar core," *Astrophys. J.* **458**, 832–849 (1996).
- Brown, T. M., Christensen-Dalsgaard, J., Dziembowski, W. A., Goode, P., Gough, D. O. and Morrow, C. A., "Inferring the sun's internal angular velocity from observed p-mode frequency splittings," *Astrophys. J.* **343**, 526–546 (1989).

- Brummell, N., Cline, K. and Cattaneo, F., "Formation of buoyant magnetic structures by a localized velocity shear," *Mon. Not. R. Astron. Soc.* **329**, L73–L76 (2002a).
- Brummell, N. H., Clune, T. L. and Toomre, J., "Penetration and overshooting in turbulent compressible convection," *Astrophys. J.* **570**, 825–854 (2002b).
- Brun, A. S. and Toomre, J., "Turbulent convection under the influence of rotation: Sustaining a strong differential rotation," *Astrophys. J.* **570**, 865–885 (2002).
- Bushby, P. J., "Modelling dynamos in rapidly rotating late-type stars," *Mon. Not. R. Astron. Soc.* **342**, L15–L19 (2003).
- Cally, P. S., "A sufficient condition for instability in a sheared incompressible magnetofluid," *Solar Phys.* **194**, 189–196 (2000).
- Cally, P. S., "Nonlinear evolution of 2d tachocline instabilities," *Solar Phys.* **199**, 231–249 (2001).
- Cally, P. S., "Three-dimensional magneto-shear instabilities in the solar tachocline," *Mon. Not. R. Astron. Soc.* **339**, 957–972 (2003).
- Cally, P. S., Dikpati, M. and Gilman, P. A., "Clamshell and tipping instabilities in a two-dimensional magnetohydrodynamic Tachocline," *Astrophys. J.* **582**, 1190–1205 (2003).
- Carrington, R. C., "On the distribution of the solar spots in latitudes since the beginning of the year 1854, with a map," *Mon. Not. R. Astron. Soc.* **19**, 1–3 (1858).
- Chandrasekhar, S., *Hydrodynamic and Hydromagnetic Stability*, International Series of Monographs on Physics, Oxford: Clarendon (1961).
- Charbonneau, P. and MacGregor, K. B., "On the generation of equipartition-strength magnetic fields by turbulent hydromagnetic dynamos," *Astrophys. J. Lett.* **473**, L59 (1996).
- Charbonneau, P. and MacGregor, K. B., "Solar interface dynamos. II. Linear, kinematic models in spherical geometry," *Astrophys. J.* **486**, 502–520 (1997).
- Charbonneau, P., Christensen-Dalsgaard, J., Henning, R., Larsen, R. M., Schou, J., Thompson, M. J. and Tomczyk, S., "Helioseismic constraints on the structure of the solar tachocline," *Astrophys. J.* **527**, 445–460 (1999a).
- Charbonneau, P., Dikpati, M. and Gilman, P. A., "Stability of the solar latitudinal differential rotation inferred from helioseismic data," *Astrophys. J.* **526**, 523–537 (1999b).
- Charbonneau, P., Tomczyk, S., Schou, J. and Thompson, M. J., "The rotation of the solar core inferred by genetic forward modeling," *Astrophys. J.* **496**, 1015–1030 (1998).
- Cline, K. S., Brummell, N. H. and Cattaneo, F., "Dynamo action driver by shear and magnetic buoyancy," *Astrophys. J.* **599**, 1449–1468 (2003a).
- Cline, K. S., Brummell, N. H. and Cattaneo, F., "On the formation of magnetic structures by the combined action of velocity shear and magnetic buoyancy," *Astrophys. J.* **588**, 630–644 (2003b).

- Corfield, C. N., "The magneto-boussinesq approximation by scale analysis," *Geophys. Astrophys. Fluid Dynam.* **29**, 19–28 (1984).
- Cowling, T. G., *Magnetohydrodynamics*, Interscience, New York (1957).
- Deluca, E. E. and Gilman, P. A., "Dynamo theory for the interface between the convection zone and the radiative interior of a star. I. Model equations and exact solutions," *Geophys. Astrophys. Fluid Dynam.* **37**, 85–127 (1986).
- Diamond, P. H., Hughes, D. and Kim, E., "Self-consistent mean field electrodynamics in two and three dimensions," this volume, Chapter 6 (2004).
- Dikpati, M. and Gilman, P. A., "Joint instability of latitudinal differential rotation and concentrated toroidal fields below the solar convection zone," *Astrophys. J.* **512**, 417–441 (1999).
- Dikpati, M. and Gilman, P. A., "Analysis of hydrodynamic stability of solar tachocline latitudinal differential rotation using a shallow-water model," *Astrophys. J.* **551**, 536–564 (2001a).
- Dikpati, M. and Gilman, P. A., "Flux-transport dynamos with α -effect from global instability of tachocline differential rotation: A solution for magnetic parity selection in the sun," *Astrophys. J.* **559**, 428–442 (2001b).
- Dorch, S. B. F. and Nordlund, Å., "On the transport of magnetic fields by solar-like stratified convection," *Astron. Astrophys.* **365**, 562–570 (2001).
- Drazin, P. G. and Reid, W. H., *Hydrodynamic Stability*, Cambridge University Press, Cambridge (1981).
- Durney, B. R., "On the torsional oscillations in Babcock-Leighton solar dynamo models," *Solar Phys.* **196**, 1–18 (2000).
- Dziembowski, W. and Kosovichev, A., "Low frequency oscillations in slowly rotating stars — Part Three — Kelvin-Helmholtz instability," *Acta Astronomica* **37**, 341–348 (1987).
- Elliott, J. R., "Aspects of the solar tachocline," *Astron. Astrophys.* **327**, 1222–1229 (1997).
- Elliott, J. R. and Gough, D. O., "Calibration of the thickness of the solar tachocline," *Astrophys. J.* **516**, 475–481 (1999).
- Emonet, T., Moreno-Insertis, F. and Rast, M. P., "The zigzag path of buoyant magnetic tubes and the generation of vorticity along their periphery," *Astrophys. J.* **549**, 1212–1220 (2001).
- Fan, Y., Abbott, W. P. and Fisher, G. H., "The dynamic evolution of twisted magnetic flux tubes in a three-dimensional convecting flow. I. Uniformly buoyant horizontal tubes," *Astrophys. J.* **582**, 1206–1219 (2003).
- Fjørtoft, R., "Application of integral theorems in deriving criteria of stability for laminar flows and for the baroclinic circular vortex," *Geophys. Publ. Oslo* **17**, 1–52 (1950).
- Galloway, D. J. and Weiss, N. O., "Convection and magnetic fields in stars," *Astrophys. J.* **243**, 945–953 (1981).
- Garaud, P., "Latitudinal shear instability in the solar tachocline," *Mon. Not. R. Astron. Soc.* **324**, 68–76 (2001).

- Garaud, P., "Dynamics of the solar tachocline — I. An incompressible study," *Mon. Not. R. Astron. Soc.* **329**, 1–17 (2002).
- Gilman, P. A., "Magnetohydrodynamic ‘shallow water’ equations for the solar tachocline," *Astrophys. J. Lett.* **544**, L79–L82 (2000).
- Gilman, P. A. and Dikpati, M., "Joint instability of latitudinal differential rotation and concentrated toroidal fields below the solar convection zone. II. Instability of narrow bands at all latitudes," *Astrophys. J.* **528**, 552–572 (2000).
- Gilman, P. A. and Dikpati, M., "Analysis of instability of latitudinal differential rotation and toroidal field in the solar tachocline using a magnetohydrodynamic shallow-water model. I. Instability for broad toroidal field profiles," *Astrophys. J.* **576**, 1031–1047 (2002).
- Gilman, P. A. and Fox, P. A., "Joint instability of latitudinal differential rotation and toroidal magnetic fields below the solar convection zone," *Astrophys. J.* **484**, 439–454 (1997).
- Gilman, P. A. and Fox, P. A., "Joint instability of latitudinal differential rotation and toroidal magnetic fields below the solar convection zone. II. Instability for toroidal fields that have a node between the equator and pole," *Astrophys. J.* **510**, 1018–1044 (1999).
- Gilman, P. A. and Fox, P. A., "Erratum: Joint instability of latitudinal differential rotation and toroidal magnetic fields below the solar convection zone. II. Instability for toroidal fields that have a node between the equator and pole," *Astrophys. J.* **534**, 1020 (2000).
- Gough, D., "Random remarks on solar hydrodynamics," in: *IAU Colloq. 36: The Energy Balance and Hydrodynamics of the Solar Chromosphere and Corona* (Eds. R.-M. Bonnet and Ph. Delache), pp. 3 (1977).
- Gough, D. and McIntyre, M. E., "Inevitability of a magnetic field in the Sun’s radiative interior," *Nature* **394**, 755–757 (1998).
- Howard, L. N., "Note on a paper of John W. Miles," *J. Fluid Mech.* **10**, 509–512 (1961).
- Howe, R., Christensen-Dalsgaard, J., Hill, F., Komm, R. W., Larsen, R. M., Schou, J., Thompson, M. J. and Toomre, J., "Dynamic variations at the base of the solar convection zone," *Science* **287**, 2456–2460 (2000).
- Hughes, D. W., "Magnetic buoyancy," in: *Advances in Solar System Magnetohydrodynamics* (Eds. E. R. Priest and A. W. Hood), pp. 77–104. Cambridge University Press (1991).
- Hughes, D. W. and Cattaneo, F., "A new look at the instability of a horizontal field," *Geophys. Astrophys. Fluid Dynam.* **39**, 65 (1987).
- Hughes, D. W. and Falle, S. A. E. G., "The rise of twisted magnetic flux tubes: A high reynolds number adaptive grid calculation," *Astrophys. J. Lett.* **509**, L57–L60 (1998).
- Hughes, D. W. and Proctor, M. R. E., "Magnetic fields in the solar convection zone — Magnetoconvection and magnetic buoyancy," *Annu. Rev. Fluid Mech.* **20**, 187–223 (1988).
- Hughes, D. W. and Tobias, S. M., "On the instability of magnetohydrodynamic shear flows," *Proc. R. Soc. Lond. A* **457**, 1365–1384 (2001).

- Kim, E. and Dubrulle, B., "Turbulent transport and equilibrium profiles in two-dimensional magnetohydrodynamics with background shear," *Phys. Plasmas* **8**, 813–824 (2001).
- Kim, E. and MacGregor, K. B., "Gravity wave-driven flows in the solar tachocline," *Astrophys. J. Lett.* **556**, L117–L120 (2001).
- Kim, E., Hahm, T. S. and Diamond, P. H., "Eddy viscosity and laminarization of sheared flow in three-dimensional reduced magnetohydrodynamic turbulence," *Phys. Plasmas* **8**, 3576–3582 (2001).
- Knobloch, E. and Spruit, H. C., "Stability of differential rotation in stars," *Astron. Astrophys.* **113**, 261–268 (1982).
- Kosovichev, A. G., Schou, J., Scherrer, P. H., Bogart, R. S., Bush, R. I., Hoeksema, J. T., Aloise, J., Bacon, L., Burnette, A., de Forest, C., Giles, P. M., Leibrand, K., Nigam, R., Rubin, M., Scott, K., Williams, S. D., Basu, S., Christensen-Dalsgaard, J., Dappen, W., Rhodes, E. J., Duvall, T. L., Howe, R., Thompson, M. J., Gough, D. O., Sekii, T., Toomre, J., Tarbell, T. D., Title, A. M., Mathur, D., Morrison, M., Saba, J. L. R., Wolfson, C. J., Zayer, I. and Milford, P. N., "Structure and rotation of the solar interior: initial results from the MDI medium-L program," *Solar Phys.* **170**, 43–61 (1997).
- Lanza, A. F., Rodonò, M. and Rosner, R., "Dynamo action and the period gap in cataclysmic variables," *Mon. Not. R. Astron. Soc.* **314**, 398–402 (2000).
- Libbrecht, K. G., "Solar p-mode frequency splittings," in: *ESA SP-286: Seismology of the Sun and Sun-Like Stars*, pp. 131–136 (1988).
- MacGregor, K. B. and Charbonneau, P., "Solar interface dynamos. I. Linear, kinematic models in cartesian geometry," *Astrophys. J.* **486**, 484 (1997).
- MacGregor, K. B. and Charbonneau, P., "Angular momentum transport in magnetized stellar radiative zones. IV. Ferraro's theorem and the solar tachocline," *Astrophys. J.* **519**, 911–917 (1999).
- Markiel, J. A. and Thomas, J. H., "Solar interface dynamo models with a realistic rotation profile," *Astrophys. J.* **523**, 827–837 (1999).
- Mason, J., Hughes, D. W. and Tobias, S. M., "The competition in the solar dynamo between surface and deep-seated α -effects," *Astrophys. J. Lett.* **580**, L89–L92 (2002).
- Mazumdar, A. and Antia, H. M., "Seismic detection of stellar tachoclines," *Astron. Astrophys.* **368**, L8–L12 (2001).
- McIntyre, M. E., "Solar tachocline dynamics: eddy viscosity, anti-friction, or something in between," in: *Stellar Astrophysical Fluid Dynamics* (Eds. M. J. Thompson and J. Christensen-Dalsgaard), pp. 111–130, Cambridge University Press (2003).
- Mestel, L., *Stellar Magnetism*, Oxford: Clarendon (International Series of Monographs on Physics; 99) (1999).
- Mestel, L. and Weiss, N. O., "Magnetic fields and non-uniform rotation in stellar radiative zones," *Mon. Not. R. Astron. Soc.* **226**, 123–135 (1987).
- Miesch, M. S., "Numerical modeling of the solar tachocline. I. Freely evolving stratified turbulence in a thin rotating spherical shell," *Astrophys. J.* **562**, 1058–1075 (2001).

- Miesch, M. S., "Numerical modeling of the solar Tachocline. II. Forced turbulence with imposed shear," *Astrophys. J.* **586**, 663–684 (2003).
- Montesinos, B., Thomas, J. H., Ventura, P. and Mazzitelli, I., "A new look at the relationship between activity, dynamo number and Rossby number in late-type stars," *Mon. Not. R. Astron. Soc.* **326**, 877–884 (2001).
- Newcomb, W. A., *Phys. Fluids* **4**, 391 (1961).
- Ogilvie, G. I. and Pringle, J. E., "The non-axisymmetric instability of a cylindrical shear flow containing an azimuthal magnetic field," *Mon. Not. R. Astron. Soc.* **279**, 152–164 (1996).
- Ossendrijver, M., "The solar dynamo," *Astron. Astrophys. Rev.* **11**, 287–367 (2003).
- Parker, E. N., "Hydromagnetic dynamo models," *Astrophys. J.* **122**, 293–314 (1955).
- Parker, E. N., "The dynamical state of the interstellar gas and field," *Astrophys. J.* **145**, 811–833 (1966).
- Parker, E. N., "A solar dynamo surface wave at the interface between convection and nonuniform rotation," *Astrophys. J.* **408**, 707–719 (1993).
- Roald, C. B., "A two-layer alphaomega-dynamo model with dynamic feedback on the omega-effect," *Mon. Not. R. Astron. Soc.* **300**, 397–410 (1998).
- Rosner, R., "Magnetic fields of stars: using stars as tools for understanding the origins of cosmic magnetic fields," *Phil. Trans. R. Soc. Lond. A* **358**, 689–708 (2000).
- Rüdiger, G. and Brandenburg, A., "A solar dynamo in the overshoot layer: cycle period and butterfly diagram," *Astron. Astrophys.* **296**, 557 (1995).
- Rüdiger, G. and Kitchatinov, L. L., "The slender solar tachocline: a magnetic model," *Astron. Nachr.* **318**, 273 (1997).
- Schatzman, E., Zahn, J.-P. and Morel, P., "Shear turbulence beneath the solar tachocline," *Astron. Astrophys.* **364**, 876–878 (2000).
- Schmitt, D., Schuessler, M. and Ferriz-Mas, A., "Intermittent solar activity by an on-off dynamo," *Astron. Astrophys.* **311**, L1–L4 (1996).
- Schou, J., Antia, H. M., Basu, S., Bogart, R. S., Bush, R. I., Chitre, S. M., Christensen-Dalsgaard, J., di Mauro, M. P., Dziembowski, W. A., Eff-Darwich, A., Gough, D. O., Haber, D. A., Hoeksema, J. T., Howe, R., Korzennik, S. G., Kosovichev, A. G., Larsen, R. M., Pijpers, F. P., Scherrer, P. H., Sekii, T., Tarbell, T. D., Title, A. M., Thompson, M. J. and Toomre, J., "Helioseismic studies of differential rotation in the solar envelope by the solar oscillations investigation using the michelson doppler imager," *Astrophys. J.* **505**, 390–417 (1998).
- Spiegel, E. A. and Weiss, N. O., "Magnetic buoyancy and the Boussinesq approximation," *Geophys. Astrophys. Fluid Dynam.* **22**, 219–234 (1982).
- Spiegel, E. A. and Zahn, J.-P., "The solar tachocline," *Astron. Astrophys.* **265**, 106–114 (1992a).
- Spiegel, E. A. and Zahn, J.-P., "The turbulent Tacholine," in: *Cool Stars, Stellar Systems, and the Sun*, Proceedings of the 7th Cambridge Workshop, ASP Conference Series (ASP: San Francisco), 26, p. 188 (1992b).

- Steenbeck, M., Krause, F. and Rädler, K. H., "A calculation of the mean electromotive force in an electrically conducting fluid in turbulent motion under the influence of coriolis forces," *Zeitschrift Naturforschung Teil A* **21**, 369–376 (1966).
- Stix, M., *The Sun: an Introduction*, Berlin: Springer (2002).
- Talon, S., Zahn, J.-P., Maeder, A. and Meynet, G., "Rotational mixing in early-type stars: the main-sequence evolution of a $9M_{\odot}$ -star," *Astron. Astrophys.* **322**, 209–217 (1997).
- Thelen, J.-C., "A mean electromotive force induced by magnetic buoyancy instabilities," *Mon. Not. R. Astron. Soc.* **315**, 155–164 (2000a).
- Thelen, J.-C., "Non-linear $\alpha\omega$ -dynamos driven by magnetic buoyancy," *Mon. Not. R. Astron. Soc.* **315**, 165–183 (2000b).
- Thompson, M. J., "Helioseismology present and future," in: *Proceedings of SOHO 12/GONG+ 2002. Local and Global Helioseismology: the Present and Future*, 27 October–1 November 2002, Big Bear Lake, CA, USA (Ed. H. Sawaya-Lacoste), ESA SP-517, pp. 209–215, Noordwijk, Netherlands: ESA Publications Division, ISBN 92-9092-827-1 (2003).
- Tobias, S. M., "Diffusivity quenching as a mechanism for Parker's surface dynamo," *Astrophys. J.* **467**, 870–880 (1996a).
- Tobias, S. M., "Grand minima in nonlinear dynamos," *Astron. Astrophys.* **307**, L21–L24 (1996b).
- Tobias, S. M., "The solar cycle: parity interactions and amplitude modulation," *Astron. Astrophys.* **322**, 1007–1017 (1997).
- Tobias, S. M., "Relating stellar cycle periods to dynamo calculations," *Mon. Not. R. Astron. Soc.* **296**, 653–661 (1998).
- Tobias, S. M., "Modulation of solar and stellar dynamos," *Astron. Nachr.* **323**, 417–423 (2002a).
- Tobias, S. M., "The solar dynamo," *Phil. Trans. R. Soc. Lond. A* **360**, 2741–2756 (2002b).
- Tobias, S. M. and Hughes, D. W., "The influence of velocity shear on magnetic buoyancy instability in the solar tachocline," *Astrophys. J. submitted* (2003).
- Tobias, S. M., Brummell, N. H., Clune, T. L. and Toomre, J., "Pumping of magnetic fields by turbulent penetrative convection," *Astrophys. J. Lett.* **502**, L177–L180 (1998).
- Tobias, S. M., Brummell, N. H., Clune, T. L. and Toomre, J., "Transport and storage of magnetic field by overshooting turbulent compressible convection," *Astrophys. J.* **549**, 1183–1203 (2001).
- Toomre, J., "Overview—where do we stand with helioseismology?", in: *Proceedings of SOHO 12/GONG+ 2002. Local and Global Helioseismology: the Present and Future*, 27 October–1 November 2002, Big Bear Lake, CA, USA (Ed. H. Sawaya-Lacoste), pp. 3–14, ESA SP-517, Noordwijk, Netherlands: ESA Publications Division, ISBN 92-9092-827-1 (2003).
- Vainshtein, S. I. and Cattaneo, F., "Nonlinear restrictions on dynamo action," *Astrophys. J.* **393**, 165–171 (1992).
- Velikhov, E. P., *Sov. Phys. JETP* **36**, 995 (1959).

- Watson, M., "Shear instability of differential rotation in stars," *Geophys. Astrophys. Fluid Dynam.* **16**, 285–298 (1981).
- Weiss, N. O., "Solar and stellar dynamos," in: *Lectures on Solar and Planetary Dynamos* (Eds. M. R. E. Proctor and A. D. Gilbert), pp. 59–96, Cambridge University Press (1994).
- Weiss, N. O. and Tobias, S. M., "Physical causes of solar activity," *Space Sci. Rev.* **94**, 99–112 (2000).
- Wissink, J. G., Matthews, P. C., Hughes, D. W. and Proctor, M. R. E., "Numerical simulations of buoyant magnetic flux tubes," *Astrophys. J.* **536**, 982–997 (2000).
- Zahn, J.-P., Talon, S. and Matias, J., "Angular momentum transport by internal waves in the solar interior," *Astron. Astrophys.* **322**, 320–328 (1997).

8 Magnetoconvection

M.R.E. Proctor

*Department of Applied Mathematics and Theoretical Physics,
University of Cambridge, Centre for Mathematical Sciences,
Wilberforce Road, Cambridge CB3 0WA, UK*

1. Introduction

Magnetoconvection, the study of the interaction between thermal convection and magnetic fields, was originally motivated by the dynamics of magnetic fields in the solar photosphere. [Figure 1](#) (see color insert following page 234) shows a recent image from the Swedish Solar Telescope (courtesy of T.E. Berger) showing the solar granulation, sunspots (dark) and bright points in downdrafts between supergranular and intergranular lanes. The last two of these are signatures of magnetic field structures. The challenge to theorists and numericists is to understand these structures, to relate them to observations of emission spectra etc., to use them as diagnostics for behavior deeper down in the solar convection zone, and ultimately to relate the phenomena to the solar dynamo responsible for the sunspot cycle. This last is beyond the scope of the present review, but I have tried to survey in some breadth work done in the last twenty years on the effects of an imposed field. In fact, interest in magnetoconvection as a bifurcation problem, similar to double-diffusive convection, has led to a large number of papers that are not directly astrophysical, and I have tried to cover these as well, and crave the reader's indulgence for emphasizing work with which I have been involved. For more details, the reviews by Proctor and Weiss (1982), Hughes and Proctor (1988), Weiss (1997) and Schüssler and Knölker (2001) may be consulted.

2. The induction equation and the Lorentz force

The interaction of a conducting fluid with a magnetic field consists of a body force (the *Lorentz force*) that the field exerts on the fluid and the effects of electromagnetic induction on the field through Faraday's law

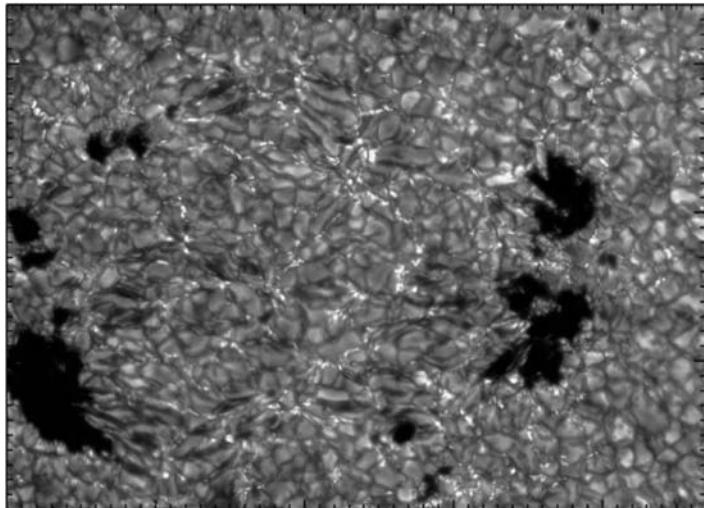


Figure 1. Filtergram of an active region. (See color insert.) Sunspots, pores and magnetic bright points can be clearly seen. (Courtesy T. E. Berger.)

in a moving medium. The latter takes the form (assuming Ohm's Law)

$$\frac{\partial \mathbf{B}}{\partial t} = -\nabla \times \mathbf{E}; \quad \sigma(\mathbf{E} + \mathbf{u} \times \mathbf{B}) = \mathbf{j}, \quad (2.1)$$

where σ is the electrical conductivity, and \mathbf{B} , \mathbf{E} are the magnetic and electric fields. If the displacement current is ignored we have the relation $\mu_0 \mathbf{j} = \nabla \times \mathbf{B}$ between the current and magnetic field, so these equations can be combined into the *induction equation*

$$\frac{\partial \mathbf{B}}{\partial t} = \nabla \times (\mathbf{u} \times \mathbf{B}) - \nabla \times (\eta \nabla \times \mathbf{B}), \quad (2.2)$$

where $\eta = (\mu_0 \sigma)^{-1}$ is the *magnetic diffusivity*. An important measure of the importance of induction compared with diffusion is the *magnetic Reynolds number* $R_m \equiv \mathcal{U}\mathcal{L}/\eta$, where \mathcal{U} , \mathcal{L} are appropriate velocity and length scales. For large R_m , appropriate to conditions in the Sun and stars, advection dominates and the field lines are effectively frozen into the fluid (Alfvén's theorem). The field ends up finely braided and twisted and reaches small scales where diffusion can operate. This scale can be estimated from the balance between diffusion $\sim \eta B''$ and induction near a local (instantaneous) stagnation point $\sim \mathcal{U}'x B'$ (the primes denoting differentiation along some normal coordinate x), giving $B'/B \propto R_m^{-1/2}$. For example, we can look at the effects of a steady flow

$\mathbf{u} = (\mathcal{U}/\mathcal{L})(-r/2, 0, z)$ in polar coordinates (r, ϕ, z) . Then the field can be supposed to be everywhere parallel to the z -axis, so $\mathbf{B} = (0, 0, B(r, t))$ with B satisfying (from (2.2))

$$\frac{\partial B}{\partial t} - \frac{\mathcal{U}}{\mathcal{L}} \left(\frac{r}{2} \frac{\partial B}{\partial r} + B \right) = \eta \left(\frac{\partial^2 B}{\partial r^2} + \frac{1}{r} \frac{\partial B}{\partial r} \right), \quad (2.3)$$

and this is solved by

$$B(r, t) = B_0 f^2(t) e^{-r^2 f^2(t)}; \quad f(t) = \exp \left(\frac{\mathcal{U}t}{2\mathcal{L}} \right) / \sqrt{\frac{4\eta\mathcal{L}}{\mathcal{U}} \exp \left(\frac{\mathcal{U}t}{\mathcal{L}} \right) + 1}. \quad (2.4)$$

This shows clearly (i) that the final scale for the field is $f^{-1}(\infty) = \sqrt{4\eta\mathcal{L}/\mathcal{U}} = 2\mathcal{L}R_m^{-1/2}$; (ii) that the total integrated flux πB_0 is conserved; and (iii) that the field strength on the axis therefore increases. This is the result of *flux stretching*, a process analogous to vortex stretching in hydrodynamics.

In persistent overturning flows any magnetic field present is advected into local regions of convergence, but is stretched out along the diverging directions which must also emanate from such regions. These loops develop high curvature and so diffusion becomes important locally, leading to reconnection of field lines. If the velocity field is laminar, this leads to closed loops of field that eventually decay due to the action of Ohmic diffusion. In a cellular flow pattern the unstable manifolds of individual fixed points can lie in the stable manifolds of other fixed points, and so the isolated flux tube theory above needs modification. This can be seen in the calculations of Galloway and Proctor (1983) for a cellular velocity field with hexagonal symmetry, in which at large times a significant proportion of flux appears near the top of the cell in a region of *divergence* (Fig. 2). The fate of disconnected flux is very different in a time-dependent chaotic flow, where there is typically exponential divergence of fluid trajectories. Firstly, coherent recirculation cannot occur and so the downdrafts remain the sites of the strongest fields. In addition, the magnetic field is continually stretched and so magnetic energy rises rapidly at first. There is also considerable folding of the field lines, however, and the outturn for the total flux (and ultimately the energy) is not certain. If there is too much folding then the action of diffusion will cause decay of energy at small scales (as happens always in two-dimensional flow fields confined to a plane), while in other cases the field can continue to grow—this is dynamo action. These issues are explored in detail in the monograph of Childress and Gilbert (1995).

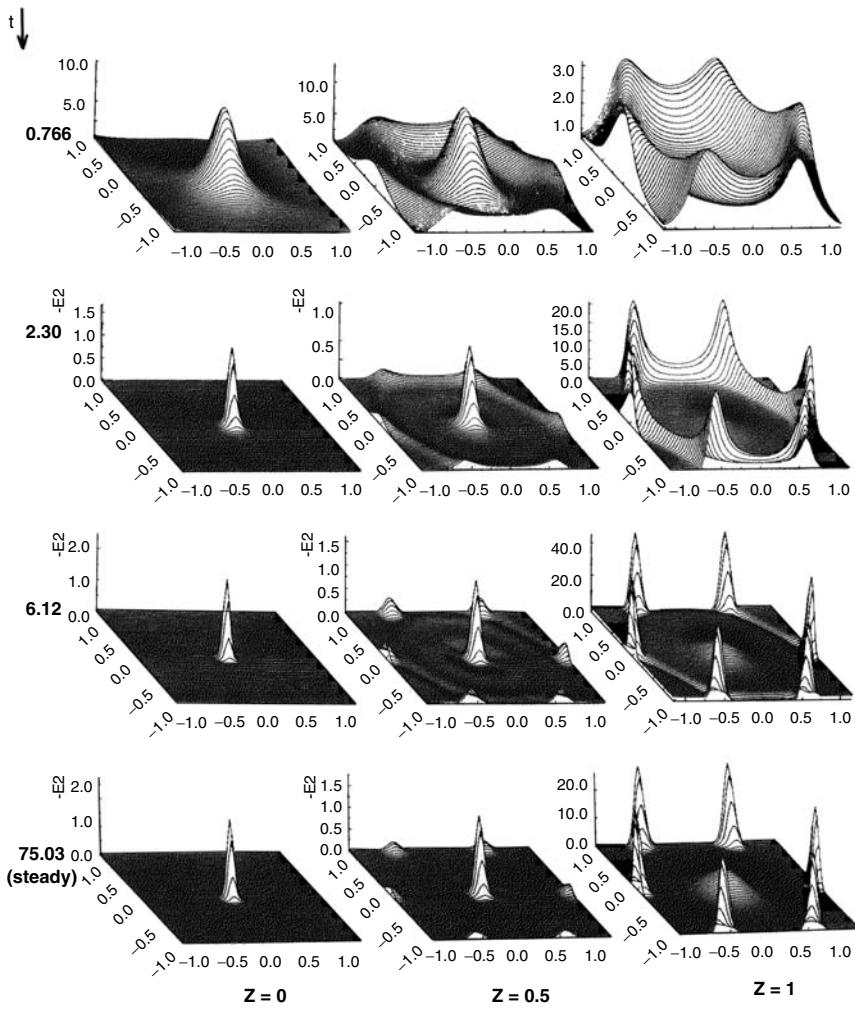


Figure 2. Flux expulsion in a hexagonal cell (from Galloway and Proctor, 1983). At small times flux is expelled to the periphery of the cell while much later a strong central tube appears.

Another important kinematic effect of motion is *flux pumping*. This was first discussed by Drobyshevski and Yuferev (1974) in a low- R_m context (see also Arter *et al.*, 1982; Galloway and Proctor, 1983). Consider for example a flow field in the form of a regular array of hexagonal cells, with isolated upwellings and connected downwellings. An initial

imposed horizontal field will be preferentially drawn downward and become concentrated near the base of the cell. This mechanism is of great current interest in connection with sunspots and the large-scale dynamo (see Section 8 below).

The dynamical effects of the magnetic field are expressed in two ways. Magnetic energy dissipated by Ohmic loss appears in the energy equation as *Ohmic heating* ($|\mathbf{j}|^2/\sigma$). While this term is essential for total energy conservation it is not usually important for driving convective flow so we do not discuss it further. There is also a direct body force on the fluid, the Lorentz force $\mathbf{j} \times \mathbf{B} = \mu_0^{-1}(\nabla \times \mathbf{B}) \times \mathbf{B} = \mu_0^{-1}[\mathbf{B} \cdot \nabla \mathbf{B} - \nabla(|\mathbf{B}|^2/2)]$. The second term in the last relation is called the *magnetic pressure*. In an incompressible fluid it can be balanced by a pressure gradient, but in a gas the relation between pressure and density at fixed temperature or entropy can lead to the evacuation of regions of strong magnetic field. The magnetic pressure gradients can take the place of pressure gradients, leading to lower pressures and hence lower densities. This phenomenon can be seen in numerical simulations as described below. By implication there is an effective upper bound to the magnetic field strength in concentrated field structures, namely $B_M \equiv \sqrt{2\mu_0 p_e}$, where p_e is the external pressure.

The remaining part $\mathbf{B} \cdot \nabla \mathbf{B}$ of the Lorentz force can be best understood by supposing the magnetic intensity $|\mathbf{B}|$ to be uniform along a field line. Then $\mathbf{B} \cdot \nabla \mathbf{B} \sim |\mathbf{B}|^2 \mathbf{n}/R$, where \mathbf{n} is the unit normal to the field line and R is the radius of curvature. Thus the field line resists bending. In consequence the magnetic field lines can be thought of as strings under tension, and can support waves (both transverse and longitudinal) when R_m is sufficiently large. If we express the magnetic field as a velocity (the *Alfvén velocity*) by writing $\mathbf{B} = (\mu_0 \rho)^{1/2} \mathbf{V}$, where ρ is the fluid density, and write $\mathbf{V} = \mathbf{V}_0 + \mathbf{v}$, $\rho = \rho_0 + \rho'$ with \mathbf{V}_0 , ρ_0 uniform, then the linearized momentum, induction and continuity equations take the form, when diffusion and viscosity are neglected:

$$\frac{\partial \mathbf{u}}{\partial t} = -c^2 \nabla \rho'/\rho_0 + \mathbf{V}_0 \cdot \nabla \mathbf{v} - \nabla(\mathbf{V}_0 \cdot \mathbf{v}), \quad (2.5)$$

$$\frac{\partial \mathbf{v}}{\partial t} = \mathbf{V}_0 \cdot \nabla \mathbf{u} - \mathbf{V}_0 \nabla \cdot \mathbf{u}, \quad \frac{\partial \rho'}{\partial t} = -\rho_0 \nabla \cdot \mathbf{u}, \quad (2.6)$$

where c is the sound speed. If we seek solutions $\propto \exp[i(\mathbf{k} \cdot \mathbf{x} - \omega t)]$ then the dispersion relation is

$$[\omega^2 - (\mathbf{V}_0 \cdot \mathbf{k})^2] [\omega^4 - \omega^2 k^2 (c^2 + |\mathbf{V}_0|^2) + c^2 k^2 (\mathbf{V}_0 \cdot \mathbf{k})^2] = 0. \quad (2.7)$$

The first bracket corresponds to *torsional waves* — the true Alfvén

waves. They travel at velocity \mathbf{V}_0 , and have no flow or perturbation field parallel to \mathbf{V}_0 ; they are therefore hard to excite by convection. The other bracket corresponds to so-called *slow* and *fast* waves according to their phase speed. There are two cases. (i) $c^2 \gg |\mathbf{V}_0|^2$; fast waves are compressional with speeds $\approx c$ while slow waves are almost transverse with velocities $\approx \mathbf{V}_0$. These latter waves become essentially the same as the torsional waves (with a different polarization) when $c \rightarrow \infty$. (ii) $c^2 \ll |\mathbf{V}_0|^2$; the slow waves are highly anisotropic, taking the form of 1D sound waves with flow parallel to \mathbf{V}_0 , while the fast waves are almost transverse. For convection in the Sun, case (i) applies except very near the surface, and the slow waves are the ones that can be destabilized by convection.

3. Boussinesq magnetoconvection in a vertical field

We start by considering the simplest problem; that of Boussinesq fluid. This limit, in which the Mach number of the flow is supposed small (corresponding to $c \rightarrow \infty$) and the fluid properties are considered constant across the layer (this can be justified provided the temperature difference is small compared with the absolute temperature), provides a basic platform to study many of the important effects of magnetic fields on convection. We consider a fluid layer lying between horizontal boundaries at $z = 0, d$, permeated by a uniform magnetic field $\mathbf{B}_0 = B_0\hat{\mathbf{z}}$. The temperature boundary conditions can vary, but generally we fix the temperature T at the top and bottom of the layer, so that $T(0) = T_0 + \Delta T$, $T(d) = T_0$. Since a uniform field exerts no Lorentz force, there is a static solution for which $T(z) = T_0 + \Delta T(1 - z/d)$. We nondimensionalize distances with d (so that the layer lies between $z = 0$ and $z = 1$), the velocity \mathbf{u} with κ/d , and write the magnetic field as $\mathbf{B} = B_0(\hat{\mathbf{z}} + \mathbf{b})$, and temperature as $T = T_0 + \Delta T(1 - z + \theta)$. Time t is scaled with d^2/κ , and fluid pressure p with $\kappa v/d^2$. We then obtain the following system (Proctor and Weiss, 1982):

$$\frac{1}{\sigma} \left(\frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \cdot \nabla \mathbf{u} \right) = -\nabla p + R\theta \hat{\mathbf{z}} + Q\xi \left(\frac{\partial \mathbf{b}}{\partial z} + \mathbf{b} \cdot \nabla \mathbf{b} \right) + \nabla^2 \mathbf{u}, \quad (3.1)$$

$$\frac{\partial \theta}{\partial t} + \mathbf{u} \cdot \nabla \theta = \mathbf{u} \cdot \hat{\mathbf{z}} + \nabla^2 \theta, \quad (3.2)$$

$$\frac{\partial \mathbf{b}}{\partial t} + \mathbf{u} \cdot \nabla \mathbf{b} = \frac{\partial \mathbf{u}}{\partial z} + \mathbf{b} \cdot \nabla \mathbf{u} - \xi \nabla \times (\nabla \times \mathbf{b}), \quad (3.3)$$

$$\nabla \cdot \mathbf{u} = \nabla \cdot \mathbf{b} = 0, \quad (3.4)$$

where $\sigma = \nu/\kappa$ is the Prandtl number, $\zeta = \eta/\kappa$ is the diffusivity ratio (inverse Roberts number), $R = g\alpha\Delta T d^3/\kappa\nu$ is the Rayleigh number (α is the coefficient of thermal expansion) and $Q = B_0^2 d^2/(\mu_0\rho_0\nu\eta)$ is the Chandrasekhar number (μ_0 is the magnetic permeability and ρ_0 the base density). The boundary conditions at $z = 0, 1$ are the standard ones, namely $\theta = u_z = 0$, $\hat{\mathbf{z}} \times \mathbf{D}\mathbf{u} = \hat{\mathbf{z}} \times \mathbf{b} = \mathbf{0}$, where \mathbf{D} denotes a derivative w.r.t. z . The mechanical condition corresponds to no tangential stress, while the magnetic field condition does not correspond to any obvious physical situation, but makes a minimal assumption about conditions outside the layer, while facilitating the algebra. Choosing other boundary conditions has little qualitative effect in general. For some comparisons see Tildesley (2003a).

3.1. Linear theory

If we linearize the above equations, we can investigate the stability of the static state to small disturbances. We can separate the velocity field \mathbf{u} into poloidal and toroidal parts, $\mathbf{u}_P = \nabla \times (\nabla \times \phi \hat{\mathbf{z}})$ and $\mathbf{u}_T = \nabla \times \psi \hat{\mathbf{z}}$, and similarly write $\mathbf{b} = \mathbf{b}_P + \mathbf{b}_T = \nabla \times (\nabla \times \xi \hat{\mathbf{z}}) + \nabla \times \chi \hat{\mathbf{z}}$. Substituting into the linearized version of (3.1)–(3.4), and taking the z components of the curl and the curlcurl of the momentum equation, we obtain two decoupled sets of equations

$$\frac{1}{\sigma} \frac{\partial}{\partial t} \nabla^2 \phi = -R\theta + \zeta Q \frac{\partial}{\partial z} \nabla^2 \xi + \nabla^4 \phi, \quad (3.5)$$

$$\frac{\partial \theta}{\partial t} = -\nabla_H^2 \phi + \nabla^2 \theta, \quad (3.6)$$

$$\frac{\partial \xi}{\partial t} = \frac{\partial \phi}{\partial z} + \zeta \nabla^2 \xi, \quad (3.7)$$

where ∇_H^2 is the horizontal laplacian, and

$$\frac{1}{\sigma} \frac{\partial \psi}{\partial t} = \zeta Q \frac{\partial \chi}{\partial z} + \nabla^2 \psi, \quad (3.8)$$

$$\frac{\partial \chi}{\partial t} = \frac{\partial \psi}{\partial z} + \zeta \nabla^2 \chi. \quad (3.9)$$

The last two equations describe *torsional oscillations*, with no vertical motion or field perturbation. There is no driving by buoyancy, and it may be verified that all solutions of this equation decay for the boundary conditions above (and indeed any sensible boundary conditions at a

plane interface). These solutions are in (fact nothing) more than damped Alfvén waves, guided by the horizontal boundaries.

We therefore focus attention on (3.5)–(3.7). We suppose all variables are proportional to $\exp(i\mathbf{k} \cdot \mathbf{x} + st)$, with $\mathbf{k} = (k_x, k_y)$ a horizontal vector; this reduces the equations to a set of o.d.e.'s with constant coefficients in z . The boundary conditions we have used then permit solutions such that $\theta, \psi \propto \sin \pi z$, $\xi \propto \cos \pi z$. This leads to the following dispersion relation:

$$\beta^2(s+\beta^2)(s+\sigma\beta^2)(s+\zeta\beta^2)+\sigma Q\beta^2\pi^2(s+\beta^2)-R\sigma k^2(s+\zeta\beta^2)=0, \quad (3.10)$$

where $k = |\mathbf{k}|$, $\beta^2 = k^2 + \pi^2$ (there are other solutions with more nodes in the vertical, but these have lower growth rates than the fundamental mode above). There is a steady-state bifurcation when $s = 0$, giving

$$R = R^{(e)} = k^{-2}(\beta^6 + Q\pi^2\beta^2). \quad (3.11)$$

Minimizing over k , we obtain the implicit relation for $R_c^{(e)} = \min_k [R^{(e)}(k)] = R^{(e)}(k_c^{(e)})$ (Dangelmayr *et al.*, 1996; see also Pollicott *et al.*, 2003):

$$\pi^2 Q = R_c^{(e)} - (R_c^{(e)2} R_0)^{1/3}, \quad k_c^{(e)2} = \pi^2 \left[\frac{3}{2} \left(\frac{R_c^{(e)}}{R_0} \right)^{1/3} - 1 \right], \quad (3.12)$$

where $R_0 = 27\pi^4/4$ is the critical value of R when $Q = 0$. It can be seen that $R_c^{(e)}$ is an increasing function of Q , and that for large Q , $R_c^{(e)} \sim \pi^2 Q$, $k_c^{(e)2} \sim (3\pi^2/2)(\pi^2 Q/R_0)^{1/3}$. This reduction in the critical wavelength of convection reflects the importance of Lorentz forces in resisting movement across the field lines.

There is also the possibility of a Hopf bifurcation if (3.10) has purely imaginary roots ($s = \pm i\omega$). This occurs when $R = R^{(o)}(k)$, where

$$R^{(o)}k^2 = A\beta^6 + B\pi^2Q\beta^2, \quad A = 1 + \frac{\zeta}{\sigma}(1 + \sigma + \zeta), \quad B = \frac{(\sigma + \zeta)\zeta}{1 + \sigma}, \quad (3.13)$$

while the frequency $\omega(k)$ satisfies

$$\omega^2 = \frac{k^2\sigma\zeta}{\beta^2(1 + \sigma + \zeta)}(R^{(e)}(k) - R^{(o)}(k)) = -\zeta^2\beta^4 + \frac{\sigma\zeta(1 - \zeta)}{1 + \sigma}\pi^2 Q. \quad (3.14)$$

This shows immediately that for any k there can be no Hopf bifurcation unless $\zeta < 1$, and that when it exists the critical Rayleigh number is always less than for the steady-state transition. If $\zeta < 1$, then for any fixed k there will first be a Hopf bifurcation as R is increased if $\pi^2 Q > \beta^4\zeta(1 + \sigma)/\sigma(1 - \zeta)$. However, the condition that there is a Hopf bifurcation for some wavenumber before any steady-state bifurcation

as R is increased is more complicated than this, since the critical wavenumbers for the two types of instability are different. Indeed, comparison of the steady state results with (3.13) shows that

$$B\pi^2 Q = R_c^{(o)} - (AR_c^{(o)2}R_0)^{1/3}, \quad k_c^{(o)2} = \pi^2 \left[\frac{3}{2} \left(\frac{R_c^{(e)}}{AR_0} \right)^{1/3} - 1 \right]. \quad (3.15)$$

The two critical values are identical when (3.12) and (3.15) hold simultaneously. This gives (Q^*, R^*) , where

$$\frac{Q^*}{R_0} = \frac{1}{\pi^2} \frac{(A^{1/3} - B)^2(A^{1/3} - 1)}{(1 - B)^3}, \quad \frac{R^*}{R_0} = \frac{(A^{1/3} - B)^3}{(1 - B)^3}. \quad (3.16)$$

It may be checked that, since $B < 1 < A$, the concomitant values of $k^{(o)2}$ and ω^2 are both positive. If $Q > Q^*$ then instability sets in first as oscillations.

Earlier work, motivated by computations in small boxes, was concerned with the interaction of oscillations and steady-state modes with the same value of k . The Takens-Bogdanov point, at which the dispersion relation (3.10) has a double-zero for a particular value of k , is in fact invisible in an extended system, but there have been many papers written considering the situation in small periodic domains, for which k is quantized, so we give some discussion of this below.

4. Nonlinear Boussinesq magnetoconvection

4.1. Pattern selection near onset

When the static state is unstable, the eventual form of convection is controlled by the nonlinear terms. Early work on the nonlinear problem was confined to small boxes, with reflection (Neumann) type or periodic boundary conditions. This ensures that there is only one cell (or at least a very small number) in the box. The bifurcating solution can then be taken to have a single wavenumber, and at least in the stress-free case we can perform perturbation theory to obtain analytical results close to onset. If we consider periodic boundary conditions in x and y , we can choose (*inter alia*) solutions whose Fourier components lie either on a square or hexagonal lattice. This problem has been thoroughly addressed by Clune and Knobloch (1994). We concentrate on the square lattice as it is simpler, and so choose a square periodic domain in x, y with period corresponding to the optimal wavenumber for instability given in the previous section. For $R = R^{(e)} + R_2\epsilon^2$, i.e., close to a

steady-state bifurcation we make the expansion

$$\begin{aligned}\mathbf{u} &= \epsilon \sum_{j=1}^2 (\hat{\mathbf{u}}_j A_j(T) e^{i\chi_j} + c.c.) \sin \pi z + \epsilon^2 \mathbf{u}_2 + O(\epsilon^3), \\ \theta &= \epsilon \sum_{j=1}^2 (\hat{\theta}_j A_j(T) e^{i\chi_j} + c.c.) \sin \pi z + \epsilon^2 \theta_2 + O(\epsilon^3), \\ \mathbf{b} &= \epsilon \sum_{j=1}^2 (\hat{\mathbf{b}}_j A_j(T) e^{i\chi_j} + c.c.) \cos \pi z + \epsilon^2 \mathbf{b}_2 + O(\epsilon^3),\end{aligned}\quad (4.1)$$

where the hatted quantities are complex constants, $\chi_1 = k_c x$, $\chi_2 = k_c y$. Use of standard perturbation theory and symmetry arguments then yield a coupled pair of evolution equations, namely

$$\begin{aligned}\frac{dA_1}{dt} &= \mu A_1 - (\nu_1 |A_1|^2 + \nu_2 |A_2|^2) A_1, \\ \frac{dA_2}{dt} &= \mu A_2 - (\nu_1 |A_2|^2 + \nu_2 |A_1|^2) A_2,\end{aligned}\quad (4.2)$$

where $\mu \propto R_2$. The real parameters ν_1, ν_2 depend on Q, ζ (ν_2 depends also on σ). There are two types of nonlinear steady state: *rolls* (stripes) in which e.g., $A_1 \neq 0, A_2 = 0$; and *squares* in which $|A_1| = |A_2|$. Clune and Knobloch (1994) show that squares are never stable, while rolls are supercritical and so stable when $\zeta^2 k_c^4 (\pi^2 + k_c^2) > \pi^2 (2k_c^2 - \pi^2)(\pi^2 - k_c^2)$. Broadly speaking, rolls are supercritical for sufficiently large Q and ζ .

When $Q > Q^*$ the first bifurcation as R increases is to oscillations. Now we set $R = R^{(o)} + \epsilon^2 R_2$, and the expansion (4.1) then has to be replaced by

$$\dots, \quad \theta = \epsilon \sum_{j=1}^4 (\hat{\theta}_j A_j(T) e^{i\chi_j} + c.c.) \sin \pi z + \dots, \text{ etc.}, \quad (4.3)$$

where now $\chi_{1,2} = i(\omega_c t \pm k_c x)$, $\chi_{3,4} = i(\omega_c t \pm k_c y)$. Again by standard techniques we have the set of four coupled equations

$$\begin{aligned}\frac{dA_1}{dt} &= (\mu + i\delta) A_1 \\ &\quad - (\nu_1 |A_1|^2 + \nu_2 |A_2|^2 + \nu_3 (|A_3|^2 + |A_4|^2)) A_1 + \nu_4 \bar{A}_2 A_3 A_4,\end{aligned}\quad (4.4)$$

$$\begin{aligned}\frac{dA_3}{dt} &= (\mu + i\delta) A_3 \\ &\quad - (\nu_1 |A_3|^2 + \nu_2 |A_4|^2 + \nu_3 (|A_1|^2 + |A_2|^2)) A_3 + \nu_4 \bar{A}_4 A_1 A_2,\end{aligned}\quad (4.5)$$

together with two other equations obtained from these by interchanging respectively A_1, A_2 and A_3, A_4 . The coefficients v_i are now complex. Now several types of stable temporally periodic solution are possible near onset, depending on the parameters. They are: *Travelling Rolls* (TR), where, e.g., $|A_1| \neq 0, A_i = 0, i > 1$; *Standing Rolls* (SR), $|A_1| = |A_2|, A_3 = A_4 = 0$; *Standing Squares* (SS), $A_1 = A_2 = A_3 = A_4$; and *Alternating Rolls* (AR), $A_1 = A_2 = iA_3 = iA_4$. The actual realized pattern now depends strongly on σ . In addition there are areas where there is a stable solution that is not periodic but quasiperiodic. As a broad statement, though, for sufficiently large Q the AR are always stable while TR can be stable also for large σ . The various regions of parameter space are shown in Fig. 3. A similar but more bewildering zoo of possibilities is found on the hexagonal lattice, yielding once more a stable three-dimensional periodic solution analogous to AR at large Q . Dawes (2000) has clarified ambiguities in the diagram near $Q = Q^*$.

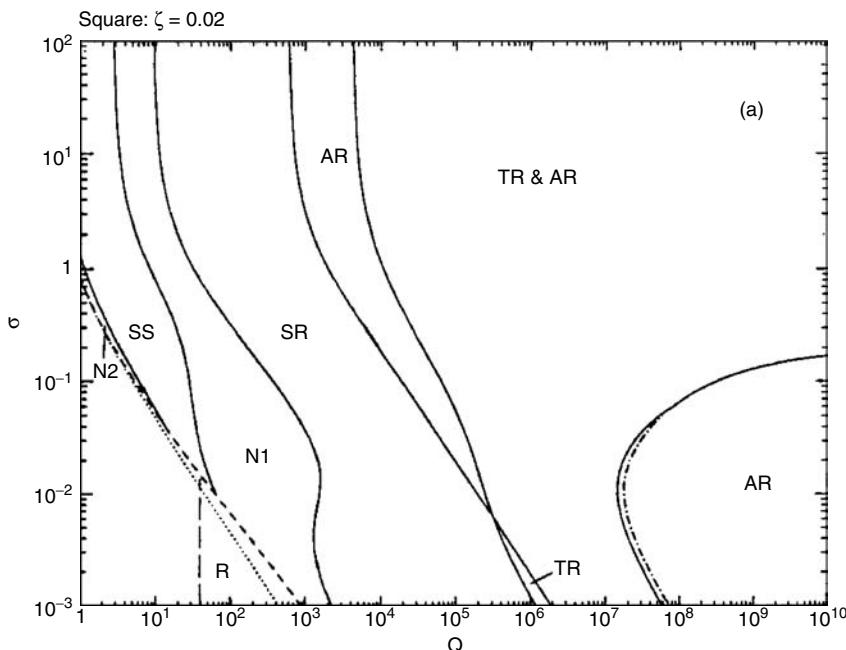


Figure 3. Bifurcation diagram on the square lattice for Boussinesq magnetoconvection for $\zeta = 0.02$ (from Clune and Knobloch, 1994). The locus of Q^* is the dashed line. In regions marked $N1, N2$ none of the primary branches is stable.

4.2. Fully nonlinear convection

A systematic survey of the two-dimensional Boussinesq problem with reflective boundary conditions has been given by Weiss (1981a,b). These conditions preclude the appearance of traveling waves, and so only standing waves and steady motion are possible near onset. The equations are solved numerically using finite differences. Two questions of interest concern the interaction between the steady and oscillatory solution branches, and the form of convection when R_m is large and flux expulsion occurs. When conditions are such that oscillations are possible, oscillations always set in first as R is increased. These oscillations, which are inefficient in transporting heat, typically end on the steady branch. If the steady branch is subcritical, then the oscillations end at a homoclinic bifurcation, where the period of oscillation tends to infinity. When the branch is supercritical the stable oscillations either end at a saddle-node (turning-point) bifurcation, or else become vacillatory (small oscillations in velocity without sign change) and then join the steady branch at a Hopf bifurcation. A fuller study of the qualitative behavior has been undertaken by Knobloch *et al.* (1981) using a truncated model in which only a few Fourier components are retained. This model reproduces conditions at onset correctly, and gives a good qualitative guide elsewhere. When Q is very close to the codimension-2 point where $R^{(e)} = R^{(o)}$ the equations can be reduced near onset to a second-order set of ODEs, as shown by Knobloch and Proctor (1981). In that case the oscillatory solutions can be found in terms of Jacobian elliptic functions, and entire solution branches found by quadrature. The results show the same qualitative behavior as for the general case. Various possibilities are shown in Fig. 4. The effects of large R_m on the steady branch are of two kinds. When flux is expelled from the interiors of the cells, its dynamical effect diminishes. The field is confined to a boundary layer of thickness $O(R_m^{-1/2})$, and the current in this layer is of order R_m . Thus the dimensionless dissipation $\langle |\mathbf{j}|^2 \rangle$ is of order $R_m^{3/2}$. Manipulation of (3.1)–(3.4) for a (statistically) steady state gives the power integral $\langle \theta \mathbf{u} \cdot \hat{\mathbf{z}} \rangle = \langle |\nabla \mathbf{u}|^2 \rangle + \zeta^2 Q \langle |\mathbf{j}|^2 \rangle$. If we suppose that σ is of order unity, and the amplitude of the velocity (in fact the Reynolds number) is $\epsilon \ll 1$, then $R_m = \epsilon/\zeta$, while $R = R_0 + \Delta R$ with $\Delta R \ll 1$. Expanding the power integral we find that, when $\zeta \ll \epsilon \ll 1$,

$$\Delta R = A\epsilon^2 + BQ(\zeta/\epsilon)^2 Q \langle |\mathbf{j}|^2 \rangle = A\epsilon^2 + C(\zeta/\epsilon)^{1/2} Q, \quad (4.6)$$

where A, B, C are constants of order unity. This ansatz is valid if $\zeta^{1/2} Q \sim \epsilon^{5/2}$ (for further details see Busse, 1975; Proctor and Weiss, 1982). Clearly ΔR has a minimum of order $(\zeta^{1/2} Q)^{4/5}$, and so unless

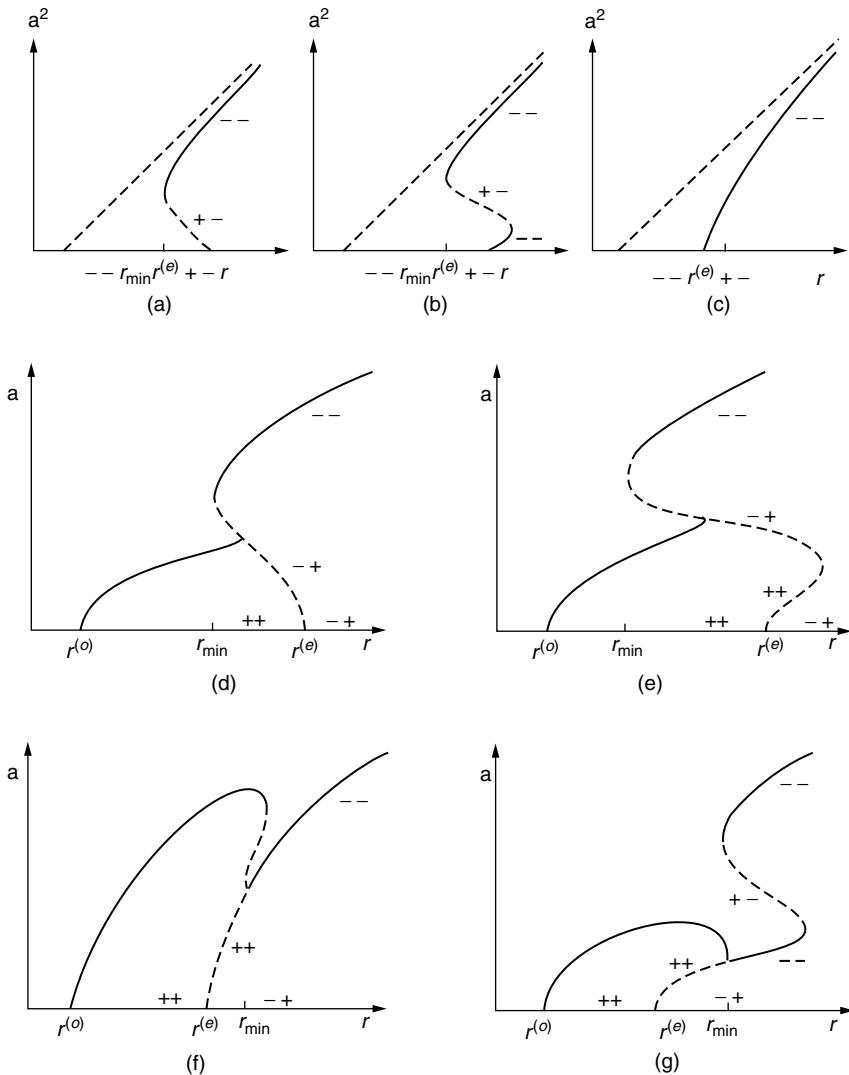


Figure 4. Amplitude-Rayleigh number diagrams for Boussinesq magnetoconvection (from Proctor, 1992). The Rayleigh number is denoted by r . (a)–(c) show situations in which there is no Hopf bifurcation. (d)–(f) show possibilities when oscillations become unstable first, with (d) and (f) giving the canonical pictures when $R^{(e)}$ is very close to $R^{(o)}$ (cf., Knobloch and Proctor, 1981).

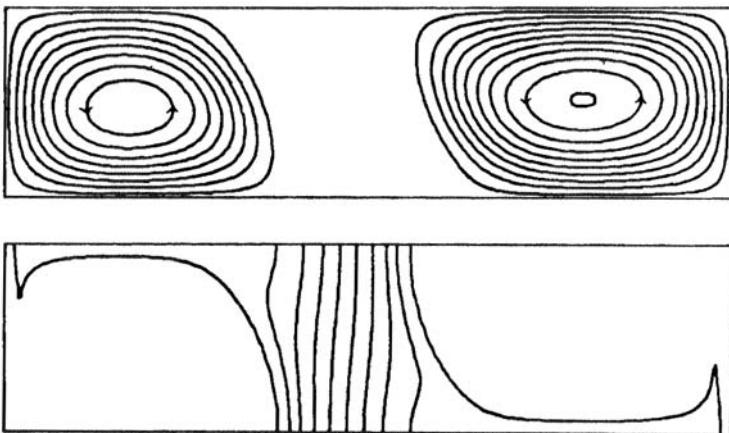


Figure 5. Convection with flux expulsion (from Proctor and Weiss, 1982). The region of strong field is nearly stagnant, and the dynamical coupling so low that the two gyres have the same sense.

$Q < O(\zeta^2)$ convection can occur for $R < R^{(e)}$. This explains the “double turning-point” structure of some of the figures. Depending on the parameters the onset of oscillations at $R^{(o)}$ may lie above or below the minimum value of R for steady convection.

Another effect of flux expulsion, more apparent at larger values of Q , is that the flux tubes between the convection cells become dynamically significant, and the fluid velocity within becomes much smaller than in the gyres. In extreme cases the interior of the sheets is essentially stagnant, and neighboring gyres are dynamically decoupled. This leads to a preference for wider cells than linear theory would suggest. Indeed for small ζ the largest value of Q for which convection can occur at fixed R is maximal as a function of the aspect ratio when the latter is about 2. Furthermore, the original symmetries of the Boussinesq problem may be broken at finite amplitude, with every other flux sheet having different amounts of flux (Fig. 5). This can happen both for steady and oscillatory motion. Related experiments and theoretical investigations have been carried out in an axisymmetric geometry by Galloway (1976, 1978), Galloway *et al.* (1978), Galloway and Moore (1979), and Proctor and Galloway (1979). In this case the field is constrained to be vertical at the outer boundary, and in the steady state it can be shown that the mean vertical field at this boundary is unaffected by the motion. For large R_m , in consequence, almost all the flux is confined to a central tube, which exerts a powerful but local dynamical

effect. The dynamical effect of a thin, but dynamically active tube can be calculated accurately, with the result that (Proctor and Galloway, 1979) (4.6) is replaced by $\Delta R = A\epsilon^2 + C[\ln(\epsilon/\zeta)]^{-1}$ (independent of Q). Thus subcritical convection can occur in this case also for small ζ . The numerical results at larger amplitudes show that the peak field (in the central tube) has a maximum as a function of Q in the flux tube regime, since the amplification factor (R_m in the axisymmetric case) is reduced by local retardation due to the Lorentz force.

Recently there has been renewed interest in three-dimensional nonlinear Boussinesq convection as a model for dynamo action in extended layers (see, e.g., Cattaneo, 1999). Cattaneo and his collaborators have investigated highly nonlinear convection in a layer of large aspect ratio. Most of this work relates to dynamo action, which is beyond the scope of this review, but Cattaneo *et al.* (2002) have investigated the transition from a dynamo (when there is no imposed flux) to what should be termed magnetoconvection, when there is a large mean flux and the scale of the convection is significantly reduced (Fig. 6; [See color insert](#) following page 234). The motion is columnar in this case and there is greater coherence in depth than when the field is absent. In three-dimensional convection it is impossible to avoid dynamo action in general, and so even when there is imposed flux the relationship of imposed field to statistically steady state may not replicate that found for two-dimensional flows.

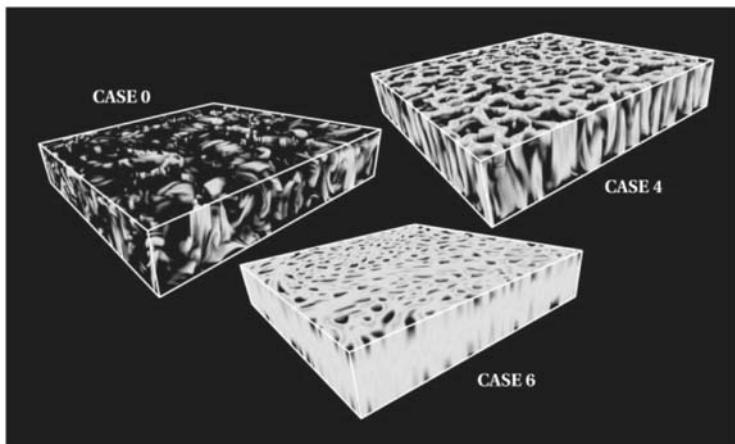


Figure 6. Volume rendering for magnetic fields in Boussinesq magnetoconvection (from Cattaneo *et al.*, 2002). ([See color insert.](#)) Case 0 is a dynamo, with no imposed field, while cases 4 and 6 have moderate and large imposed fields, respectively.

5. Strong magnetic fields

It is noted above that in the limit of large Q the critical wavelength for the onset of convection is very small – of order $\epsilon = Q^{-1/6}$. Because of the disparity of scales it is possible by making an expansion in powers of ϵ to produce a set of reduced equations that give information about the nonlinear behavior. The first attempt at this reduction was made by Proctor (1986), who considered the onset of steady-state convection and of oscillations in the Boussinesq case. Close to onset there is a regime in which the timescale for the secular change of the cellular convection and the vertical diffusion time for the horizontally averaged temperature are comparable. The principal nonlinearity is due to the alteration of the horizontally averaged temperature profile. The appropriate scalings are

$$\left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y} \right) = \epsilon^{-1} \nabla_H, \quad \mathbf{u} = (\epsilon \tilde{\mathbf{u}}_H, \tilde{u}_z), \quad p = \epsilon^{-2} \tilde{p}; \quad (5.1)$$

$$\theta = \epsilon^2 (\tilde{\theta}(\tilde{x}, \tilde{y}, z) + \Theta(z)), \quad \mathbf{b} = (\epsilon^3 \tilde{\mathbf{b}}_H, \epsilon^2 \tilde{b}_z), \quad (5.2)$$

and the tildes are now dropped. Note that the horizontally averaged temperature Θ is of the same order as the fluctuating temperature θ . We write $R = r\epsilon^{-6}$. Concentrating first on the steady-state case, and substituting into the governing equations it is found that the leading order balance in the momentum equation is between the buoyancy force and the vertical Lorentz force. The linear stability problem reduces to second order in z , and it is found that $r_0 = \pi^2 + \epsilon^2 r_2 + \dots$, independently of the horizontal planform of the motion, which is left undetermined. We write, e.g., $\theta = -\sin \pi z \nabla_H^2 F(x, y, t)$; then at the next order in the expansion we obtain a pair of coupled evolution equations for the interaction of the convection and the mean temperature:

$$(1 - \zeta^{-1} \pi^2) \nabla_H^2 \left(\frac{\partial F}{\partial t} \right) = r_0 H(t) \nabla_H^4 F - r_0 \pi^2 \nabla_H^2 F + \nabla_H^8 F - r_2 \nabla_H^4 F, \quad (5.3)$$

$$\frac{\partial \Theta}{\partial t} = \pi \sin 2\pi z \overline{\nabla_H^4 F \nabla_H^2 F} + \frac{\partial^2 \Theta}{\partial z^2}, \quad (5.4)$$

where $H(t) = \int_0^1 \sin^2 \pi z (\partial \Theta / \partial z) dz$. This coupled system selects the critical wavenumber of the convection, but still contains no information concerning the preferred planform, which can only be determined at yet higher order. Nor is the time behavior very exciting: above the critical value of r_2 the stable states are all time independent, so that the extra dynamical possibilities provided by the mean temperature are not

realized. The same methods can be used on the oscillatory bifurcation, though there is the complication that the frequencies near onset are on a different timescale from the secular growth of disturbances, so that the expansion method is more complicated.

The situation for the steady-state transition is quite different when non-Boussinesq effects are present. Halford and Proctor (2002) modeled this situation by supposing the magnetic diffusivity to be a function of z , though the results do not depend on the precise choice of the asymmetry. The equations are then (3.1)–(3.4) except that the last term in (3.3) becomes $\zeta_0 \nabla \times (h(z) \nabla \times \mathbf{b})$ where $\zeta_0 = \zeta(z = 0)$. Repeating the earlier analysis leads to equations essentially identical to (5.3), (5.4) (with slightly different coefficients), except that (5.3) has extra quadratic terms in F . These new terms change the character of the solutions entirely. There is now strong planform selection in favor of hexagons, as is to be expected in the non-symmetric case. Furthermore, the branch of steady hexagons can become unstable to oscillations, leading to large amplitude pulsations. These rely for their existence not only on the quadratic terms, but also on the extra time derivative provided by the mean temperature term. In addition, the uniform hexagonal pattern (whether steady or oscillatory) can become unstable due to modulational Turing or Benjamin-Feir instabilities, depending on the parameters. The Turing instability leads (when the basic state is linearly stable) to isolated regions of convection surrounded by nearly static fluid.

The above analysis is only valid when $R/R_c = 1 + O(\epsilon^2)$, and so although the temperature is significantly modified we are really in the weakly nonlinear regime. Matthews (1999) has noted that if the planform is not required, the reduction can be effective when $R/R_c = O(1)$. He repeats the above analysis (in the Boussinesq case) but with \mathbf{u} , \mathbf{b} , p , θ multiplied by ϵ^{-1} and Θ by ϵ^{-2} . If steady solutions are sought then at leading order the system supports solutions with a single horizontal wavenumber, which can be removed by scaling. The mean temperature balance in the steady-state case gives $\bar{u}_z \bar{\theta} = d\bar{\Theta}/dz + N - 1$, where N is the Nusselt number and the bar denotes horizontal average. The other equations take the form

$$u_z \left(\frac{d\Theta}{dz} - 1 \right) = \nabla_H^2 \theta, \quad R \nabla_H^2 \theta = -\zeta Q \frac{\partial}{\partial z} \nabla_H^2 b_z, \quad 0 = \frac{\partial u_z}{\partial z} + \zeta \nabla_H^2 b_z. \quad (5.5)$$

Redefining $W = u_z k^{-1}$, where k is the horizontal wavenumber, we can reduce the system to a nonlinear equation for $W(z)$, namely

$$\frac{d^2 W}{dz^2} + \frac{RN}{Q} \frac{W}{1 + W^2} = 0. \quad (5.6)$$

This gives a solution for each value of $RN > \pi^2 Q$, and N is then determined by the relation (which derives from the boundary condition $\Theta(1) = \Theta(0) = 0$) $\int_0^1 (1 + W^2)^{-1} dz = N^{-1}$. The solution of (5.6) can be found numerically and provides a connection with the small-amplitude theory of Proctor (1986) and large amplitude results showing the confinement of temperature gradients and magnetic field to thin boundary layers of thickness $\sim N^{-1}$. A similar set of equations can be obtained for nonlinear periodic convection when $\zeta < 1$. In this case the scaling is more complicated, with the high-frequency oscillatory parts of the variables having larger amplitudes than the (temporally) constant parts. Writing the dominant fluctuating vertical velocity as $W e^{i\omega t} \omega \epsilon^3 / \sqrt{2k} + c.c.$, we get a nonlinear equation similar to (5.6), namely

$$\frac{d^2 W}{dz^2} + \frac{RN}{\zeta Q} \frac{W}{1 + |W|^2} + \frac{\omega^2}{\sigma \zeta Q} W = 0 \quad (5.7)$$

with the integral conditions, analogous to the steady case, $\int_0^1 (1 + |W|^2)^{-1} dz = N^{-1}$, $\omega^2 (1 + \sigma) \int_0^1 |W|^2 dz = \sigma \zeta Q (1 - \zeta) \int_0^1 |W'|^2 dz$. It can be shown from these equations that the frequency of oscillation increases with R , so that the oscillatory branch does not end at the steady branch (with $\omega = 0$) within this scaling. There is a singular limit when $\zeta = O(\epsilon^2)$, in which both branches could appear together, but this has not been investigated.

Another approach has been taken by Julien *et al.* (1999, 2000). In the first of these papers the vertical field problem is considered, but the horizontal wavenumbers are considered to have the larger value $Q^{1/4}$. Making this scaling allows the frequency to be addressed on the same basis as other variables, though the critical wavenumber lies outside the range. The outturn is an equation similar to (5.7), namely

$$\frac{d^2 W}{dz^2} - \frac{k^4}{\zeta \sigma} (i\Omega + \sigma)(i\Omega + \zeta) W + \frac{RN}{\zeta} \frac{(i\Omega + \zeta)(-i\Omega + 1)}{\Omega^2 + 1 + |W|^2} W = 0, \quad (5.8)$$

where now $\int_0^1 (\Omega^2 + 1)(\Omega^2 + 1 + |W|^2)^{-1} dz = N^{-1}$, and Ω is proportional to the frequency. It can be shown that both the Hopf bifurcation (if any) and the steady-state bifurcation are supercritical, and that the frequency increases with amplitude. In fact we have the exact result

$$\Omega^2 = -\zeta^2 + \frac{\zeta \sigma (1 - \zeta)}{k^4 (1 + \sigma)} \frac{\int_0^1 |W'|^2 dz}{\int_0^1 |W|^2 dz} \quad (5.9)$$

and since $\int_0^1 |W'|^2 dz \geq \pi^2 \int_0^1 |W|^2 dz$, with equality only if $W(z) \propto \sin \pi z$, we can see that Ω is always greater than its linear value. This has

the interesting consequence that oscillations may be possible at finite amplitude when k is too large to permit an initial Hopf bifurcation. The situation is somewhat similar to that described in Rucklidge *et al.* (1993), though in that paper there is a Hopf bifurcation from the steady branch, rather than a symmetry-breaking steady-state one.

The effects of a strong imposed *oblique* field have been considered in a further paper by Julien *et al.* (2000). The same asymptotics is used, but now there is anisotropy for tilted fields. Both the Boussinesq limit and the effects of space-dependent ζ are considered. The large Q limit misses the most important effect of the interaction between compressibility and field obliquity; that due to insufficient symmetry there are no steady solutions and the initial bifurcation is always to a drifting mode (see Section 8, below). However, an interesting new effect is identified; a transition as the tilt increases between vigorous convection similar to that for a vertical field, and less vigorous motion, with the magnetic field expelled to (relatively thick) horizontal boundary layers. This transition is hysteretic, and it is suggested that for a sunspot, where field tilt increases from the center, the abrupt transition from umbra to penumbra can be related to this change. It is again assumed that the solutions are steady or are periodic sinusoidal oscillations; the results of Halford and Proctor (2002) suggest that these solutions might be unstable in the non-Boussinesq case.

6. Shearing instabilities and streaming flows

The above results, with reflective boundary conditions, have removed the possibility of mean horizontal flows. It turns out that when the vertical boundaries are replaced by periodic ones, the usual cellular flows can become unstable to horizontal shears. This instability has been investigated in the Boussinesq case by Matthews *et al.* (1993), Rucklidge and Matthews (1996) and Proctor *et al.* (1994) (in the compressible case). The instability manifests itself either as a pitchfork bifurcation leading to steady horizontal shear, with tilted convection cells (STC), with every other one enhanced in vigor due to the definite sign of the vorticity induced by the shear. Alternatively (and this appears to be the norm for larger Q and sufficiently small ζ) a Hopf bifurcation occurs leading to periodically oscillating shear, known as pulsating waves (PW) ([Fig. 7](#)). In this case the shear can be thought of as the velocity of an Alfvén wave, which can gain energy from the cellular flow to balance its natural dissipative decay. In the compressible case the shear breaks enough symmetries that the resulting tilted cells travel, but the

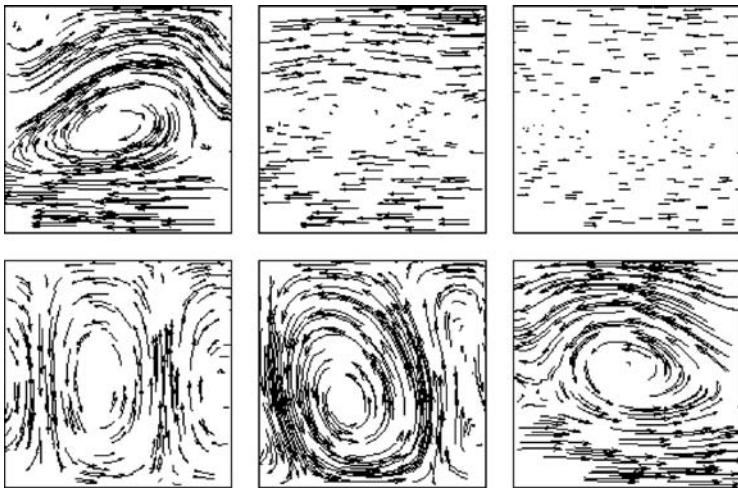


Figure 7. Velocity field for a pulsating wave (from Matthews *et al.*, 1993). The two-cell convective pattern is dominated by the shear at each end of the sequence. At intermediate times only the shear remains.

bifurcation sequences are similar otherwise. When there is an initial steady-state bifurcation the pattern can become time dependent via a secondary Hopf bifurcation, leading to vacillations. PW can then reappear via a complicated sequence of further transitions, which have only been investigated fully in a truncated model. In the Boussinesq case the principal interaction is between the first three permitted vertical wavenumbers, so that we write (cf. (3.5)–(3.7))

$$\mathbf{u} = \nabla \times (\psi_{11} \sin kx \sin \pi z + \psi_{01} \sin \pi z + \psi_{12} \cos kx \sin 2\pi z) \hat{\mathbf{y}}, \quad (6.1)$$

$$\theta = \theta_{11} \cos kx \sin \pi z + \theta_{02} \sin 2\pi z + \theta_{12} \sin kx \sin 2\pi z, \quad (6.2)$$

$$\begin{aligned} \mathbf{b} = \nabla \times & (A_{11} \sin kx \cos \pi z + A_{01} \cos \pi z + A_{12} \cos kx \cos 2\pi z \\ & + A_{10} \cos kx + A_{20} \sin 2kx) \hat{\mathbf{y}}. \end{aligned} \quad (6.3)$$

Substituting into the governing equations and suppressing all other modes yields an eleventh order system of ODEs, which can be thought of as the extension to shear of the model of Knobloch *et al.* (1981) referred to above. Rucklidge and Matthews (1996) undertake a very careful study of the region of parameter space near the global bifurcation leading from vacillations to PW (Fig. 8). There are in fact four equivalent steady tilted solutions (two directions of tilt, two signs of the velocity), and so this can happen in a large number of different ways. Only

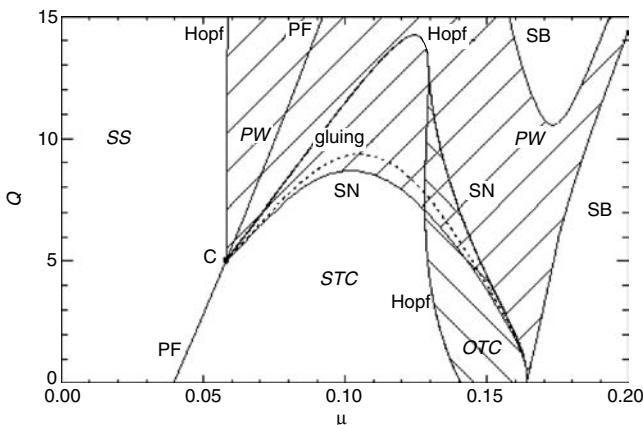


Figure 8. Unfolding diagram for pulsating waves in the reduced model of Matthews *et al.* (1993); μ is the bifurcation parameter corresponding to R , and $\sigma = 0.5$, $\zeta = 0.2$. The hatched region denotes pulsating waves (PW). SS denotes steady state, and STC steady tilted cells (with steady shear).

close study of reductions of the equation to 1D maps is sufficient to yield the full picture. For three-dimensional convection the situation is still more complex. Matthews *et al.* (1996) have considered the Boussinesq problem confined to a square periodic domain in the horizontal. A similar reduction to the above yields a 23-mode system which can be reduced to seven modes by a further approximation. This has then been analyzed in a similar way to the two-dimensional problem. If the initial state is two-dimensional rolls, the effect of a shearing instability is to greatly reduce the amplitude of those rolls, thus allowing orthogonal rolls to appear. Then these rolls in turn become unstable. The sequence of roll patterns is again extremely complex. A more formal treatment of the problem as an example of “cycling chaos” has been given by Ashwin and Rucklidge (1998).

Although the shearing instability yields fascinating dynamical behavior, its interest is somewhat diminished by the fact that the phenomenon only seems to occur in narrow periodic boxes. Simulations for the same parameters in wider domains usually result in the breakup of the convective pattern. However, it has been shown by Julien *et al.* (2004) that in the case of large Q , for which convection onsets as narrow rolls, and small σ so that $\sigma^{3/2}Q$ is of order unity, an instability to shear can be found not far from onset, for rolls which have optimum (large) wavenumber $\sim Q^{1/6}$.

7. Convection in extended regions

Almost all moderately nonlinear simulations of magnetoconvection have been undertaken in relatively narrow domains with periodic or reflective (Neumann type) boundary conditions. If the periodicity of the domain is much larger than the critical wavelength of the motion, then periodic solutions of the original wavelength are still possible but they may be unstable. Possible instabilities are of two kinds; there is a “period doubling” transition in which alternate pairs of cells are of different amplitude; this is associated with the expulsion of flux into the interstices of the cells, and the concomitant reduction in the restoring Lorentz force. The other is a long-wavelength instability, associated with the fact that the total magnetic flux in the layer is a conserved quantity under the usual boundary conditions. This latter mechanism can occur essentially at onset; when the Boussinesq symmetry is broken, it is found under very general circumstances.

7.1. Symmetry breaking instabilities and convectons

The results reported above for two-dimensional convection in a box of width the order of the wavelength show that the symmetry between alternate flux sheets can be broken when the convection is vigorous enough. A simplified model of this process was given by Proctor (1983). Suppose that there is a two-dimensional cellular flow in a periodic box of width $2L$ and consider half the domain $0 < x < L$. There is a single cell in this domain, and flux sheets near $x = 0, x = L$. In the absence of field the cell would have characteristic velocity U , while in the neighborhood of $x = 0, x = L$ the local velocities are $V_1 = U - \Delta V_1$, $V_2 = U - \Delta V_2$, respectively. The induced velocities ΔV_i can be related to the vorticity induced by the flux sheets near $x = 0, L$. Following Galloway *et al.* (1978), we can write the velocities in terms of the peak fields B_i and the boundary layer thicknesses ϵ_i : $\Delta V_1 \propto B_1^2 \epsilon_1 + \alpha B_2^2 \epsilon_2$, and similarly for ΔV_2 ; here $0 < \alpha < 1$ is a coupling constant (~ 0.5 in the original theory of Galloway *et al.* (1978)). The ϵ_i are proportional to the local magnetic Reynolds number so that $\epsilon_i \propto V_i^{-1/2}$, and flux conservation and symmetry constraints lead to $B_1 = B_2 = B_0 L / (\epsilon_1 + \epsilon_2)$, where $B_0 L$ is the total vertical magnetic flux. After rescaling we arrive at the coupled equations for $r_{1,2} \propto \epsilon_{1,2}$:

$$r_1^2 = 1 - 4\Lambda \frac{(r_1 + \alpha r_2)}{(r_1 + r_2)^2}, \quad r_2^2 = 1 - 4\Lambda \frac{(r_2 + \alpha r_1)}{(r_2 + r_1)^2}, \quad (7.1)$$

where $\Lambda \propto B_0^2 U^{-1/2}$. The symmetric solution to (7.1), which holds for all values of Λ , bifurcates to give asymmetric solutions if Λ is sufficiently large. The model can be extended to a long line of cells, and complicated solutions can be found (Houghton, 2004, in preparation). However the model cannot easily be extended to three space dimensions, e.g., to a lattice of hexagonal cells, since there is no simple way to determine the peak fields in that case. Furthermore, the possibility of total suppression of convection by field that is expelled from a region of vigorous convection is outside the scope of the model.

Blanchflower *et al.* (1998) (see also Blanchflower, 1999a) have conducted an extensive study of two-dimensional compressible magnetohydroconvection in wide boxes, with varying temperature and magnetic field boundary conditions. The atmosphere employed is the standard one used for earlier two- and three-dimensional studies by Hurlbut and Toomre (1988) and Weiss *et al.* (1990, 1996). By varying Q at fixed R the transition from narrow, ordered cellular motion to time-dependent convection can be studied. It is found that for sufficiently weak fields the convection is dominated by very wide cells, which act to expel the magnetic flux. Furthermore, for some values of the parameters there is the possibility that the convection is confined to a small region of the domain, with the remainder being occupied by low-velocity fluid with a strong, almost uniform magnetic field. This phenomenon is known as *flux separation*; see Weiss *et al.* (2002) and below. We can identify two mechanisms here: the tendency of cells in compressible convection to attain very large horizontal size, and the ability of vigorous convection to expel magnetic field and so become more vigorous still. The propensity of compressible convection with constant dynamic viscosity to have large horizontal scales is not fully understood, but the phenomenon of flux separation can be investigated even within the Boussinesq theory. This problem is treated in Blanchflower (1999a) and by Blanchflower (1999b) and Blanchflower and Weiss (2002). Numerical simulations were performed on the full Boussinesq equations in two dimensions in a long periodic box, with aspect ratio ≤ 6 . The parameter choices were $\sigma = 1$, $\zeta = 0.1$, $R = 20000$. In this regime the initial bifurcation to steady convection is subcritical for sufficiently large Q . It is found that as Q is increased the cellular pattern obtaining for smaller Q becomes unstable, and isolated states of convection (*convectons*) appear. The horizontal extent of motion decreases as Q is increased, until ultimately there is only one half-cell (with just one gyre) in the box; away from these localized cells there is almost static fluid, with a magnetic field that is almost uniform, and of sufficient strength to stabilize any local instabilities.

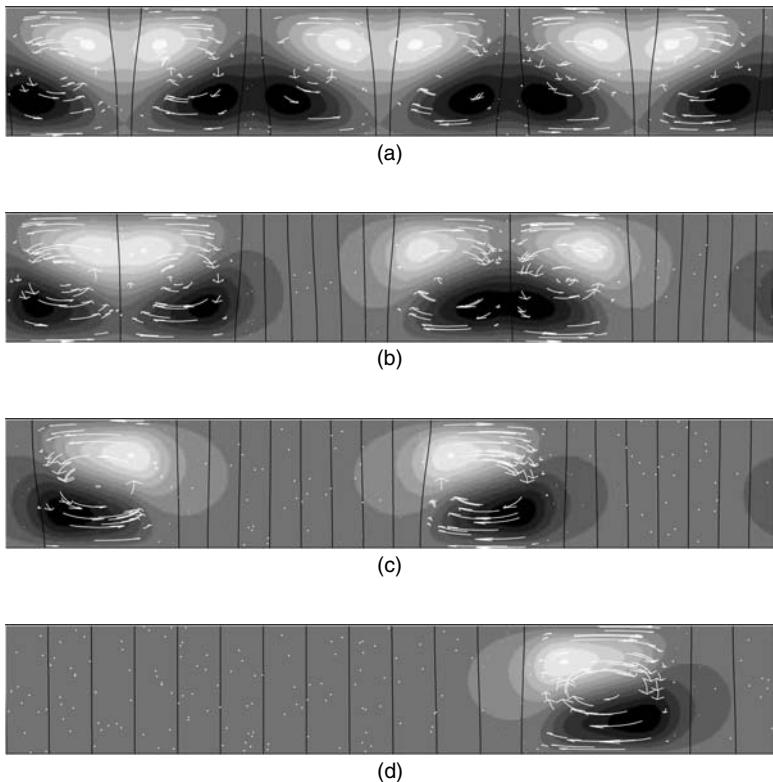


Figure 9. Formation of convection cells in a truncated model of Boussinesq magnetoconvection (from Blanchflower, 1999b). Shown are steady states of an initial-value simulation. The magnetic field strength increases downward.

In order to cover a wider range of parameters Blanchflower developed a reduced model of Boussinesq convection, in which only a few vertical modes in the velocity and magnetic fields were retained, with full horizontal resolution. Then very wide boxes with aspect ratio 16 and larger can be considered. Although the model cannot cope with horizontal boundary layers, at the relatively low Nusselt numbers employed it gives results similar to those of the fully resolved model (Fig. 9). The way in which the total number of gyres decreases as the magnetic field increases is affected by the accompanying changes in symmetry. In particular, the final change from one full cell (two gyres) to a single gyre breaks a reflectional symmetry and has to happen via a pitchfork bifurcation. Unresolved questions include a theoretical estimate for the

range of parameters in which different types of isolated solutions can exist, and so of the maximum value of Q for which any kind of convection is possible. The reduced model described above has also been investigated for fields and flows that vary in both horizontal dimensions (Blanchflower and Weiss, 2002). Isolated solutions can still be found though these are oscillatory rather than the steady motion found in the two-dimensional case. Various types of pulsation with different symmetries have been detected, though more detailed work is needed for proper classification.

7.2. Modulational instabilities

The linear theory of Section 2 shows that there is no long-wavelength instability of magnetoconvection with the usual boundary conditions (though the situation is different for thermally insulating boundaries and sufficiently small imposed fields). Nonetheless there is a marginal mode at infinite wavelength, associated with the conservation of magnetic flux. Indeed, taking the vertical average of (3.7) and applying the boundary conditions, we get $\partial\bar{\xi}/\partial t = \zeta\nabla_H^2\bar{\xi}$, where $\bar{\xi} = \int_0^1 \xi dz$. Thus long wavelength magnetic modes vary slowly in time. In fact if $R - R^{(e)} = O(\epsilon^2)$, we expect both the usual convective modes and any long-wavelength mode with length scales of order ϵ^{-1} (which is also the standard modulation wavelength of the convective modes) to vary on the same timescale. This suggests a nontrivial coupling between the long-wavelength mode and the convection, which can affect the stability of uniform patterns of convection.

If we restrict attention to two-dimensional solutions depending only on x and z , then we can introduce the modulation scale $X = \epsilon x$ and slow timescale $T = \epsilon^2 t$. Then Cox and Matthews (2001) make the *ansatz* as in (4.1), but now with just x -dependence and long-wavelength modulation:

$$\mathbf{u} = \epsilon(\hat{\mathbf{u}}A(X, T)e^{ik_c x} + c.c.) \sin \pi z + \epsilon^2 \mathbf{u}_2 + O(\epsilon^3), \quad (7.2)$$

$$\theta = \epsilon(\hat{\theta}A(X, T)e^{ik_c x} + c.c.) \sin \pi z + \epsilon^2 \theta_2 + O(\epsilon^3), \quad (7.3)$$

$$\xi = \Xi(X, T) + \epsilon(\hat{\xi}A(X, T)e^{ik_c x} + c.c.) \cos \pi z + \epsilon^2 \xi_2 + O(\epsilon^3). \quad (7.4)$$

Here $\hat{\mathbf{u}}$, $\hat{\theta}$, etc. are (complex) constants. The first term in the last expression is the potential for a vertical magnetic field depending only on X . Substituting into the nonlinear equations, and performing the usual manipulations a pair of evolution equations is obtained for the

envelope function $A(X, T)$ and the mean field perturbation $B(X, T) \equiv -\partial^2 \Sigma / \partial X^2$:

$$\frac{\partial A}{\partial T} = a_1 A + a_2 \frac{\partial^2 A}{\partial X^2} - a_3 |A|^2 A - a_4 AB, \quad (7.5)$$

$$\frac{\partial B}{\partial T} = \frac{\partial^2}{\partial X^2} \left(\zeta B + \frac{\pi^2}{\zeta(\pi^2 + k_c^2)} |A|^2 \right). \quad (7.6)$$

The constants a_i have simple expressions as functions of σ , ζ and k_c . The physical meaning of the new terms involving B is that a local increase in the mean magnetic field strength leads to a reduction in the effective Rayleigh number, while vigorous convection (larger $|A|^2$) locally expels magnetic field. Matthews and Cox (2000) have investigated the stability of spatially periodic convection rolls of the form $B = 0$, $A = Re^{i\ell X}$. They find that whatever the value of ℓ there is a long-wavelength instability, leading to modulation of the pattern, if the parameters are such that

$$\pi^2(2k_c^2 - \pi^2)(3\pi^2 + k_c^2) > \zeta^2 k_c^4 (\pi^2 + k_c^2) > \pi^2(2k_c^2 - \pi^2)(\pi^2 - k_c^2). \quad (7.7)$$

If the first inequality is violated there is a band of stable rolls with $|\ell|$ sufficiently small, while when the second fails the bifurcation to rolls is subcritical, and small amplitude steady convection cannot occur. Of course there may be a prior instability to oscillatory convection if $\zeta < 1$ and $Q > Q^*(\sigma, \zeta)$. It seems likely that there is a similar mode of instability of standing or traveling waves, but this has not yet been investigated.

The instability turns out to be supercritical in long finite domains. However, the bifurcation becomes degenerate at infinite wavelength and when higher order terms are included it is in fact generically subcritical (see Proctor, 2001). Numerical computations of the full PDEs in long boxes show that the instability saturates, giving significant variations in roll amplitudes. The results can be extended without much difficulty to two horizontal dimensions, as shown by Pollicott *et al.* (2003) for a reduced model of convection similar to that used by Blanchflower (1999a,b). The results show that “flux separation” as described in an earlier section, with isolated domains of vigorous convection with small magnetic field can appear in Boussinesq convection at least in some parameter ranges.

All these results concern Boussinesq convection. If the layer is stratified the regular patterns near onset are typically hexagonal, and for such patterns quadratic terms of the form A^{*2} will appear in (7.5). It can then be shown that provided $a_4 > 0$ there is always a region of instability of the hexagonal pattern. Some results have been found by Cox and Matthews (2003). Cox *et al.* (2004) have shown that the instability

can be highly subcritical, making modulated states the norm in many circumstances.

8. Effects of stratification

The Boussinesq approximation cannot give an accurate representation of field behavior in the Sun, since the photosphere is an ionized plasma, with highly non-uniform properties. There are two major strands to recent work on compressible magnetoconvection. Nordlund, Stein, Schüssler and their collaborators have made ambitious attempts to give an accurate model of the solar convection zone, with all possible physics included. An alternative approach that we consider first is to set up simple model problems (by analogy with the Boussinesq problem studied for so many years), to include new physical processes incrementally, and to aim for fundamental understanding.

The principal effects of compressibility on magnetoconvection in a layer are (i) the effect of the magnetic pressure, which acts to reduce the local gas pressure when fields are strong. This will in turn lead to a reduction in density, which will help to evacuate regions of strong magnetic field. Indeed, since gas pressures cannot be negative, and $p + B^2/2\mu_0 \approx \text{const.}$ in a quasi-static tube of flux there is a theoretical upper limit to any magnetic concentration of $\sqrt{2\mu_0 p_e}$, where p_e is the external pressure. When magnetic fields are concentrated by convection the resulting evacuation tends to reduce the peak fields that can be produced. (ii) Because the centers of the flux structures are partially evacuated, they have increased buoyancy. This can lead, near the top of the layer where flux concentrations occur in downflow regions, to convective counter-cells, which serve further to inhibit concentration. (iii) In three dimensions, the lack of up-down symmetry favors hexagonal convection at onset. Even far from onset, where the flow is aperiodic and disordered, the motion is characterized by broad upwellings and narrow downflows, with important implications for the vertical transport of horizontal flux (e.g., Drobyshevski and Yuferev, 1974).

8.1. *Fundamental* models

The standard model of compressible magnetoconvection was introduced by Hurlbert and Toomre (1988), and has since been used by many others, and so we give brief details here. The domain is a layer of perfect gas, heated from below and with stress-free velocity boundary conditions. As for the Boussinesq calculations the field is constrained to be vertical

at the top and bottom of the layer, though in some cases the field matches a potential field at the top of the layer. In the absence of motion the field is uniform and vertical and the atmosphere is a polytrope with index m . If the dynamic viscosity and thermal conductivity are constant the Prandtl number σ is uniform, but $\zeta \propto \rho$ (density) and so increases with depth. After non-dimensionalizing we obtain the equations

$$P = T \rho, \quad \frac{\partial \rho}{\partial t} = -\nabla \cdot (\rho \mathbf{u}), \quad (8.1)$$

$$\nabla \cdot \mathbf{B} = 0, \quad \frac{\partial \mathbf{B}}{\partial t} = \nabla \times (\mathbf{u} \times \mathbf{B}) + \zeta_0 \bar{K} \nabla^2 \mathbf{B}, \quad (8.2)$$

$$\frac{\partial}{\partial t}(\rho \mathbf{u}) = -\nabla \cdot (\rho \mathbf{u} \mathbf{u} + F \mathbf{B} \mathbf{B}) - \nabla \left(P + \frac{1}{2} F |\mathbf{B}|^2 \right) + (m+1)\rho \hat{\mathbf{z}} + \nabla \cdot \boldsymbol{\tau}, \quad (8.3)$$

and

$$\begin{aligned} \frac{\partial}{\partial t} \left\{ \rho \left[\frac{T}{\gamma-1} + \frac{1}{2} |\mathbf{u}|^2 - (m+1)z \right] + \frac{1}{2} F |\mathbf{B}|^2 \right\} = \\ \nabla \cdot \left\{ \rho \left[\frac{\gamma T}{\gamma-1} + \frac{1}{2} |\mathbf{u}|^2 - (m+1)z \right] \mathbf{u} + \right. \\ \left. F \mathbf{B} \times (\mathbf{u} \times \mathbf{B} - \zeta_0 \bar{K} \nabla \times \mathbf{B}) - \bar{K} \nabla T + \mathbf{u} \cdot \boldsymbol{\tau} \right\}, \end{aligned} \quad (8.4)$$

where $\boldsymbol{\tau}$ is the viscous stress tensor. The constant $\zeta_0 = \zeta \rho_0 / \rho$ is the value of ζ at the top of the layer. The dimensionless thermal conductivity $\bar{K} = K / [c_p \rho_0 d (\mathfrak{R} \Delta T)^{1/2}]$ while the field strength is measured by the dimensionless quantity $F = B_0^2 / (\mu_0 \mathfrak{R} \rho_0 \Delta T) = \sigma \zeta_0 \bar{K}^2 Q$. Once the properties of the reference atmosphere are fixed, the two parameters \bar{K} and F define the state of the system. The static polytrope is given by $T = z$, $\rho = (z/z_0)^m$, $p = z^{m+1}/z_0^m$ and the corresponding ratio of the gas pressure to the magnetic pressure at mid-layer is $\beta = 2(z_0 + 1/2)^{m+1}/(F z_0^m)$, while the superadiabatic gradient is measured by the Rayleigh number

$$R = (m+1)^2 (\nabla - \nabla_{ad}) \frac{(z_0 + \frac{1}{2})^{2m+1}}{\sigma \bar{K}^2 z_0^{2m}}, \quad (\nabla - \nabla_{ad}) \equiv \frac{1}{m+1} - \frac{\gamma-1}{\gamma}.$$

For more details see Proctor (1992). In the original calculations of Hurlburt and Toomre (1988), the phenomenon of flux sheet evacuation was clearly shown. While for weak fields the sheets are approximately gaussian, for larger fields there are internal counterflows and the peak fields are off-center (Fig. 10). In Weiss *et al.* (1990) the density ratio

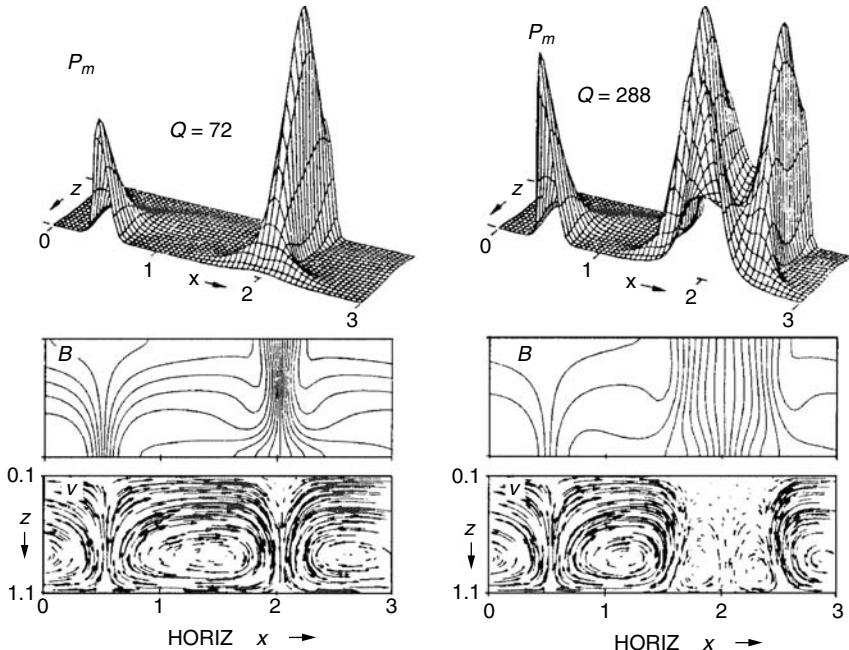


Figure 10. Magnetic pressure ($\propto |\mathbf{B}|^2$) and velocity fields for $R = 10^5$. The flux sheet is dynamically active in the right-hand picture (from Hurlburt and Toomre, 1988).

$[(z_0+1)/z_0]^m = 11$, and ζ_0 was chosen to be 1.2 so that the effective value of ζ was less than unity at the top of the layer and greater than unity at greater depths, thus modeling conditions in sunspot umbrae (see, e.g., Meyer *et al.*, 1974). Indeed, although the initial bifurcation was to steady convection, in the mildly nonlinear regime a mixed mode oscillation was established, with velocities changing sign at the top and being almost steady at the base. This was interpreted in terms of umbral dots, but the results seemed to be strongly influenced by the aspect ratio selected. More recent computations in very wide cells (Blanchflower *et al.*, 1998) show that in many cases very long horizontal scales develop in the flow, separated by large stagnant flux sheets. In further papers (Proctor *et al.*, 1994; Brownjohn *et al.*, 1995), with imposed horizontal and vertical fields, respectively, the occurrence of waves and oscillations in a shallow layer was investigated. Standing, modulated and traveling waves were found for different parameter values; again, however, these suggestive results may well be partly due to the interaction between modes with different critical wavenumbers, and therefore once more the size of the periodic domain in the horizontal might play an important role.

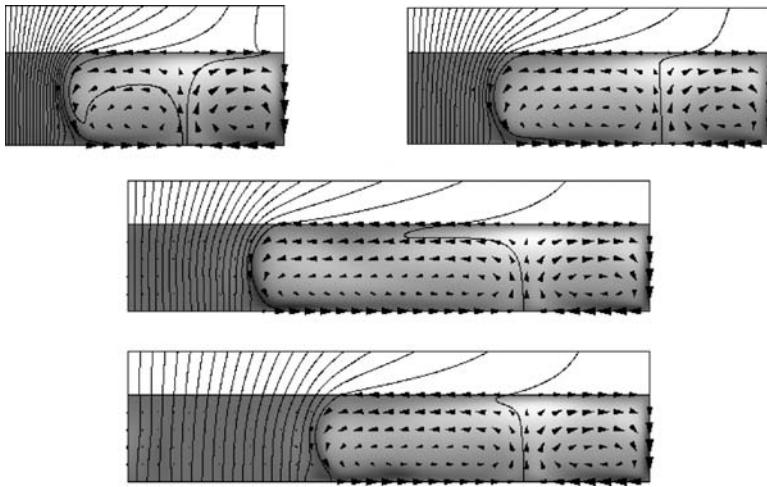


Figure 11. Cylindrical compressible magnetohydroconvection (adapted from Hurlburt and Rucklidge, 2000). Various aspect ratios and values of Q are shown. In all cases the flow near the flux tube is inward at the top of the layer.

The study of flux concentrations in the photosphere suggests that axisymmetric configurations might repay investigation. This has been done in the compressible case by Hurlburt and Rucklidge (2000). They consider a layer of polytropic gas in a cylinder, heated from below. The field at the bottom of the layer is vertical; that at the top matches to a potential field that tends to a uniform field at infinity. For narrow cylinders there is a single annular gyre, inward at the top of the layer, while for larger aspect ratios there is a more complex pattern, with a second outer gyre (Fig. 11). In this laminar situation the flux tube is such that the peak field takes a more or less constant value independent of the flux (as in the Boussinesq case), while convection occupies the rest of the cell and tries to have a cellular aspect ratio of order unity. In Hurlburt *et al.* (2002) these ideas are taken rather further, and the heating of the corona is modeled by regarding the latter as due to the Poynting flux into the coronal structures. The model is of course simplified in that it lacks a transition region and chromosphere. Nonetheless the results show the way toward providing a useful two-way interface between numerical simulations and detailed observational data such as that from TRACE. Three-dimensional computations in reasonable sized domains with decent resolution have only become available relatively recently. Matthews *et al.* (1995) considered a shallow stratified layer with the aim of studying the effects of weak stratification.

In narrow boxes alternating pulsating waves were found similar to those in the Boussinesq case. For larger values of ζ_0 and in wider boxes steady near-hexagonal convection cells were preferred near onset.

The deep layer has been considered by Weiss *et al.* (1996), in a square periodic domain of order one aspect ratio. At large Q there is steady convection with a nearly hexagonal planform, while for smaller imposed fields we see a sequential loss of symmetry leading to mixed mode oscillations similar to the two-dimensional case, and ultimately to disordered flow which is clearly much affected by the boundary conditions. The bifurcation theory aspects of this work were taken forward by Rucklidge *et al.* (2000), who looked in detail at various box sizes and values of ζ_0 in the deep layer case, varying R for fixed values of Q . The way in which the original patterns break down as R increases or Q decreases involves a complicated sequence of bifurcations, which can be understood only by considering the changes in symmetry. A new type of solution is the so-called “wobbler”, which takes the form of standing waves involving oscillations of the cell boundaries about a vertical axis. They can only appear as a result of a number of secondary bifurcations, and the complex bifurcation sequences that occur can only be understood by means of group theoretic ideas. This paper also contains a short section on dynamics in boxes of larger aspect ratio in the fully nonlinear regime. Recent advances in computational power mean that aspect ratios of 8 are possible; this seems to be wide enough to remove any boundary influence in the selection of the widest plume sizes. These solutions exhibit what Tao *et al.* (1998) termed “flux separation”; the division of the convective region into two types of region. In one, magnetic field is expelled from vigorously convecting plumes, which then join together near the surface to form very wide structures. In the other, the field is large enough to provide a strong effect on convection, which then takes the form of narrow cells that are much less vigorous (Fig. 12). This coexistence can occur for a range of values of Q with fixed R , and the transition is hysteretic; it is possible for the same parameters to have either flux separation or relatively uniform small scales of motion. A careful survey of the phenomenon has been undertaken by Weiss *et al.* (2002), who have also looked at quantitative features of the solutions. It is clear from flux conservation that if flux is expelled from parts of the domain the effective value of Q rises elsewhere. By estimating the size of the region from which flux is expelled they conclude that an upper limit to the extent of this region is given by an *effective Q* which is just less than that which would be sufficient to suppress convection entirely. The observed transitions have some of the character of a phase transition, but there is as yet no proper theory to account for the hysteresis.

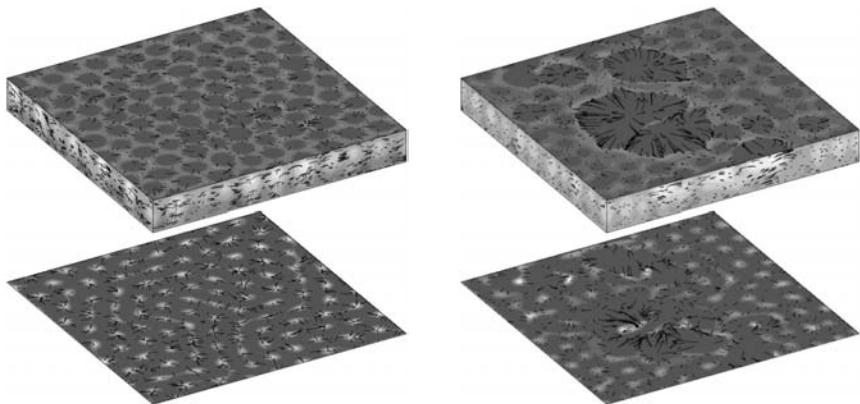


Figure 12. Flux separation (from Weiss *et al.*, 2002). In each case the aspect ratio is 8, $R + 10^5$ and $Q = 1600$. In the second figure there are two scales of motion. Shading on top and bottom indicates field strength, on sides temperature fluctuation.

The flux separation transition has some affinities with the phenomenon of *convectons* discussed above. It might be thought that the phenomenon could be found also in Boussinesq convection, though there has been little success in showing analogous bimodality of cellular scales. Very recent results by Pollicott *et al.* (2003), using a reduced model, do show suggestive similar features.

There have also been investigations of non-axisymmetric instabilities of axisymmetric compressible convection. Hurlburt and Alexander (2003) have studied the problem posed by Hurlburt and Rucklidge (2000), but now allow azimuthal variation within a 30° wedge. They find that if there is sufficient flux in the cylinder the axisymmetric configuration breaks down into a fluted structure. The non-axisymmetric parts are insignificant near the center for geometrical reasons, and the picture bears a suggestive resemblance to the structure of sunspots, with a transition between umbra and penumbra. The preferred wavenumber of the instability may be influenced by the choice of wedge angle. Similar results were obtained in a Boussinesq cartesian model by Tildesley (2003a,c). Examination of the solutions shows local outflows in rising regions of flow, similar to the observed Evershed flows in penumbras.

8.2. Nonlinear pumping

The phenomenon of downward pumping of flux is important both near the surface (see below) as an inhibitor of dynamo action and in many

other contexts. Tobias *et al.* (2001) have investigated the effects of non-linear convection on an initially horizontal field near the base of the convection zone (Fig. 13; see color insert following page 234). They show that the net downward pumping effect produces a flow of field toward and into the stable overshoot region supposed to underly the convective zone. Such a confinement effect is likely to be an important part of any coherent model of the solar cycle. The same mechanism has been invoked in a different context (at the top of the convection zone) to explain the complex structure of sunspot penumbrae. Some of the field lines in the outer penumbra appear to drop below the solar surface, while others continue upward and only return at great distances. Thomas *et al.* (2002) have argued that the locally reentrant fields must be subject to downward pumping to counteract their buoyancy.

8.3. *Oblique fields and traveling waves*

When the imposed field is neither vertical nor horizontal, and the layer is stratified, the left-right symmetry associated with either vertical or horizontal fields is broken, and steady convection becomes impossible. If the vertical case favors oscillations at onset, these will be replaced by two distinct sets of traveling waves, with critical Rayleigh numbers depending linearly on the tilt angle provided the latter is small. If the vertical case yields steady convection, this is replaced with a traveling wave. The linear two-dimensional situation, in which the field angle was specified at the horizontal boundaries, was investigated by Matthews *et al.* (1992). They showed that the direction of the traveling waves could change sign as the tilt angle changed. Hurlburt *et al.* (1996) looked at fully nonlinear solutions using a variety of boundary conditions (not all such conditions are satisfactory; there can be a horizontal Maxwell stress exerted on the boundary in the nonlinear regime). They found that when there is no such stress, there is a velocity at the top of the layer in the direction of the field tilt. For weak imposed fields the traveling waves have a phase velocity in the opposite direction to this flow, while for strong fields the phase velocity and the flow velocity are in the same sense. It might be argued that these two-dimensional solutions would be irrelevant since in three dimensions the convection rolls would line up with the the tilt direction. In fact, as shown by Matthews *et al.* (1992), for a range of small tilt angles the rolls preferentially align with the axes *perpendicular* to the tilt; there is then a bifurcation to nearly aligned rolls at larger angles. Preliminary computations undertaken by Hurlburt *et al.* (2000) in a geometry where rolls are not preferred for

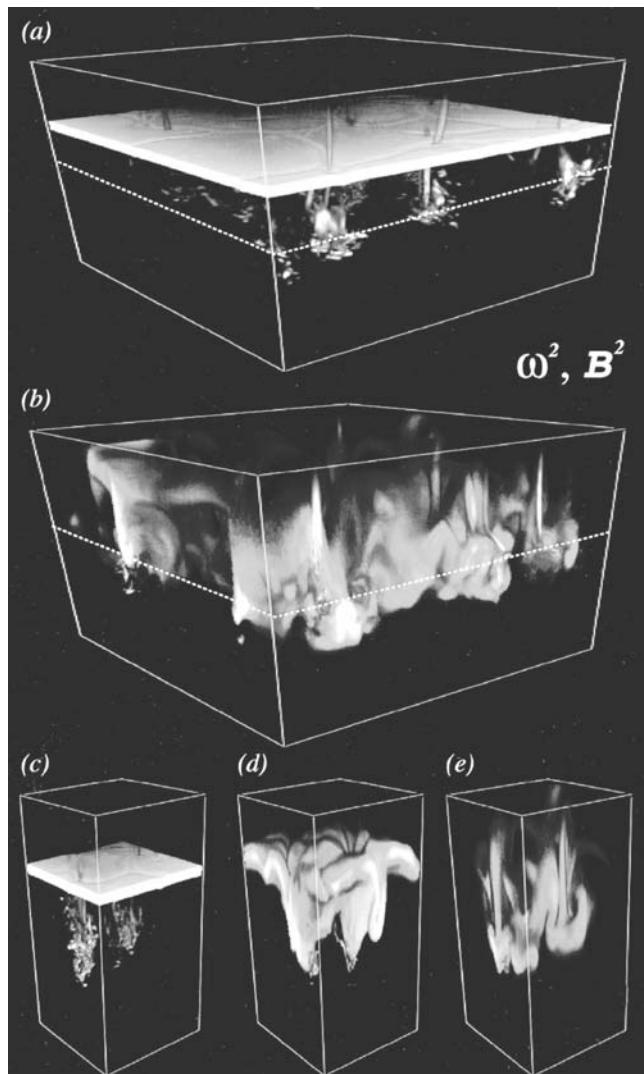


Figure 13. Pumping of horizontal fields by three-dimensional convection above a stably stratified layer (from Tobias *et al.*, 2001). (See color insert.) (a) Initial configuration with a layer of magnetic field inserted in the unstable convection zone. (b) Later time, showing concentrations of the magnetic field in the stable region. (c)–(e) Volume renderings for a subvolume of the full domain centered around a couple of coherent downflows. The strong plumes pump magnetic flux downward and amplify the magnetic energy locally by inductive processes.

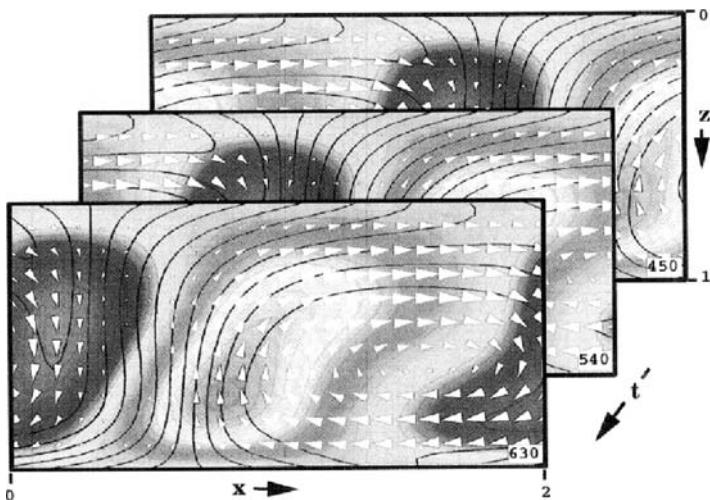


Figure 14. Traveling waves in oblique fields (from Hurlburt *et al.*, 1996). The field points up and to the right, but the phase velocity of the wave is to the left for this moderate field strength.

vertical fields show that for small tilt the pattern drifts without much change of form while roll-like structures appear at larger angles. They suggest that the results may shed light on the flows of bright points in sunspot penumbrae.

8.4. “Realistic” models

Another important area of research is given by so-called “realistic models” of solar surface magnetoconvection (cf., the recent review of Schüssler and Knölker, 2001). In this approach, pioneered by Nordlund and collaborators (Nordlund, 1983) the aim is to include effects of radiative transfer and ionization so as to produce an accurate model of the solar atmosphere, with diagnostic information in terms of line profiles, etc., that can be directly compared with observation. An important difference from the simulations described above is that the boundaries are “open”; fluid can leave and enter the domain. This has the potential to give a more realistic picture of the fluxes, etc., in the upper layers, with the disadvantage that the results for deeper seated fields may depend sensitively on the prescription for the in- and outflows. The earlier results were necessarily of limited resolution but more recent computations (Nordlund and Stein, 1990) were able convincingly to show the

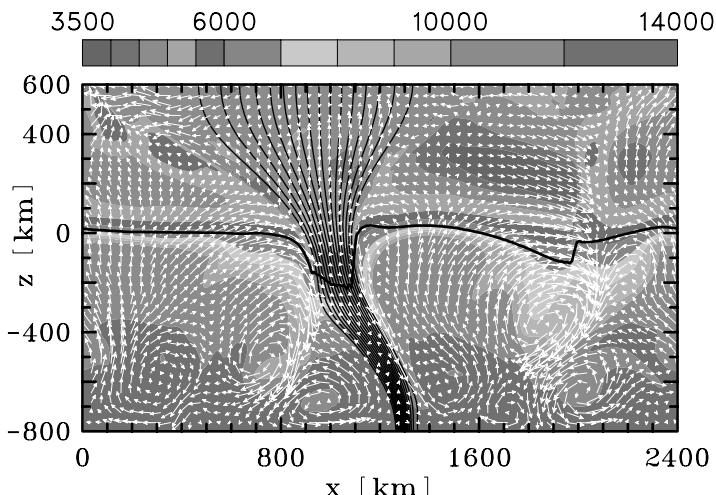


Figure 15. Two-dimensional magnetoconvection with radiative transfer (courtesy of O. Steiner). (See color insert.) An intense flux tube can be seen, together with the depression of the surface due to evacuation and cooling.

stabilization of the granulation pattern at depth by the magnetic fields that are drawn into the downdrafts (Fig. 15; see color insert following page 234). The diagnostics emerging from these calculations are in excellent agreement with observations, thus showing that the surface field features are insensitive to conditions at depth due to the strong stratification.

The alternatives to moderate resolution three-dimensional calculations are extremely well-resolved two-dimensional ones; these have been carried out by Steiner and collaborators (Steiner *et al.*, 1994; Steiner *et al.*, 1996; Steiner *et al.*, 1998; Grossmann-Dörth *et al.*, 1998) using an adaptive code. They are able very accurately to resolve small-scale flux structures and in particular to show how the concentration of flux leads to evacuation of the flux sheet, consequent falling plasma and the enhancement of concentration (“convective collapse”). For strong fields there is a strong rebound from this evacuation, leading possibly to spicule-like outflows above intense flux sheets.

Very recent three-dimensional work by Stein *et al.* (2002, 2003) has looked at the effect on magnetic fields of mesoscale granular flows. In one study a vertical magnetic field is allowed to be swept into a downdraft. Cooling, evacuation, etc., proceeds not dissimilarly to that for the two-dimensional case. In the other study the fate of a horizontal flux distribution released at the base of the convection zone is followed in

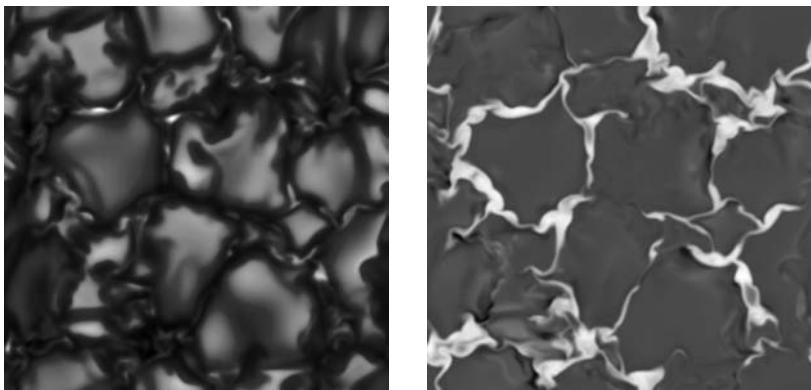


Figure 16. Three-dimensional magnetoconvection with radiative transfer (from Vögler and Schüssler, 2003). (See color insert.) Left image shows field intensity, right image vertical field strength. Averaged imposed field is 200G. Note the prominent mesogranular scale of the magnetic field.

time. It is shown that individual fluid elements spend very little time in the upper convection zone, and so the scope for local flux stretching and amplification is limited. Since this simulation has non-zero imposed field, it cannot shed light on local dynamo generation, but does pose the interesting question of whether surface dynamos can exist in the presence of strong downward pumping.

Finally, Vögler and Schüssler (2003), in collaboration with Cattaneo and others, have developed a new three-dimensional code including radiative transfer. Preliminary results of a calculation with an initial vertical field show the occurrence in the magnetic intensity plots of the mesoscale (Fig. 16; see color insert following page 234), comprising several granular scales, already reported by Cattaneo *et al.* (2002) for Boussinesq simulations of convective dynamos. They also bear a strong resemblance to the results of Tao *et al.* (1998) and Weiss *et al.* (2002). A detailed explanation of this mesoscale will be an active topic of research.

9. Discussion

The increasing speed of computers has transformed the study of magnetoconvection, as large-scale numerical simulations are now well within reach. Nonetheless the simplified models of the past can be seen to be relevant in the interpretation of the modern calculations. The bifurcation problem will continue to excite fluid dynamicists and bifurcation theorists; for direct solar application, the outstanding question is the

connection between magnetoconvection and dynamo action. All vigorous three-dimensional flows are likely to act as dynamos if there is no imposed field; when there is an ambient field the distinction between dynamo action and magnetoconvection becomes blurred, though some progress has been made in the Boussinesq case (Cattaneo *et al.*, 2002). The next years will be spent sorting out this interaction, as well as providing better and better detailed models of flux tube and sunspot dynamics. A longer term challenge is to use the surface data to learn about the solar interior; the robustness of patterns of convection with different boundary conditions at depth means that such probing will require subtlety. New challenges will also be faced in interpreting the burgeoning streams of data. As a result of this synergy, one might hope that the Holy Grail of a complete model of the photosphere may yet be reached in the next decade.

Acknowledgments

I am happy to acknowledge discussions with many colleagues, but especially with Nigel Weiss.

References

- Arter, W. A., Proctor, M. R. E. and Galloway, D. J., "New results on the mechanism of magnetic flux pumping by three-dimensional convection," *Mon. Not. R. Astron. Soc.* **201**, 57P–61P (1982).
- Ashwin, P. A. and Rucklidge, A. M., "Cycling chaos: its creation, persistence and loss of stability in a model of nonlinear magnetoconvection," *Physica D* **122**, 134–154 (1998).
- Blanchflower, S. M., *Modelling Photospheric Magnetoconvection*, Ph.D. Dissertation; University of Cambridge (1999a).
- Blanchflower, S. M., "Magnetohydrodynamic convective cells," *Phys. Lett. A.* **261**, 74–81 (1999b).
- Blanchflower, S. M. and Weiss, N. O., "Three-dimensional magnetohydrodynamic convective cells," *Phys. Lett. A.* **294**, 297–303 (2002).
- Blanchflower, S. M., Rucklidge, A. M. and Weiss, N. O., "Modelling photospheric magnetoconvection," *Mon. Not. R. Astron. Soc.* **301**, 593–608 (1998).
- Brownjohn, D. P., Hurlburt, N. E., Proctor, M. R. E. and Weiss, N. O., "Nonlinear compressible magnetoconvection Part 3. Traveling waves in a horizontal field," *J. Fluid Mech.* **300**, 287–309 (1995).
- Busse, F. H., "Nonlinear interaction of magnetic field and convection," *J. Fluid Mech.* **71**, 193–206 (1975).

- Cattaneo, F., "On the origin of magnetic fields in the quiet photosphere," *Astrophys. J.* **515**, L39–L42 (1999).
- Cattaneo, F., Emonet, T. and Weiss, N. O., "On the interaction between convection and magnetic-fields," *Astrophys. J.* **588**, 1183–1198 (2003).
- Cattaneo, F., Lenz, D. and Weiss, N. O., "On the origin of solar mesogranulation," *Astrophys. J.* **563**, L91–L94 (2002).
- Childress, S. and Gilbert, A. D., *Stretch, Twist, Fold: The Fast Dynamo*, Springer-Verlag, Berlin (1995).
- Clune, T. and Knobloch, E., "Pattern selection in three-dimensional magnetoconvection," *Physica D* **74**, 151–176 (1994).
- Cox, S. M. and Matthews, P. C., "New instabilities in two-dimensional rotating convection and magnetoconvection," *Physica D* **149**, 210–229 (2001).
- Cox, S. M. and Matthews, P. C., "Instability and localisation of patterns due to a conserved quantity," *Physica D* **175**, 196–219 (2003).
- Cox, S. M., Matthews, P. C. and Proctor, M. R. E. "Modulation instabilities of hexagon patterns," to be submitted (2004).
- Dangelmayr, G., Fiedler, B., Kirchgässner, K. and Mielke, A., *Dynamics of Non-linear Waves in Dissipative Systems: Reduction, Bifurcation and Stability*, Pitman Research Notes in Mathematics Series 352, pp. 44–46, Longman, (1996).
- Dawes, J. H. P., "The $1:\sqrt{2}$ Hopf/steady-state mode interaction in three-dimensional magnetoconvection," *Physica D* **139**, 109–36 (2000).
- Drobyshevski, E. M. and Yuferev, V. S., "Topological pumping of magnetic flux by three-dimensional convection," *J. Fluid Mech.* **65**, 33 (1974).
- Galloway, D. J., "Axisymmetric convection with a magnetic field," in: *Problems of Stellar Convection* (Eds. E. A. Spiegel and J. P. Zaha), Springer-Verlag, Berlin (1976).
- Galloway, D. J., "The origin of running penumbral waves," *Mon. Not. R. Astron. Soc.* **184**, 49P–52P (1978).
- Galloway, D. J. and Moore, D. R., "Axisymmetric convection in the presence of a magnetic field," *Geophys. Astrophys. Fluid Dynam.* **12**, 73–105 (1979).
- Galloway, D. J. and Proctor, M. R. E., "The kinematics of hexagonal magnetoconvection," *Geophys. Astrophys. Fluid Dynam.* **24**, 109–136 (1983).
- Galloway, D. J., Proctor, M. R. E. and Weiss, N. O., "Magnetic flux ropes and convection," *J. Fluid Mech.* **87**, 243–261 (1978).
- Grossmann-Dörth, U., Schüssler, M. and Steiner, O., "Convective intensification of solar surface magnetic fields: results of numerical experiments," *Astron. Astrophys.* **337**, 928–939 (1998).
- Halford, A. R. and Proctor, M. R. E., "An oscillatory secondary bifurcation for magnetoconvection and rotating convection at small aspect ratio," *J. Fluid Mech.* **467**, 241–257 (2002).
- Hughes, D. W. and Proctor, M. R. E., "Magnetic fields in the solar convection zone: magnetoconvection and magnetic buoyancy," *Annu. Rev. Fluid Mech.* **20**, 187–223 (1988).
- Hurlburt, N. E. and Alexander, D., "Sunspot dynamics and Coronal Heating," *Adv. Space Res.* **14**, 19–24 (2003).

- Hurlburt, N. E. and Rucklidge, A. M., "Development of structures in pores and sunspots: flows around axisymmetric flux tubes," *Mon. Not R. Astron. Soc.* **314**, 793–807 (2000).
- Hurlburt, N. E. and Toomre, J., "Magnetic fields interacting with nonlinear compressible convection," *Astrophys. J.* **327**, 920–932 (1988).
- Hurlburt, N. E., Alexander, D. and Rucklidge, A. M., "Complete models of axisymmetric sunspots: magnetoconvection with coronal heating," *Astrophys. J.* **577**, 993–1005 (2002).
- Hurlburt, N. E., Matthews, P. C. and Proctor, M. R. E., "Nonlinear compressible convection in oblique magnetic fields," *Astrophys. J.* **437**, 933–938 (1996).
- Hurlburt, N. E., Matthews, P. C. and Rucklidge, A. M., "Solar magnetoconvection," *Solar Phys.* **192**, 109–118 (2000).
- Julien, K., Knobloch, E. and Tobias, S. M., "Strongly nonlinear magnetoconvection in three dimensions," *Physica D* **128**, 105–129 (1999).
- Julien, K., Knobloch, E. and Tobias, S. M., "Nonlinear magnetoconvection in the presence of strong oblique fields," *J. Fluid Mech.* **410**, 285–322 (2000).
- Julien, K. A., Prat, J., Proctor, M. R. E. and Rucklidge, A. M., "Mean flow instabilities of two-dimensional convection in strong magnetic fields," *J. Fluid. Mech.*, to be submitted (2004).
- Knobloch, E. and Proctor, M. R. E., "Nonlinear periodic convection in double-diffusive systems," *J. Fluid Mech.* **108**, 291–316 (1981).
- Knobloch, E., Weiss, N. O. and Da Costa, L. N., "Oscillatory and steady convection in a magnetic field," *J. Fluid Mech.* **113**, 153–186 (1981).
- Matthews, P. C., "Asymptotic solutions for nonlinear magnetoconvection," *J. Fluid Mech.* **387**, 397–409 (1999).
- Matthews, P. C. and Cox, S. M., "Pattern formation with a conservation law," *Nonlinearity*, **13**, 1293–1320 (2000).
- Matthews, P. C., Hurlburt, N. E., Proctor, M. R. E. and Brownjohn, D. P., "Compressible magnetoconvection in oblique fields: linearised theory and simple nonlinear models," *J. Fluid Mech.* **240**, 559–569 (1992).
- Matthews, P. C., Proctor, M. R. E. and Weiss, N. O., "Compressible magnetoconvection in three dimensions: planforms and nonlinear behaviour," *J. Fluid Mech.* **305**, 281–305 (1995).
- Matthews, P. C., Proctor, M. R. E., Rucklidge, A. M. and Weiss, N. O., "Pulsating waves in nonlinear magnetoconvection," *Phys. Lett. A.* **183**, 69–75 (1993).
- Matthews, P. C., Rucklidge, A. M., Weiss, N. O. and Proctor, M. R. E., "The three-dimensional development of the shearing instability of convection," *Phys. Fluids* **8(6)**, 1350–1352 (1996).
- Meyer, F., Schmidt, H. U., Weiss, N. O. and Wilson, P. R., "The growth and decay of sunspots," *Mon. Not. R. Astron. Soc.* **169**, 35–57 (1974).
- Nordlund, Å., "Numerical 3-D simulations of the collapse of photospheric flux tubes," in: *IAU Symp. 102—Solar and Stellar Magnetic Fields: Origins and Coronal Effects* (Ed. J. O. Stenflo), Kluwer, Dordrecht (1983).
- Nordlund, Å. and Stein, R. F., "Solar magnetoconvection," in: *IAU Symp. 138: Solar Photosphere: Structure, Convection and Magnetic Fields* (Ed. J. O. Stenflo), pp. 191–211, Kluwer, Dordrecht (1990).

- Pollicott, S. L., Cox, S. M. and Matthews, P. C., "A two-dimensional model for magnetoconvection," *Phys. Rev. E*, submitted (2003).
- Proctor, M. R. E., "Collective bifurcation of flux sheets in two-dimensional magnetoconvection," *Mon. Not. R. Astron. Soc.* **204**, 935–943 (1983).
- Proctor, M. R. E., "Columnar convection in doubly-diffusive systems," *Cont. Math.* **56**, 267–276 (1986).
- Proctor, M. R. E., "Magnetoconvection," in: *Theory and Observations of Sunspots* (Eds. N. O. Weiss and J. H. Thomas), pp. 221–241, Kluwer, (1992).
- Proctor, M. R. E., "Finite amplitude behaviour of the Matthews-Cox instability," *Phys. Lett. A* **292**, 181–187 (2000).
- Proctor, M. R. E. and Galloway, D. J., "The dynamic effect of flux ropes on Rayleigh-Benard convection," *J. Fluid Mech.* **90**, 273–287 (1979).
- Proctor, M. R. E. and Weiss, N. O., "Magnetoconvection," *Rep. Prog. Phys.* **45**, 1317–1379 (1982).
- Proctor, M. R. E. and Weiss, N. O., "Amplification and maintenance of thin magnetic flux tubes by compressible convection," in: *The Hydrodynamics of the Sun*, pp. 77–80, ESA SP-220, Noordwijk (1984).
- Proctor, M. R. E., Weiss, N. O., Brownjohn, D. P. and Hurlburt, N. E., "Nonlinear compressible magnetoconvection. Part 2. Streaming instabilities in two dimensions," *J. Fluid Mech.* **280**, 227–253 (1994).
- Rucklidge, A. M. and Matthews, P. C., "Analysis of the shearing instability in nonlinear convection and magnetoconvection," *Nonlinearity* **9**(2), 311–351 (1996).
- Rucklidge, A. M., Weiss, N. O., Brownjohn, D. P. and Proctor, M. R. E., "Oscillations and secondary bifurcations in nonlinear magnetoconvection," *Geophys. Astrophys. Fluid Dynam.* **68**, 133–150 (1993).
- Rucklidge, A. M., Weiss, N. O., Brownjohn, D. P., Matthews, P. C. and Proctor, M. R. E., "Compressible magnetoconvection in three dimensions: pattern formation in a strongly stratified layer," *J. Fluid Mech.* **419**, 283–323 (2000).
- Schüssler, M. and Knölker, M., "Magneto-convection," in: *Magnetic Fields across the Hertzsprung-Russell Diagram: Astron. Soc. Pacific Conf. Ser.* **248** (Eds. G. Mathys, S. K. Solanki and D. T. Wickramasinghe), pp. 115–124 (2001).
- Stein, R. F. and Nordlund, Å., "Realistic solar convection simulations," *Solar Phys.* **192**, 91–108 (2000).
- Stein, R. F., Bercik, D. and Nordlund, Å., "Solar convection and magnetoconvection simulations," *Il Nuovo Cimento C, Geophys. Space Phys.* **25**, 513–522 (2002).
- Stein, R. F., Bercik, D. and Nordlund, Å., "Solar Surface Magneto-Convection," in: *Current Theoretical Models and High Resolution Solar Observations: Preparing for ATST*, *Astron. Soc. Pacific, Conf. Ser.* **286** (Eds. A. A. Pevtsov and H. Uitenbroek), pp. 121–132 (2003).
- Steiner, O., Knölker, M. and Schüssler, M., "Dynamic Interaction of convection with magnetic flux sheets: First results of a new MHD code," in: *Solar Surface Magnetism* (Eds. R. Rutten and C. Schrijver), pp. 441–470, Kluwer, Dordrecht (1994).

- Steiner, O., Grossmann-Dörth, U., Schüssler, M. and Knölker, M., "Polarised radiation diagnostics of magnetohydrodynamic models of the solar atmosphere," *Solar Phys.* **164**, 223–242 (1996).
- Steiner, O., Grossmann-Dörth, U., Knölker, M. and Schüssler, M., "Dynamical interaction of solar magnetic elements and granular convection: Results of a numerical simulation," *Astrophys. J.* **495**, 468–484 (1998).
- Tao, L., Weiss, N. O., Brownjohn, D. P., and Proctor, M. R. E., "Flux separation in stellar magnetoconvection," *Astrophys. J.* **496**, L39–L42 (1998).
- Thomas, J. H., Weiss, N. O., Tobias, S. M. and Brummell, N. H., "Downward pumping of magnetic flux as the cause of filamentary structures in sunspot penumbrae," *Nature* **420**, 390–393 (2002).
- Tildesley, M. J., *Modelling Filamentary Structure in Sunspots*, Ph.D. Dissertation; University of Cambridge (2003a).
- Tildesley, M. J., "On the origin of filamentary structure in sunspot penumbrae: Linear instabilities," *Mon. Not. R. Astron. Soc.* **338**, 497–507 (2003b).
- Tildesley, M. J., "Convectively driven filamentary instabilities in sunspot penumbrae," *Astron. Soc. Pacific Conf. Ser.* **286**, 147 (2003c).
- Tobias, S. M., Brummell, N. H., Clune, T. L. and Toomre, J., "Transport and Storage of Magnetic Field by Overshooting Turbulent Compressible Convection," *Astrophys. J.* **549**, 1183–1203 (2001).
- Vögler, A. and Schüssler, M., "Studying magneto-convection by numerical simulation," *Astron. Nachr.* **324**, 399–404 (2003).
- Weiss, N. O., "Convection in an imposed magnetic field. Part 1. The development of nonlinear convection," *J. Fluid. Mech.* **108**, 247–272 (1981a).
- Weiss, N. O., "Convection in an imposed magnetic field. Part 2. The dynamical regime," *J. Fluid. Mech.* **108**, 273–289 (1981b).
- Weiss, N. O., "Magnetoconvection," in: *Advances in the Physics of Sunspots* (Eds. B. Schmieder, J. C. del Toro Iniesta and M. Vázquez), pp. 21–33, Astron. Soc. Pacific, San Francisco (1997).
- Weiss, N. O., Proctor, M. R. E. and Brownjohn, D. P., "Magnetic flux separation in photospheric convection," *Mon. Not. R. Astron. Soc.* **337**, 293–304 (2002).
- Weiss, N. O., Brownjohn, D. P., Hurlburt, N. E. and Proctor, M. R. E., "Oscillatory convection in sunspot umbrae," *Mon. Not. R. Astron. Soc.* **245**, 434–452 (1990).
- Weiss, N. O., Brownjohn, D. P., Matthews, P. C. and Proctor, M. R. E., "Photospheric convection in strong magnetic fields," *Mon. Not. R. Astron. Soc.* **283**, 1153–1164 (1996).

9 Alfvén waves within the Earth's core

D. Jault and G. Légaut

*Laboratoire de Géophysique Interne et Tectonophysique, CNRS,
Université Joseph Fourier de Grenoble, BP 53, 38041 Grenoble
Cedex 9, France*

Alfvén waves within the Earth's fluid core have been detected through the monitoring of the magnetic field at observatories situated on the Earth's surface. Inside the core, geostrophic motions consist of oscillations of cylindrical annuli about the rotation axis and propagate as Alfvén waves, with periods in the range 10–100 years. These waves cause torques acting on the mantle and yield time changes of the rotation period of the solid Earth that are also monitored. From a modeling of these waves, Zatman and Bloxham had initiated the inversion of the magnetic field threading the coaxial geostrophic cylinders in the core interior.

We give an account of existing theoretical models of torsional Alfvén waves within the Earth's core. We complement the previous studies that have considered standing waves only by investigating the propagation of torsional Alfvén waves as an initial value problem instead. Interestingly, the direction of propagation of the waves may indicate where dissipation takes place. In this respect, including a solid inner core in the model is crucial.

1. Introduction

The Coriolis force is a key ingredient of large scale dynamics of the Earth's fluid core, where the waves that are the more important in the dynamo process and have periods on the order of the magnetic diffusion time are strongly influenced by rotation. An exact balance between the Coriolis and pressure forces set apart the geostrophic motions, which are readily accelerated by Lorentz and other forces. They can form fast Alfvén waves, weakly influenced by magnetic diffusion. These transverse waves propagate in the directions perpendicular to the rotation axis. Their frequency is linearly dependent on the intensity of the

magnetic field transverse to the wave motions (Roberts, 1967). Their study, which was initiated by Braginsky (1970), has thus the potential to give precious information on the otherwise hidden magnetic field within the core (Zatman and Bloxham, 1998).

Models of the magnetic field at the Earth's surface can be downward continued to the surface of the metallic core because the solid mantle of the Earth is almost electrically insulating. At the Earth's surface, the declination of the magnetic field has been mapped for a few centuries (Jonkers *et al.*, 2003). It turns out that this is the appropriate timescale to study the evolution of the large scale (up to harmonic degree 3-4) and nondipolar parts of the magnetic field. Measurements at magnetic observatories have enabled to characterize changes in the Earth's magnetic field up to degree 6-8 for the last century (Langel, 1987). Finally, variations of the magnetic field up to degree 13 between 1980 and 2000 (Hulot *et al.*, 2002) and up to degree 11 during the last few years (Olsen, 2002) have been recently retrieved from satellite measurements. We note that the variation of the magnetic field of harmonic degree 13 between 1980 and 2000 at the Earth's surface is comparable to but still smaller than the total intensity of the magnetic field of this degree at either of the two epochs (Hulot *et al.*, 2002). It is now well established that the duration of magnetic field features is decreased in proportion to their lengthscale, as expected from a physical standpoint. Unfortunately, the rapid attenuation of small scale features of the Earth's magnetic field throughout the mantle will severely limit our ability to map it at the Core-Mantle Boundary (CMB).

In addition to these variations, sudden changes of the large scale part of the magnetic field have been detected in observatories. They occur simultaneously throughout large regions of the Earth's surface (Alexandrescu *et al.*, 1996). These events can be characterized as very rapid changes of the rate of variation of the Earth's magnetic field. It is customary to call them "jerks". The first global one to be detected occurred in 1969. It has been followed by similar events in 1978, 1991, and 1999. Their recurrence period is reminiscent of the solar cycle. However, the magnetic signal has predominantly an internal origin according to spherical harmonic analyses (Malin and Hodder, 1982) and the magnetic jerks were first detected after isolation of the solar-cycle related variation (Ducruix *et al.*, 1980). With one exception (Duhau and Martinez, 1995), studies of the internal part of the solar-cycle related variation have considered only magnetic induction in the solid mantle, leaving out processes in the liquid core. Bloxham *et al.* (2002) have just argued that the regularity of the occurrence of jerks during the last

thirty years may result from the propagation of Alfvén torsional waves within the Earth's core.

It is only recently that a host of geophysical applications followed up the initial study of Braginsky (1970). The trace of Alfvén waves has been searched in models of core surface flow \mathbf{u} , of typical speed a few 10^{-4} m.s^{-1} , derived from the Secular Variation (SV) of the Earth's magnetic field through the equation

$$\frac{\partial B_r}{\partial t} = -\nabla_H \cdot (\mathbf{u} B_r), \quad (1)$$

where B_r is the radial component of the magnetic field at the core-mantle boundary, inferred from models of the magnetic field at the Earth's surface, assuming an insulating mantle. This approach has been validated when it has been shown that changes in core angular momentum carried by geostrophic motions extracted from surface flow models balance fairly well-observed changes in the angular momentum of the solid Earth (Jault *et al.*, 1988; Jackson *et al.*, 1993; Hide *et al.*, 2000; Pais and Hulot, 2000). The basic tenet of these studies is that time-dependent zonal motions are geostrophic. The main difficulty is that zonal motions are not dominant at the core surface. It is compounded by the lack of models accounting for the non-zonal part of the motions. Recent maps of flow at the core surface show strong westward winds in the equatorial region of the Atlantic hemisphere and few motions in the Pacific hemisphere. After these early studies, Zatman and Bloxham (1997) noted that most of the zonal motions $u_\phi(\theta)$ (u_ϕ orthoradial component, θ colatitude) symmetrical about the equator of their velocity models from 1900 to 1990 can be fitted as the sum of two standing oscillations, with periods 76 and 53 years. Redoing the same analysis for the shorter epoch 1957–2001 with data better resolved in time, Bloxham *et al.* (2002) found three oscillatory motions, with periods 45, 20 and 13 years.

In the case of axisymmetrical solid boundaries, the torsional Alfvén equation amounts to an equation for the density of angular momentum within the core. Hide *et al.* (2000) argued, from the velocity models of Jackson *et al.* (1993), that there is propagation of core angular momentum density from the equatorial to polar regions. Equations of Alfvén waves, in an unbounded fluid, are not modified under time inversion. When dissipative terms at the boundaries are omitted, this remains true for the special waves studied here and no direction of propagation is privileged. The observation of Hide *et al.* (2000), if it is confirmed, thus gives an indication of the importance of dissipation for the evolution of

torsional Alfvén waves inside the core. The question of a favored direction of propagation has been left untouched by previous models as they were restricted to standing oscillations. In contrast, we choose here to set an initial value problem.

The geometry of the fluid outer core very much dictates the characteristics of these waves. Geostrophic motions \mathbf{u}_g obey the balance

$$2\rho(\Omega \times \mathbf{u}_g) = -\nabla p_g, \quad \mathbf{u}_g \cdot \mathbf{n} |_{\Sigma} = 0, \quad (2)$$

where \mathbf{n} is the outward normal to the boundary Σ of the fluid volume, and p_g the pressure. Geostrophic motions are independent of the coordinate z in the direction of the rotation axis, and are entirely defined by their streamlines on Σ , the pair of geostrophic contours Γ . Denote respectively z_T and z_B the z -coordinates along each upper and lower geostrophic contours. The length $H = z_T - z_B$ is an invariant of each pair of contours and the geostrophic cylinders \mathcal{C} are defined in a unique way by their total height H . The pressure p_g is constant on each cylinder \mathcal{C} , parallel to the rotation axis, generated by geostrophic contours. Neither the Coriolis force nor the pressure force enter directly the equation governing the time evolution of the geostrophic velocity. Indeed, integrating the momentum equation on \mathcal{C} to eliminate these forces, as suggested by Bell and Soward (1996), yields

$$\rho H \oint \left(\frac{\partial \mathbf{u}_g}{\partial t} + \frac{d\Omega}{dt} \mathbf{e}_z \times \mathbf{r} \right) \cdot d\Gamma = \int_0^H \left(\oint_{\Gamma(H,z)} (\mathbf{j} \times \mathbf{B} + \rho \mathbf{g}) \cdot d\Gamma \right) dz, \quad (3)$$

taking into account possible fluctuations of the spin rate Ω of the mantle and noting the position vector \mathbf{r} , the unit vector along the rotation axis \mathbf{e}_z , the geostrophic contour at the height z above the bottom boundary $\Gamma(H, z)$, the core density ρ , the magnetic field \mathbf{B} , the electric current density \mathbf{j} , and the buoyancy force $\rho \mathbf{g}$. We omit here both the viscous and the nonlinear terms. Assuming a quasi-static ambient magnetic field, we linearize the expression for the Lorentz force. In presence of the geostrophic motions, a magnetic field $\tilde{\mathbf{b}}$ is induced:

$$\frac{\partial \tilde{\mathbf{b}}}{\partial t} = \nabla \times (\mathbf{u}_g \times \mathbf{B}) \quad (4)$$

in the interior of the core, where the time changes of $\tilde{\mathbf{b}}$ are fast enough to make diffusion negligible. Equations (3) and (4), with $\tilde{\mathbf{j}} = \nabla \times \tilde{\mathbf{b}}/\mu_0$ and $\mathbf{j} \times \mathbf{B} \simeq \tilde{\mathbf{j}} \times \mathbf{B} + \mathbf{j} \times \tilde{\mathbf{b}}$, are analogous to the equations of Alfvén waves (μ_0 is the magnetic permeability).

In the case of an axisymmetrical container, the geostrophic contours are circular, the buoyancy term vanishes from (3), and the magnetic

term can be transformed into

$$\int_{-z_T}^{z_T} \oint (\mathbf{j} \times \mathbf{B})_\phi s \, d\phi \, dz = \frac{1}{s\mu_0} \frac{\partial}{\partial s} \left(s^2 \int_{-z_T}^{z_T} \oint B_s B_\phi s \, d\phi \, dz \right) + \frac{1}{\mu_0} \frac{dl}{ds} \left(\oint B_N B_\phi s \, d\phi(z_T) + \oint B_N B_\phi s \, d\phi(-z_T) \right), \quad (5)$$

where (s, ϕ, z) are cylindrical coordinates, dl is the element of length along the boundary in a meridional section, and B_N is the outward normal component of the magnetic field at the boundary. In writing (5), we have assumed that the container is also symmetrical with respect to the equatorial plane. In the axisymmetrical case, the angular momentum density is constant on each geostrophic cylinder. Multiplying (3) by the distance s to the rotation axis and integrating it in the fluid volume gives the equation governing the core angular momentum. In this transformation, the first term on the right-hand side of (5) vanishes and the last term gives the total electromagnetic torque acting on the core. It vanishes also if the mantle is electrically insulating. Then, there is conservation of the total core angular momentum.

In the next section, we follow the interpretation of the equation of torsional Alfvén waves within an entirely fluid spherical body in terms of transport of angular momentum. This explains well the amplification of the waves in the vicinity of the equator of the CMB. In the third section, we discuss the influence of possible bumps at the CMB. That enables us to stress that the description in terms of density of angular momentum and torques on geostrophic cylinders is convenient only in the spherical case. We consider also the pressure torque acting on the mantle throughout the propagation of Alfvén waves. The model is modified to include a solid inner core in a fourth section. At this stage, we have to reinstate the magnetic dissipation also. Finally, we list some works in progress as well as pending problems.

2. Propagation of Alfvén waves within an entirely fluid spherical body

The study of Alfvén waves in the spherical case was initiated by Braginsky (1970) based on (3)–(5). He showed how to eliminate $\tilde{\mathbf{b}}$ except for a surface term. The derivation has just been detailed by Jault (2003). A magnetic diffusion layer, set up to match the magnetic field induced in the core interior to the magnetic field in the mantle has to be taken into account. In a spherical core, the equation for Alfvén torsional

waves is

$$\begin{aligned}
s^3 z_T \frac{\partial^2}{\partial t^2} (\omega_g + \Omega) = & \frac{1}{\rho \mu_0} \frac{\partial}{\partial s} \left(z_T s^3 \frac{\partial \omega_g}{\partial s} \{B_s^2\} \right) \\
& - \frac{as^3}{4\pi\rho z_T} \frac{\partial \omega_g}{\partial t} \oint \left[(\sigma_m \Delta B_r^2)(z_T) + (\sigma_m \Delta B_r^2)(-z_T) \right] d\phi \\
& + \frac{as^2}{4\pi\rho\mu_0 z_T} \oint \left[\left(B_r \frac{\partial \tilde{b}_{m\phi}}{\partial t} + B_\phi \frac{\partial \tilde{b}_{mr}}{\partial t} \right) (z_T) \right. \\
& \left. + \left(B_r \frac{\partial \tilde{b}_{m\phi}}{\partial t} + B_\phi \frac{\partial \tilde{b}_{mr}}{\partial t} \right) (-z_T) \right] d\phi,
\end{aligned} \tag{6}$$

where (r, θ, ϕ) are spherical coordinates, a is the core radius (whence $z_T = (a^2 - s^2)^{1/2}$), \mathbf{e}_ϕ is the unit azimuthal vector, $\mathbf{u}_g = s\omega_g(s)\mathbf{e}_\phi$, Δ is the variable thickness of a hypothetical thin layer of electrical conductivity $\sigma_m(\theta, \phi)$ at the bottom of the mantle, B_r is the magnetic field component normal to the core surface, $\tilde{\mathbf{b}}_m$ is the magnetic field at the bottom of the insulating volume inside the mantle, and $\{B_s^2\}$ is a measure of the s-component of the magnetic field averaged on each geostrophic cylinder:

$$\{B_s^2\}(s) = \frac{1}{4\pi z_T} \int_{-z_T}^{z_T} \oint B_s^2 d\phi dz. \tag{7}$$

Even though an expression of $\partial \tilde{b}_{mr}/\partial t|_{r=a}$ as a function of ω_g is directly obtained from (4), the determination of $\partial \tilde{b}_{m\phi}/\partial t|_{r=a}$ necessitates an integration over the entire core surface. It can be achieved through the calculation of the potential V , such that

$$\tilde{\mathbf{b}}_m = -\nabla V, \tag{8}$$

which is derived from \tilde{b}_{mr} . In the case of axisymmetrical \mathbf{B} (and thus $\tilde{\mathbf{b}}$), there is no poloidal ingredient in \tilde{b}_ϕ and $\tilde{b}_{m\phi}$ vanishes. In the non-axisymmetrical case, however, the last term of (6) makes the propagation mechanism non-local. Time changes of the mantle spin-rate $d\Omega/dt$, which are linearly related to the time changes of core angular momentum carried by geostrophic motions, yield another non-local coupling mechanism between the motions of the different geostrophic cylinders. Indeed, we cannot simply transform (6) into an equation for $\omega_g^* = (\omega_g + \Omega)$ because of coupling with the mantle (the second term on the right-hand side). Equation (6) is finally completed by two boundary conditions,

$$\frac{\partial \omega_g}{\partial s} = -\mu_0 \left(\oint_{\theta=\pi/2} B_r^2 d\phi \right)^{-1} \left(\oint_{\theta=\pi/2} \sigma_m \Delta B_r^2 d\phi \right) \frac{\partial \omega_g}{\partial t} \quad \text{at } s = a,$$

which is required to avoid a singularity at ($s = a$) when $\oint B_r^2 d\phi \neq 0$ at the equator (Buffett, 1998), and

$$\frac{\partial \omega_g}{\partial s} = 0 \quad \text{at} \quad s = 0, \quad (9)$$

in the case of non-axisymmetrical **B**. The mechanism of propagation is described by the first two terms of (6). As the wave propagates, there is transport of angular momentum density $z_T s^3 \omega_g$. At the edges of the domain, where either s or z_T vanishes, we expect amplification of ω_g . In the geophysical case, this does not happen near the rotation axis because of the very efficient coupling between the solid and electrically conducting inner core and the wave motions. On the other hand, Fig. 2 of Zatman and Bloxham (1997) indicates amplification of the geostrophic motions at the equator $z_T = 0$. Arguably, their representation minimizes the actual intensification at the equator because of the regularization conditions that are required for the modeling.

We set an initial-value problem. Equation (6) can be transformed into a set of equations that are readily solved numerically:

$$\begin{aligned} & s^3 z_T \frac{\partial}{\partial t} (\omega_g + \Omega) \\ &= \frac{1}{\rho \mu_0} \frac{\partial \tau}{\partial s} - \frac{as^3}{4\pi\rho z_T} \left(\oint \sigma_m \Delta B_r^2 d\phi(s, z_T) + \oint \sigma_m \Delta B_r^2 d\phi(s, -z_T) \right) \omega_g \\ &+ \frac{as^2}{4\pi\rho\mu_0 z_T} \oint [(B_r \tilde{b}_{m\phi} + B_\phi \tilde{b}_{mr})(s, z_T) \\ &+ (B_r \tilde{b}_{m\phi} + B_\phi \tilde{b}_{mr})(s, -z_T)] d\phi, \end{aligned} \quad (10)$$

$$\frac{\partial \tau}{\partial t} = z_T s^3 \frac{\partial \omega_g}{\partial s} \{B_s^2\}, \quad (11)$$

where τ is an auxiliary variable. This is completed by equations for \tilde{b}_m at the boundary (see (8)) and for Ω . As the actual magnetic field in the core interior and the distribution of $\{B_s^2\}$ are unknown, Buffett (1998) based a forward modeling of torsional waves on maps of magnetic field inferred from a numerical geodynamo model. We build a three-dimensional map of the interior magnetic field from models of the large scale magnetic field at the CMB, seeking solutions that have the geometry of free decay modes and neglecting entirely the toroidal component. This is acceptable since our calculation has only an illustrative purpose. The root mean square $\{B_s^2\}^{1/2}(s)$, shown on Fig. 1 (full curve), for this ambient magnetic field can be compared with the models of Zatman and Bloxham (1998) and Buffett (1998). At $t = 0$, we set

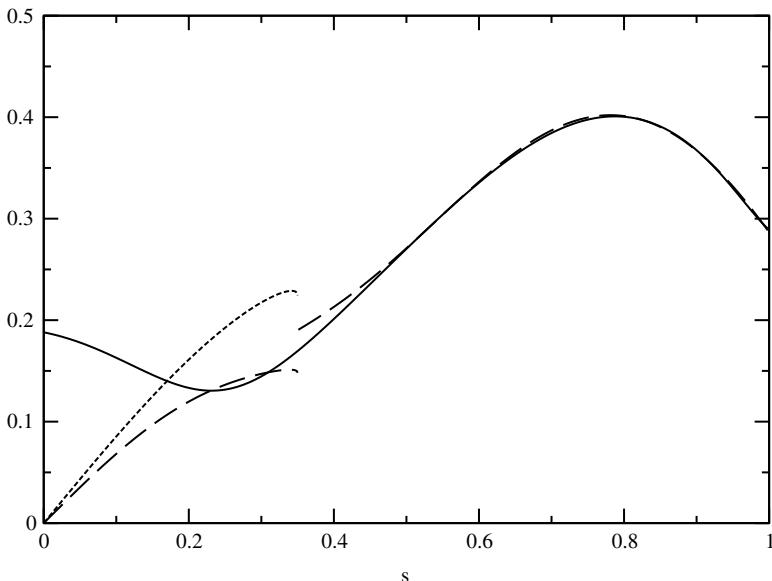


Figure 1. Cylindrical average $\{B_s^2\}^{1/2}$ in mT units respectively for the three-dimensional quasi-static magnetic field of Section 2 (full curve) and the axisymmetrical field of Section 4 (dashed curve). In the latter case, and for $s \leq b$, the dashed and dotted curves show, respectively, $\{B_s^2\}^{1/2}$ in the Northern and Southern Hemispheres.

$\tau = 0$ and $\mathbf{b}_m = \mathbf{0}$. Finally, we make a variety of choices for the initial value of ω_g . We use conservation of total energy, kinetic and magnetic as a diagnostic, when the mantle is electrically insulating. Indeed, the last portions of our runs have to be discarded as waves with ever decreasing lengthscales pile up near the rotation axis. Before this final stage, there is amplification of the motions in the equatorial region, as expected. Neither positive nor negative s -direction of propagation is favored. We find very similar results whether we include the last surface term of (6) or not. Finally, we do not dwell on the most conspicuous feature, which is the magnification of the waves as they get nearer to the axis since it is drastically modified in presence of a solid and electrically conducting inner core.

3. Topographical effects at the CMB

It is estimated from nutation models and nutation series derived from very long baseline interferometry that the ellipticity of the fluid outer

core exceeds its hydrostatic equilibrium value by 3.8% (Mathews *et al.*, 2002). That corresponds to an excess of the equatorial radius by 350 m. This value gives us a lower estimate of the height of the corrugations at the CMB. Many seismological studies have been devoted to the estimation of the CMB topography amplitude. Recently, Garcia and Souriau (2000) concluded that when averaging over regions of 20° lateral extent, the topography amplitude is in the range $\sim \pm 1.5$ km. They found also larger amplitude when averaging over smaller regions. Sze and Van der Hilst (2003) concurred but argued that the currently available seismic data can resolve CMB topography at least in some regions. They found 3 km peak-to-peak undulations.

In presence of topography at the CMB or at the inner core boundary (ICB), geostrophic contours deviate from circles. Then, both the magnetic and buoyancy forces contribute to the geostrophic acceleration because the component of the buoyancy force parallel to the geostrophic contours does not vanish. Propagation of torsional waves comes with a pressure torque acting at the CMB. This net torque acting on the mantle is nonzero and may play a role in the changes of the Earth's rate of rotation over $10-10^2$ years. This is what motivated the inclusion of CMB topography in models of torsional waves propagation (Buffett, 1998) as the origin of core-mantle coupling is not yet elucidated. Neglecting Lorentz forces at the top of the core, the non-hydrostatic pressure can be estimated at the core surface from secular variation data (Chulliat and Hulot, 2000). It is of the order of $\pm 10^3$ Pa. peak-to-peak. Then, a typical magnitude of the pressure torque acting at the CMB of area Σ on the solid mantle is $hp\Sigma$ (10^{20} N.m.), which is 10^2 larger than the actual torque between core and mantle inferred from LOD variations. On the other hand, the typical pressure B^2/μ_0 associated with magnetic fields like the one considered in the previous section is only 0.1 Pa. More important pressures are obtained when one assumes that there is a large zonal toroidal magnetic field $B_\phi(s, z)\mathbf{e}_\phi$ in the core interior, which is invisible at the core surface and is parallel to the geostrophic cylinders in spherical geometry. However, Anufriev and Braginsky (1977) found that the magnitude of the pressure torque scales then only as $h^2 B_\phi^2 \Sigma / \mu_0 a$ because the action of the zonal pressure cancels out when integrated on the core surface. More general studies are needed to decide whether the magnetic field in the core interior can be strong enough to make the pressure field carried by the torsional oscillations significant for Earth rotation studies without making the Alfvén timescale too rapid. Finally, other mechanisms, such as the transport of density anomalies by the torsional waves, can also be contemplated (Jault *et al.*, 1996).

From a theoretical viewpoint, the interesting question is whether the pressure at the CMB can provide the restoring torque for differential rotation between the core and the mantle in the same way as the magnetic field provides the restoring force for torsional oscillations in the core interior. The work of Kuang and Chao (2001, 2003) gives some indication that the pressure torque may play a role of this kind. They have recently incorporated topographic coupling in a self-consistent convective dynamo model. They account for the topography by modifying the boundary conditions for the velocity and magnetic fields. That ensures that the geostrophic contours are distorted as the result of the modified no-penetration condition. They find a large cancelation in the integral that gives the pressure torque at the core surface.

A discussion of the topographical effects at the CMB starts with (3). We lump together the magnetic and buoyancy forces as a force \mathbf{F} . The circulation of the force \mathbf{F} along a geostrophic contour $\Gamma(H, z)$ is written

$$\oint \mathbf{F} \cdot d\Gamma = \oint F_\phi \mathbf{e}_\phi \cdot d\Gamma + \oint (F_s \mathbf{e}_s + F_z \mathbf{e}_z) \cdot d\Gamma. \quad (12)$$

Assuming that the corrugations are smooth and of height h small compared with the core radius a , we use a perturbation approach ($\varepsilon = h/a$). We first note that, beyond the order $O(\varepsilon)$, the interpretation of (3) as a torque balance is not correct, as already noted by Fearn and Proctor (1992) in the course of a derivation of the Taylor's condition. The angular momentum density is not uniform on the geostrophic cylinders $\mathcal{C}(H)$. The distance to the rotation axis from $\mathcal{C}(H)$ varies and the geostrophic velocity changes along a contour because the thickness of the annulus $A(H)$ enclosed between two geostrophic cylinders $\mathcal{C}(H)$ and $\mathcal{C}(H + dH)$ is not uniform. As a result, (3) cannot be transformed into a torque budget at the $O(\varepsilon^2)$ order. Equation (3) governs the kinetic energy of the geostrophic motions rather than the angular momentum density since one way to obtain it is to take the dot product of any geostrophic-like vector field with the momentum equation. In conclusion, on the one hand, it is necessary to invoke the pressure torque to write the equation which gives the differential rotation between core and mantle but, on the other hand, it is not natural, and even not correct beyond the order $O(\varepsilon)$, to interpret the equation governing the time evolution of the geostrophic velocity as a torque budget.

Returning now to the order $O(\varepsilon)$, we would like to explain how (6) is modified when the geostrophic contours are distorted. With this aim in view, it is helpful to define a spherical reference state and to transform the contours $\Gamma(H, z)$ into circles. Anufriyev and Braginsky (1977) suggested also to study separately the effects of topographies

respectively symmetrical and antisymmetrical with respect to the equatorial plane. However, we are not aware that such a project has been achieved. Previous studies inspired by angular momentum considerations have included only a pressure term at the core surface to take into account topographical effects. A discussion of the symmetrical case shows that it is not satisfactory. Then, the geostrophic contours are distorted in the s -direction by an amount $\varepsilon s_1(s, \phi)$. Consider again the annulus $A(H)$ enclosed between two geostrophic cylinders $\mathcal{C}(H)$ and $\mathcal{C}(H + dH)$. There are pressure torques acting on the two interior sides $\mathcal{C}(H)$ and $\mathcal{C}(H + dH)$ of $A(H)$ and also an $O(\varepsilon)$ perturbation of the torque exerted by the force $F_\phi(s, \phi, z) \mathbf{e}_\phi$ on $A(H)$ because of both the variable distance to the axis along a contour and the variable thickness of $A(H)$. Thus, the action of the pressure at the top and the bottom of the annulus cannot be considered in isolation and knowing the pressure at the boundary does not suffice to calculate the influence of the boundary topography on the geostrophic acceleration.

4. Influence of a conducting inner core

In the special case of a spherical cavity enclosed between two spheres of radius respectively b and a , the geostrophic contours are circular and the geostrophic velocity is constant on each cylinder \mathcal{C} . However, inside the region delimited by the cylindrical surface Σ_T ($s = b$) tangent to the inner core, the cylindrical average $\{B_s^2\}$ calculated above and below the inner core differs because the ambient magnetic field \mathbf{B} is not symmetrical with respect to the equatorial plane. In the Northern and Southern hemispheres, we have, respectively,

$$\begin{aligned} \{B_s^2\}^+(s) &= \frac{1}{2\pi H} \int_{z_B}^{z_T} \oint B_s^2 d\phi dz, \quad \text{and} \\ \{B_s^2\}^-(s) &= \frac{1}{2\pi H} \int_{-z_T}^{-z_B} \oint B_s^2 d\phi dz \end{aligned} \quad (13)$$

with $z_T = \sqrt{a^2 - s^2}$, $z_B = \sqrt{b^2 - s^2}$ and $H = z_T - z_B$. As a result, the geostrophic velocity splits up also:

$$\mathbf{u}_g = s\omega_g^\pm(s)\mathbf{e}_\phi \quad \text{at } s \leq b \quad \text{and} \quad \pm z \geq 0. \quad (14)$$

There is first a weak singularity at ($s = b^-$) because the height of the geostrophic cylinder ($z_T - z_B$) changes rapidly as s approaches b from below. However, the main singularity arises because (3) changes from

one side to the other side of Σ_T . A discontinuity in the geostrophic velocity at Σ_T cannot be ruled out because viscosity is neglected in the fluid interior. Then, electrical currents set up at the cylinder Σ_T tangent to the inner core (and at the boundary with the electrically conducting inner core as well):

$$\eta \left(\frac{\partial b_\phi}{\partial s} \Big|_{s=b^+} - \frac{\partial b_\phi}{\partial s} \Big|_{s=b^-} \right) = -b(\omega_g(b^+) - \omega_g(b^-))B_s, \quad (15)$$

$$\eta \left(\frac{\partial b_\phi}{\partial r} \Big|_{r=b^+} - \frac{\partial b_\phi}{\partial r} \Big|_{r=b^-} \right) = -s(\omega_g(s) - \omega_{IC})B_r, \quad (16)$$

where ω_{IC} is the inner core rotation, and $\eta = 1/\mu_0\sigma$ is magnetic diffusivity. Equation (16) is written assuming equal conductivity σ for the solid and liquid parts of our core model.

The treatment of the inner core in previous studies was quite different from ours as equatorial symmetry was assumed. Braginsky (1970) argued that the magnetic coupling between the solid and fluid cores is so strong that $\omega_g(s) = \omega_{IC}$ for $(s \leq b)$. He solved (6) only for $(s \geq b)$ with the boundary condition $\omega_g(b) = \omega_{IC}$. As Zatman and Bloxham (1997) omitted the inner core in their model, they studied the propagation of torsional waves only for $(s \geq b)$ also. Finally, Buffett (1998) imposed the continuity of ω_g through $s = b$ and solved an equation analogous to (6) separately for $(s < b)$ and $(s > b)$ while Mound and Buffett (2003) assumed that, within Σ_T , ω_g is uniform but not identical to ω_{IC} . The latter authors estimated the evolution of $\omega_g(b)$ from the total electromagnetic torque acting on the region within Σ_T , which was inferred from $d\omega_g/ds|_{s=b^+}$.

If either the inner core surface or the CMB are not perfectly axisymmetrical, there are fluid lenses void of closed geostrophic contours in the vicinity of $s = b$ (Herrmann and Busse, 1998). There, geostrophic motions may be replaced by Rossby waves, low frequency z-independent inertial waves. They have a natural frequency comparable to the frequency of the torsional Alfvén waves (Jault, 2003). Their possible role in the course of the propagation of Alfvén waves within the Earth's core has not yet been investigated.

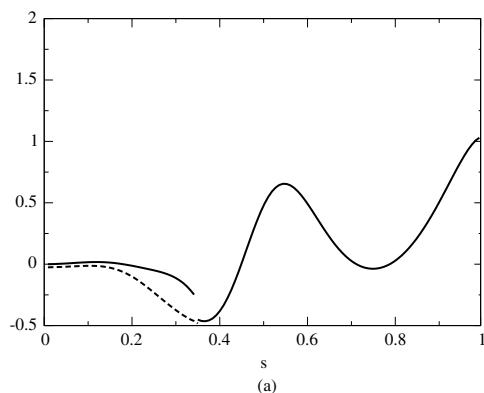
In order to give an illustrative example, we set again an initial value problem restricted, however, to axisymmetrical variables. We build an axisymmetrical model such that its distribution of $\{B_s^2\}^{1/2}$ (Fig. 1, dashed curve) resembles that calculated in Section 2 from a model of the geomagnetic field at the core surface (Fig. 1, full curve) outside Σ_T . For $(s \leq b)$, $\{B_s^2\}^+$ in the Northern Hemisphere differs from $\{B_s^2\}^-$ in the Southern Hemisphere. We assume a model for the

geostrophic velocity ω_g at $t = 0$. A zonal toroidal magnetic field b_ϕ is induced. In order to study the evolution of the two scalar variables ω_g and b_ϕ , we use a finite difference code with both spherical and cylindrical grids. The two grids coincide on Σ_T so that (15) can be easily enforced (Jault, 1996). We calculate separately ω_g at $s = b^-$ and $s = b^+$. In addition, for $s < b$, the geostrophic angular velocity ω_g^\pm can differ below and above the inner core. Finally, we reinstate magnetic diffusion because, in the vicinity of the tangent cylinder Σ_T and of the inner core surface, the large gradients of b_ϕ make it important.

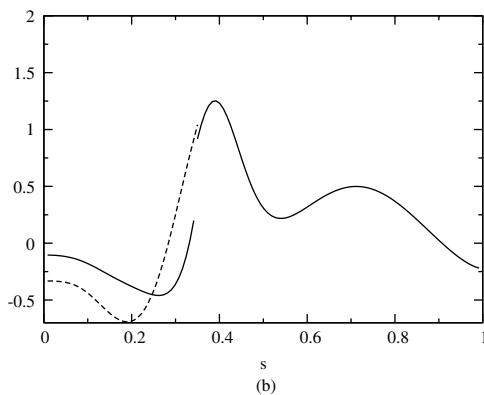
We have studied a few initial states for the geostrophic velocity. The ratio between the Alfvén wave period and the magnetic dissipation time of the solid inner core depends on the magnetic Ekman number $E_M = \eta/\Omega a^2$, on the cylindrical average $\{B_s^2\}^+(s)$ and $\{B_s^2\}^-(s)$ for ($s \leq b$), and on the spherical average $(\int \int_{r=b} B_r^2 d\Sigma)/\Sigma_b$ of the radial magnetic field at the inner-core boundary of area Σ_b . The value of $[(\int \int_{r=b} B_r^2 d\Sigma)/\Sigma_b]^{1/2}$ for our model of quasi-static magnetic field (0.4 mT) is similar to $\{B_s^2\}^{1/2}$. We have investigated a range of values of E_M down to 10^{-9} , which is the geophysically relevant value according to the conductivity estimates of Secco and Schloessin (1989). We have chosen three different epochs (respectively, $t = 6.7$ yr, $t = 14.3$ yr, and $t = 17.7$ yr) to illustrate the propagation of torsional waves in presence of an inner core (Fig. 2). We find that the discontinuities in ω_g at Σ_T are kept small by the Lorentz force coupling the geostrophic cylinders across Σ_T . In our example, the latter are much more efficient in the Southern Hemisphere because $(\{B_s^2\}^+(b) < \{B_s^2\}^-(b))$. After a short transient period, the torsional waves mainly propagate inward. We attribute that to magnetic dissipation within the inner core. The waves penetrate within the tangent cylinder and are damped on a distance comparable to the inner core radius. However, the coupling between the solid inner core and the geostrophic motions remains strong in our model. There are large oscillations of the inner core, which are apparent in the time evolution of the geostrophic motions within Σ_T .

5. Concluding remarks

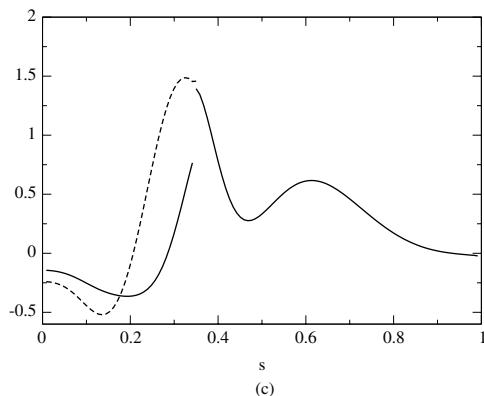
We have argued that a preferred direction of propagation, as observed by Hide *et al.* (2000) is potentially a useful constraint on the dissipation of the torsional waves, especially at the ICB. We have attempted a forward modeling of the torsional Alfvén waves without the often made assumption that they can be represented as standing oscillations. We have been able to point up sharp velocity gradients at the cylinder Σ_T



(a)



(b)



(c)

Figure 2. Snapshots of the geostrophic angular velocity at (a) $t = 6.7$ yr; (b) $t = 14.3$ yr; (c) $t = 17.7$ yr. For $s \leq b$, the full and dashed curves show respectively the geostrophic velocity in the Northern and Southern Hemisphere. $Ro = 3.10^{-8}$.

tangent to the inner core as we have not assumed equatorial symmetry of \mathbf{B} in our study of the role of the inner core. We intend now to undertake an inverse modeling. In order to recover possible sharp velocity gradients at Σ_T , we will determine directly the torsional oscillations in the Earth's fluid core from geomagnetic data. Bypassing the construction of surface flow models is indeed helpful as small scale features, such as large gradients in a narrow zone, cannot be recovered from the calculation of general core surface flows. Hopefully, the recording of large geomagnetic data sets by the Ørsted, Champ, and following satellites will promote yet other theoretical progresses.

Modeling Earth's nutation, Mathews *et al.* (2002) estimated that the rms radial magnetic field at the ICB is 7.2 mT. This finding, which required several hypotheses, does not agree well with the models of the magnetic field in the core interior that we have used. It raises a question that we have already touched in the section on topographic effects. In our models, the strength of the magnetic field is comparable at the core surface and in the interior whereas we would have expected it to be larger in the interior from considerations of the speed of core motions.

We have just sketched a discussion of topographic coupling. We especially expect from a study of torsional waves propagation in a core model including corrugated boundaries to learn whether pressure effects tend to restore corotation between core and mantle or whether other mechanisms (viscous, electromagnetic) have to be taken into account at the CMB.

Acknowledgments

This contribution is dedicated to the memory of Stephen Zatman who, together with Professor Jeremy Bloxham, rejuvenated the study of torsional Alfvén waves inside the Earth's core before his untimely death.

DJ thanks gratefully Professor Fritz Busse who suggested that torsional waves arriving at the tangent cylinder or at the CMB equator may cause sudden changes in the geomagnetic field.

References

- Alexandrescu, M., Gibert, D., Hulot, G., Le Mouël, J.-L. and Saracco, G., "World-wide wavelet analysis of geomagnetic jerks," *J. Geophys. Res.* **101**, 21975–21994 (1996).

- Anufriyev, A. P. and Braginsky, S. I., “Effects of irregularities of the boundary of the Earth’s core on the speed of the fluid and on the magnetic field. iii,” *Geomag. Aeron.* **17**, 492–496 (1977).
- Bell, P. I., and Soward, A. M., “The influence of surface topography on rotating convection,” *J. Fluid Mech.* **313**, 147–180 (1996).
- Bloxham, J., Zatman, S. and Dumberry, M., “The origin of geomagnetic jerks,” *Nature* **420**, 65–68 (2002).
- Braginsky, S. I., “Torsional magnetohydrodynamic vibrations in the Earth’s core and variations in day length,” *Geomag. Aeron.* **10**, 1–8, (1970).
- Buffett, B. A., “Free oscillations in the length of day: inferences on physical properties near the core-mantle boundary”, in: *The Core-Mantle Boundary Region* (Eds. M., Gurnis, M. E., Wysession, E. Knittle, and B. A. Buffett) vol. **28** of *Geodynamics Series*, pp. 153–165, AGU (1998).
- Chulliat, A. and Hulot, G., “Local computation of the geostrophic pressure at the top of the core,” *Phys. Earth Planet. Inter.* **117**, 309–328 (2000).
- Ducruix, J., Courtillot, V. and Le Mouél, J.-L., “The late 1960s secular variation impulse, the eleven year magnetic variation and the electrical conductivity of the deep mantle,” *Geophys. J. R. Astron. Soc.* **61**, 73–94 (1980).
- Duhau, S. and Martinez, E. A., “On the origin of the fluctuations in the length of day and in the geomagnetic field on a decadal time scale,” *Geophys. Res. Lett.* **22**, 3283–3286 (1995).
- Fearn, D. R. and Proctor, M. R. E., “Magnetostrophic balance in non-axisymmetric, non-standard dynamo models,” *Geophys. Astrophys. Fluid Dynam.* **67**, 117–128 (1992).
- Garcia, R. and Souriau, A., “Amplitude of the core-mantle boundary topography estimated by stochastic analyses of core phases,” *Phys. Earth Planet. Inter.* **117**, 345–359 (2000).
- Herrmann, J. and Busse, F. H., “Stationary and time dependent convection in the rotating cylindrical annulus with modulated height,” *Phys. Fluids* **10**, 1611–1620 (1998).
- Hide, R., Boggs, D. H. and Dickey, J. O., “Angular momentum fluctuations within the Earth’s liquid core and torsional oscillations of the core-mantle system,” *Geophys. J. Int.* **143**, 777–786 (2000).
- Hulot, G., Eymin, C., Langlais, B., Mandea, M. and Olsen, N., “Small-scale structure of the geodynamo inferred from Oersted and Magsat satellite data,” *Nature* **416**, 620–623 (2002).
- Jackson, A., Bloxham, J. and Gubbins, D., “Time-dependent flow at the core surface and conservation of angular momentum in the coupled core-mantle system,” in: *Dynamics of the Earth’s Deep Interior and Earth Rotation* (Eds. J. L. Le Mouél, D. E. Smylie, and T. Herring), pp. 97–107, AGU/IUGG, (1993).
- Jault, D., “Sur l’inhibition de la régénération du champ magnétique dans certains modèles de dynamo planétaire en présence d’une graine solide,” *C. R. Acad. Sci. Paris, série II a* **323**, 451–458 (1996).
- Jault, D., “Electromagnetic and topographic coupling, and LOD variations,” in: *Earth’s Core and Lower Mantle*, (Eds. C. A. Jones, A. M. Soward, and

- K. Zhang), *The Fluid Mechanics of Geophysics and Astrophysics*, vol. **11**, pp. 56–76, Taylor and Francis (2003).
- Jault, D., Gire, C. and Le Mouël, J.-L., “Westward drift, core motions and exchanges of angular momentum between core and mantle,” *Nature* **333**, 353–356 (1988).
- Jault, D., Hulot, G. and Le Mouël, J.-L., “Mechanical core-mantle coupling and dynamo modelling,” *Phys. Earth Planet. Inter.* **43**, 187–191 (1996).
- Jonkers, A. R. T., Jackson, A. and Murray, A., “Four centuries of geomagnetic data from historical records,” *Rev. Geophys.* **41** (1006): 10.1029/2002RG000115 (2003).
- Kuang, W. and Chao, B. F., “Topographic core-mantle coupling in geodynamo modeling,” *Geophys. Res. Lett.* **28**, 1871–1874 (2001).
- Kuang, W. and Chao, B. F., “Geodynamo modeling and core-mantle interactions,” in: *Earth’s Core Dynamics, Structure, Rotation* (Eds. V. Dehant, K. C. Creager, S.-I. Karatu, and S. Zatman), *Geodynamics Series*, pp. 193–212, AGU (2003).
- Langel, R. A., “The main field,” in *Geomagnetism, Vol. 1* (Ed. J. A. Jacobs), pp. 249–512, Academic Press (1987).
- Malin, S. R. C. and Hodder, B. M., “Was the 1970 geomagnetic jerk of internal origin?”, *Nature* **296**, 726–728 (1982).
- Mathews, P. M., Herring, T. A. and Buffett, B. A., “Modeling of nutation-precession: New nutation series for nonrigid Earth, and insights into the Earth’s interior,” *J. Geophys. Res.* **107**(B4), doi:10.1029/2001JB000390 (2002).
- Mound, J. E. and Buffett, B. A., “Interannual oscillations in length of day: Implications for the structure of the mantle and core,” *J. Geophys. Res.* **108** (B7), 10.1029/2002JB002054 (2003).
- Olsen, N., “A model of the geomagnetic field and its secular variation for epoch 2000 estimated from Ørsted data,” *Geophys. J. Int.* **149**, 454–462 (2002).
- Pais, A. and Hulot, G., “Length of day decade variations, torsional oscillations and inner core superrotation: Evidence from recovered core surface zonal flows,” *Phys. Earth Planet. Inter.* **118**, 291–316 (2000).
- Roberts, P. H. *An Introduction to Magnetohydrodynamics*, Elsevier, New York (1967).
- Secco, R. A. and Schloessin, H. H., “The electrical resistivity of solid and liquid Fe at pressures up to 7 GPa,” *J. Geophys. Res.* **94**, 5887–5894 (1989).
- Sze, E. K. M. and Van der Hilst, R. D., “Core mantle boundary topography from short period PcP, PKP, and PKKP data,” *Phys. Earth Planet. Inter.* **135**, 27–46 (2003).
- Zatman, S. and Bloxham, J., “Torsional oscillations and the magnetic field within the Earth’s core,” *Nature* **388**, 760–763 (1997).
- Zatman, S. and Bloxham, J., “A one-dimensional map of B_s from torsional oscillations of the Earth’s core,” in: *The Core-Mantle Boundary Region* (Eds. M. Gurnis, M. E. Wysession, E. Knittle, and B. A. Buffett) vol. **28**, *Geodynamics Series*, pp. 183–196, AGU (1998).

10 Turbulence models and plane layer dynamos

C.A. Jones

*Department of Mathematical Sciences, University of Exeter,
North Park Road, Exeter EX4 4QE, UK*

P.H. Roberts

*Institute of Geophysics and Planetary Physics University
of California, Los Angeles, CA 90095, USA*

The parameter ranges in which naturally occurring dynamos operate lie far beyond those that are numerically accessible. Small-scale motions that significantly influence the dynamo mechanism cannot be numerically resolved. Two approximate ways that have been suggested for incorporating these “sub-grid scales” are investigated here: the addition of hyperdiffusion and the application of the LANS alpha theory. These proposals were tested by comparing their consequences, not only for marginal convection but also for finite amplitude dynamos, with corresponding “exact” results. Plane layer models were studied because of their numerical advantages.

Both approximate theories raise the differential order of the governing equation so that additional boundary conditions are required. Physical considerations reduce the options but arbitrariness remains. Our choice was the simplest and made it possible to preserve the turbulent character of the solutions while increasing the time step by a factor of about 10, compared with that required by the exact solution at the same spatial resolution. It was found that, for optimal results, the hyperdiffusivities for velocity, momentum and magnetic field should be equal. In the case of the LANS alpha approximation, a further choice of boundary conditions was investigated. In this model, alpha represents the rms fluid displacement from the mean turbulent motion, both of which vanish at the no-slip boundaries. Analysis of the resulting depression of alpha motivated the alternative conditions that led to even more satisfactory solutions. The question of the best choice of boundary conditions remains open, but is a prime target for future research that uses these methods.

1. Introduction

Substantial progress has been made in modeling the geodynamo in recent years, with Earthlike magnetic fields emerging from convection-driven simulations of the liquid outer core of the Earth. These models are based on self-consistent numerical solutions of the Navier-Stokes equation and the magnetic induction equation. This work has been reviewed recently by Kono and Roberts (2002) and from a more observational standpoint by Dormy *et al.* (2000). There is, however, a disturbing feature of these simulations; the dimensionless parameters used are very different from those suggested by the molecular diffusion constants relevant to the outer core. For example, the Ekman number $E = \nu/2\Omega d^2$, where ν is the kinematic viscosity, Ω is the Earth's angular velocity and d is the outer core radius, is about 10^{-15} in the core, but values of order 10^{-4} are typically used in simulations. Similarly, the heat flux Rayleigh number, which is defined as $Ra = Fg\bar{\alpha}d^2/2\Omega\kappa^2\rho c_p$ where F is the heat flux per unit area emerging from the core, g is gravity, $\bar{\alpha}$ is the thermal expansion coefficient, κ is the thermal diffusivity, ρ is the density and c_p is the specific heat, is approximately 10^{15} , whereas values between the order $10^2 - 10^3$ are used in simulations. The Prandtl number $P_r = \nu/\kappa$ in the core is about 0.1, and the magnetic Prandtl number $P_m = \nu/\eta$ is estimated to be about 10^{-6} , where η is the magnetic diffusivity, but in simulations O(1) values of P_m are assumed. Indeed, no successful simulations have been performed with $q = P_m/P_r$ small; even using such modest values as $q \sim 0.1$ is currently out of reach (Christensen *et al.*, 1999). It is therefore rather surprising that geodynamo simulations give results that apparently agree well with observation.

The scaling arguments of Starchenko and Jones (2002) shed some light on this paradox. They argue that the MAC balance is the relevant one for the core, that is Coriolis, pressure, buoyancy and Lorentz forces are all of similar magnitudes. In the limit of small diffusion, this leads to an estimate for the typical velocity $u_* \sim (g\bar{\alpha}F/2\Omega\rho c_p)^{1/2}$ that is independent of all diffusivities, and has the right order of magnitude when compared with the velocities inferred from the secular variation. A similar criterion exists for compositional convection. In consequence, provided a geodynamo simulation uses the right heat flux, and provided it is in the correct asymptotic MAC regime, a reasonable magnitude for the velocity will emerge even if the thermal and viscous diffusivities are completely incorrect.

Using a plane layer dynamo model, Rotvig and Jones (2002) showed that at moderate Rayleigh number the MAC regime (as evidenced by

the progressive satisfying of Taylor's (1963) constraint) is attained for $10^{-4} < E < 10^{-5}$ at moderate Rayleigh number. This suggests that current spherical geometry geodynamo models are on the verge of entering the MAC regime, but to get fully into this regime in spherical geometry is still computationally very challenging. Even if this difficulty is overcome, a further hurdle lies ahead, that raised by the large Rayleigh number (low thermal diffusivity) of core flow. The Earth's core and all experimental dynamos are liquid metal systems in which the magnetic diffusivity is much greater than the thermal diffusivity. This is currently a very difficult numerical regime for dynamo codes, and so the issue of whether the MAC balance regime is achieved in the limit of small κ is still largely unexplored.

A wide variety of numerical methods have been employed to attack the dynamo problem, the pseudo-spectral method, finite differences and finite elements all being used. Spherical harmonic expansion in the ϑ and ϕ directions, (r, ϑ, ϕ) being spherical polar coordinates, and finite differences in r are a popular choice. All these methods have a common difficulty, which is intrinsic to the problem, that of the appearance of waves on many different timescales. This is a consequence of the large number of different forces involved in the dynamical balance. Unfortunately, this obstacle has yet to be removed, so that little further progress has been made since the initial successes in the mid nineties. The aim of this article is to analyze the difficulties and to identify promising directions for future simulations.

Since the difficulties are clearly associated with the presence of small diffusion terms, it is natural to focus on the modeling of these processes in the Earth's core. It is of course true that similar problems occur in all modeling of problems where the diffusive processes are small in dimensionless terms such as, for example, the turbulent flow in wind-tunnels. Nevertheless, there are very particular difficulties which arise in the dynamo problem over and above those of high Reynolds number non-magnetic, non-rotating flows. It is arguable that to represent the geodynamo adequately, we need to be able to resolve scales on which the magnetic Reynolds number is around unity. Since the large scale magnetic Reynolds number in the core is only a few hundred, this would not appear to be an impossible challenge, but numerical difficulties have so far prevented this goal from being attained.

The difficulties found in spherical systems appear to be similar to those found in plane layer geometry, but planar models are much better suited to experimentation, as the numerical codes are much simpler. These codes also run faster because of the availability of practical fast

Fourier transforms in all three spatial dimensions. Plane layer geometry therefore provides an excellent test bed for new numerical schemes, and for evaluating ideas of subgrid scale modeling.

We believe that the difficulties are intrinsic to the dynamo problem and arise from the nature of the wave modes inside the core. It is therefore unlikely that the problem can be avoided by using a different numerical method, though of course some methods are more stable than others, and some are more efficient than others. We adopt the view that the equations themselves have to be modified to take into account the particular nature of the dynamics of the core. One way forward is to introduce hyperdiffusion to overcome the numerical difficulties. Another is to attempt to model the small scale motions, and so add further terms to the equations. The traditional “eddy viscosity” model can be viewed as an example of this approach, though here only the magnitude of the diffusion terms are enhanced, the form of the diffusion operator still being the divergence of a gradient. Further modifications, such as the Smagorinsky method, in which the eddy viscosity is taken to be dependent on the local properties of the solution (e.g., diffusion is assumed to be enhanced by turbulence in regions of strong shear) are also possible, see for example Buffett (2003). The LANS α -model (Lagrangian Averaged Navier-Stokes model) is a more radical step in this direction, because here the form of the diffusion process is assumed to change as a result of the turbulence, so that fourth order “hyperdiffusive” terms enter the equations. There is therefore a link between the addition of these terms and the attempts to model the subgrid scale processes.

There are, however, two rather different philosophies possible here, which can affect the way in which the theory is developed. On the one hand, the problem of the subgrid scales can be attacked by purely numerical methods. Prominent among these is the introduction of hyperdiffusion. This was done by Glatzmaier and Roberts (1995, 1997) and by Kuang and Bloxham (1999) essentially as a numerical method to help stabilize the equations. No physical argument was advanced to justify the form of the hyperdiffusion used. This is a reasonable approach if it can be shown that the form of the low order large scale modes is not significantly affected by the hyperdiffusion. After all, numerical methods invariably involve truncation of modes (or equivalently limitations of the mesh size), so that there is essentially nothing new in this approach. Unfortunately, the analysis of Zhang and Jones (1997) and Zhang *et al.* (1998) showed that hyperdiffusion has a significant effect on the onset of convection, and Grote *et al.* (2000) showed that

the anisotropic hyperdiffusion used by Glatzmaier and Roberts and Kuang and Bloxham has a substantial effect on the character of the solutions. One of the aims of the present work is to see whether these objections can be overcome by the use of isotropic hyperdiffusion, and to determine, by comparing the results with non-hyperdiffusive solutions, whether hyperdiffusivity is a useful numerical tool. Finite difference CFD schemes commonly employ some form of shock-capturing algorithm, which smooths out regions in which very rapid variations of the dependent variables occur. Hyperdiffusion can play a similar role in pseudo-spectral codes. It follows from this approach that the aim should be to minimize the effects of the hyperdiffusive terms wherever possible, especially in the choice of boundary conditions. When following this philosophy, we also want to gain maximum numerical stabilization, and hence allow larger timesteps, while maintaining a minimum of added hyperdiffusion.

If, on the other hand, the additional terms are motivated by physical models of the subgrid scale processes, we might hope that the results are more realistic than those of a simple truncation of the equations at a given wavenumber. In this case, we are not constrained by having to produce optimal results from the standard equations. For example, it is possible that energy can be injected on small scales and then, instead of being dissipated by viscosity, it may flow into the large scales. This is known as “back-scatter”. If it is an active process, the idea of merely enhancing dissipation through hyperdiffusion may not be the best option. Also, if the form of the subgrid scale terms has a plausible physical basis, it may be acceptable to study systems that can run only if the timestep is comparatively small, i.e., numerical stability is not necessarily the paramount consideration. The test of whether the subgrid scale models are performing well is then through comparison of low-resolution models that include subgrid scale modeling with high resolution runs of the “raw” equations.

The methods considered here are by no means the only possible ways forward, and a number of other directions are currently being explored. One such is the similarity method, where the behavior on the smallest resolved scales in the computation are analyzed and it is assumed that the unresolved scales behave in a similar manner (e.g., Buffett, 2003). Yet another approach is the shell model method, which assumes the behavior on the smallest scales is isotropic and self-similar, and can therefore be adequately represented by blocking together the large Fourier wavenumber modes into shells in Fourier space (e.g., Frick *et al.*, 2002).

2. Background to the LANS α -models

A detailed description of the reasoning that provided motivation for the α -equations would occupy more space than could be justified; only a sketch is given here. See also Montgomery and Pouquet (2002).

The basic idea may be attributed to Andrews and McIntyre (1978a,b), and is most easily illustrated for a compressible, non-rotating, non-magnetic, inviscid, isentropic fluid moving adiabatically, what is often called a “barotropic fluid”. In turbulent flow, the paths $\mathbf{X}(\mathbf{x}_0, t)$ of fluid elements from an initial state \mathbf{x}_0 scatter about a mean trajectory $\mathbf{x}(\mathbf{x}_0, t)$, the separation between them, $\xi = \mathbf{X} - \mathbf{x}$, not being necessarily small. Similarly, the difference $\mathbf{u}^\ell = D_t \xi$ between the velocity $\mathbf{u}^*(\mathbf{X}, t)$ of the fluid element and the contemporaneous velocity $\bar{\mathbf{u}}^L(\mathbf{x}, t)$ of the corresponding point on the reference trajectory is not small in general. Here $D_t = \partial_t + \bar{\mathbf{u}}^L \cdot \nabla$, where $\partial_t = \partial/\partial t$ for fixed \mathbf{x} and $\nabla_i = \partial/\partial x_i$ for fixed t . Lagrangian averages such as $\bar{\mathbf{u}}^L$ are defined as averages for fixed \mathbf{x}_0 or equivalently by

$$\overline{\chi(\mathbf{x}, t)}^L = \overline{\chi(\mathbf{X}, t)}, \quad (2.1)$$

i.e., $\overline{\chi}^L$ is the ensemble average of $\chi(\mathbf{X}, t)$ referred back to \mathbf{x} . The technique of making such transformations was initiated by Eckart (1963) and developed further by Soward (1972). It is worth observing that $\rho(\mathbf{x}, t)$ for the reference orbit is defined as $\mathcal{J}\rho^*(\mathbf{X}, t)$, where $\mathcal{J} = \partial(X_1, X_2, X_3)/\partial(x_1, x_2, x_3)$ is the Jacobian of the transformation $\mathbf{X} \rightarrow \mathbf{x}$; it follows that $\rho^*(\mathbf{X})d^3X = \rho(\mathbf{x})d^3x$. For the transformed equation of motion, see (5.4) of Soward (1972). The averaged form of the transformed equation was given by Andrews and McIntyre (1978a) and called by them “the Generalized Lagrangian Mean equation” (or “GLM equation”); see their (3.8). This was recovered by Holm from the Lagrange-averaged Lagrangian by an application of Hamilton’s principle (see, e.g., Holm, 2002).

The application of GLM theory to practical situations assumes that a large eddy simulation (LES) gives the mean trajectory about which the subgrid scales (SGS) provide the scatter. The better the resolution of the LES, the smaller the scatter, and one can imagine that in many circumstances of interest the scatter is so small that the evolution of ξ is well represented by a linear equation, and also that only terms up to quadratic order are needed in the Lagrangian ℓ to recognize the effect of the SGS on the LES:

$$\ell = \ell_0 + \ell' + \frac{1}{2}\ell'', \quad (2.2)$$

where ℓ_0 , ℓ' and ℓ'' are of order ξ^0 , ξ and ξ^2 respectively. Holm (2002, Section 4.2) obtained

$$\begin{aligned}\ell_0 &= \int \bar{\rho} \left(\frac{1}{2} |\bar{\mathbf{u}}|^2 - \bar{U} \right) d^3x, & \ell' &= \int [\rho' \left(\frac{1}{2} |\bar{\mathbf{u}}|^2 - \bar{W} \right) + \bar{\rho} \mathbf{u}' \cdot \bar{\mathbf{u}}] d^3x \\ \ell'' &= \int [\bar{\rho} |\mathbf{u}'|^2 + 2\rho' \mathbf{u}' \cdot \bar{\mathbf{u}} - \bar{c}^2 \rho'^2 / \bar{\rho}] d^3x\end{aligned}\quad (2.3)$$

where U is the internal energy, $W = U + P/\rho$ is the enthalpy (both per unit mass) and c is the sound speed; for a barotropic fluid all these, and the pressure P , depend on ρ alone. Clearly $\bar{\ell}' = 0$ and, with the help of

$$\mathbf{u}' = \partial_t \xi + \bar{\mathbf{u}} \cdot \nabla \xi - \xi \cdot \nabla \bar{\mathbf{u}}, \quad \rho' = -\nabla \cdot (\bar{\rho} \xi), \quad (2.4, 2.5)$$

Holm (2002) argued that

$$\ell'' = \int [\partial_t \xi \cdot \tilde{\mathbf{A}} \cdot \partial_t \xi + \partial_t \xi \cdot \tilde{\mathbf{B}} \cdot \xi + \xi \cdot \tilde{\mathbf{C}} \cdot \xi] d^3x, \quad (2.6)$$

where the tensors $\tilde{\mathbf{A}}$, $\tilde{\mathbf{B}}$ and $\tilde{\mathbf{C}}$ depend only on $\bar{\mathbf{u}}$, $\bar{\rho}$ and their gradients. Clearly $\bar{\ell}''$ involves statistical averages of quadratic products involving ξ and/or $\partial_t \xi$. To make progress it is necessary to obtain information about these. Once this has been done, an application of Hamilton's principle to $\bar{\ell}$ will provide an equation of motion governing the LES with the mean effects of the SGS built in.

Two possible ways of closing the theory may be attempted. The first is to solve the linear equation governing the evolution of ξ in terms of $\bar{\mathbf{u}}$ and $\bar{\rho}$. This was the approach of Gjaja and Holm (1996), in which the ξ equation was attacked by WKBJ methods. Alternatively the new equation for $\bar{\mathbf{u}}$ could be solved in conjunction with that governing ξ and the necessary quadratic averages obtained from the latter as the integration proceeded. In either case the theory has been closed in an approximate, though physically consistent, way.

A second, and much simpler, procedure is to make an *ad hoc* assumption about the unknown quadratic moments. In Section 5 of his paper, Holm (2002) lists four possible Ansätze, which he names in honor of G. I. Taylor. The first ansatz is the simplest and leads quickly to the α -model. Equation (2.4) is separated into two parts:

$$\partial_t \xi + \bar{\mathbf{u}} \cdot \nabla \xi = 0, \quad \mathbf{u}' = -\xi \cdot \nabla \bar{\mathbf{u}}. \quad (2.7, 2.8)$$

Since $\bar{\mathbf{u}}$ and $\bar{\mathbf{u}}^L$ differ only at quadratic order in ξ , (2.7) gives $\mathbf{u}' = 0$, to leading order. It also follows from (2.7) that

$$(\partial_t + \bar{\mathbf{u}} \cdot \nabla) \overline{\xi_i \xi_j} = 0, \quad (2.9)$$

so that $\overline{\xi_i \xi_j}$ is invariant following the mean motion. The easiest and most natural assumptions are that it is isotropic and homogeneous:

$$\overline{\xi_i \xi_j} = \alpha^2 \delta_{ij}, \quad \alpha = \text{constant}. \quad (2.10)$$

This defines the α -model. It is also found that (2.6) simplifies considerably to give

$$\bar{\ell} = \int \bar{\rho} \left[\frac{1}{2} (|\bar{\mathbf{u}}|^2 + \alpha^2 |\nabla \bar{\mathbf{u}}|^2) + \frac{\alpha^2}{2\bar{\rho}} (\nabla |\bar{\mathbf{u}}|^2) \cdot \nabla \bar{\rho} - \frac{\alpha^2 \bar{c}^2}{2\bar{\rho}^2} |\nabla \bar{\rho}|^2 - U(\bar{\rho}) \right] d^3x. \quad (2.11)$$

Hamilton's principle then gives

$$(\partial_t + \bar{\mathbf{u}} \cdot \nabla) \mathbf{u} + (\nabla \bar{\mathbf{u}}) \cdot \mathbf{u} = -\nabla \Pi \quad (2.12)$$

where

$$\mathbf{u} = \bar{\mathbf{u}} - \alpha^2 \left[\nabla^2 \bar{\mathbf{u}} + \frac{1}{\bar{\rho}} (\nabla \bar{\rho} \cdot \nabla) \bar{\mathbf{u}} + \frac{1}{\bar{\rho}^2} \bar{\mathbf{u}} \nabla^2 \bar{\rho} \right], \quad (2.13)$$

$$-\Pi = (1 - \alpha^2 \nabla^2) \left(\frac{1}{2} |\bar{\mathbf{u}}|^2 - \bar{W} \right) + \frac{1}{2} \alpha^2 |\nabla \bar{\mathbf{u}}|^2 - \frac{1}{2} \alpha^2 |\nabla \bar{\rho}|^2 \left(\frac{2\bar{c}}{\bar{\rho}} \frac{d\bar{c}}{d\bar{\rho}} - \frac{\bar{c}^2}{\bar{\rho}^2} \right). \quad (2.14)$$

In the remainder of this paper, we shall assume that the fluid is incompressible. To obtain the equations that then replace (2.12)–(2.14), the argument just given must be modified by introducing the constraint of constant density into the Lagrangian with a Lagrange multiplier that is interpreted as a “reduced pressure” Π . Alternatively, by setting $\bar{\rho} = \text{constant}$ in (2.12)–(2.14), we again obtain (2.12), but with a different (and irrelevant) Π and with

$$\mathbf{u} = (1 - \alpha^2 \nabla^2) \bar{\mathbf{u}}. \quad (2.15)$$

Equation (2.12) has a very attractive property (that is however not significant for the topic of this paper where the inertial terms are neglected). It preserves Kelvin's theorem in the sense that the circulation of \mathbf{u} around a closed contour carried by the mean motion $\bar{\mathbf{u}}$ is conserved.

The effects of viscous friction are ignored in (2.12). If the fluid is incompressible, the equation governing $\mathbf{u}^*(\mathbf{X})$ would include a term $\nu \nabla_{\mathbf{X}}^2 \mathbf{u}^*$, where ν is a constant kinematic viscosity. The transformation $\mathbf{X} \rightarrow \mathbf{x}$ complicates this term; see the analogous expression given by

Soward (1972) for ohmic diffusion in the induction equation governing the magnetic field \mathbf{B} . To avoid such complications, the effects of which are not necessarily serious, it is customary to recognize viscosity by adding a term such as $\nu \nabla^2 \mathbf{u}$ to the right-hand side of (2.12).

In what follows, \mathbf{u}^* will be called “the unfiltered velocity” and will usually be written simply as \mathbf{u} while $\bar{\mathbf{u}}^L$ will be termed “the filtered velocity” and written as $\bar{\mathbf{u}}$.

3. Model description and governing equations

We consider a conducting fluid layer confined between horizontal boundaries at $z = \pm \frac{1}{2}d$ as in Jones and Roberts (2000). We seek solutions in a “box” that are periodic in both the x -direction (period $2\pi d/k_x$) and the y -direction (period $2\pi d/k_y$).

The magnetic diffusivity is η , the thermal diffusivity is κ and the kinematic viscosity is ν . The layer is rotating at angular velocity Ω about the upward vertical axis, Oz. The coefficient of thermal expansion is $\bar{\alpha}$ and the temperature difference across the layer is βd .

In the Earth’s core, the inertial terms in the momentum equation are small compared with the Coriolis term (i.e., the Rossby number is small) and are often neglected, as in the high Prandtl number limit ($\nu/\kappa \rightarrow \infty$) considered here. Taking the length scale as the layer-depth d , the timescale as the magnetic diffusion time d^2/η , the magnetic field unit as $(2\Omega\rho\mu\eta)^{1/2}$ (where ρ is the fluid density and μ the permeability) and the temperature unit as βd , the governing equations are

$$\hat{\mathbf{z}} \times \mathbf{u} = -\nabla p + \mathbf{j} \times \mathbf{B} + q Ra \theta \hat{\mathbf{z}} + E \nabla^2 \mathbf{u}, \quad (3.1)$$

$$\partial_t \mathbf{B} = \nabla \times (\mathbf{u} \times \mathbf{B}) + \nabla^2 \mathbf{B}, \quad (3.2)$$

$$\partial_t \theta + \mathbf{u} \cdot \nabla \theta = q \nabla^2 \theta + \mathbf{u} \cdot \hat{\mathbf{z}}, \quad (3.3)$$

$$\nabla \cdot \mathbf{B} = 0, \quad \nabla \cdot \mathbf{u} = 0, \quad (3.4, 3.5)$$

where \mathbf{B} is the magnetic field, θ is the difference between the actual temperature gradient and the static temperature gradient and $\mathbf{j} = \nabla \times \mathbf{B}$ is the current density. The three dimensionless parameters are E , the Ekman number, Ra , the Rayleigh number and q , the diffusivity ratio defined by

$$E = \frac{\nu}{2\Omega d^2}, \quad Ra = \frac{g\bar{\alpha}\beta d^2}{2\Omega\kappa}, \quad q = \frac{\kappa}{\eta}. \quad (3.6)$$

The velocity and magnetic field are expanded in toroidal and poloidal components, but in the plane layer geometry horizontally averaged horizontal components must be added explicitly to obtain the complete representation

$$\mathbf{u} = \nabla \times e\hat{\mathbf{z}} + \nabla \times (\nabla \times f\hat{\mathbf{z}}) + U_x\hat{\mathbf{x}} + U_y\hat{\mathbf{y}}, \quad (3.7)$$

$$\mathbf{B} = \nabla \times g\hat{\mathbf{z}} + \nabla \times (\nabla \times h\hat{\mathbf{z}}) + b_x\hat{\mathbf{x}} + b_y\hat{\mathbf{y}}. \quad (3.8)$$

Here U_x and U_y define what we call “the mean velocity”, and b_x and b_y define “the mean field”; they are functions of z and t alone.

The scalar equations solved are formed by taking the z -components of the curl and of the curl of (3.1) and the z -components of (3.2) and of the curl² of (3.2):

$$\nabla_H^2(\partial_z f + E\nabla^2 e) = \hat{\mathbf{z}} \cdot \nabla \times (\mathbf{j} \times \mathbf{B}), \quad (3.9)$$

$$\nabla_H^2(\partial_z e - E\nabla^4 f + qRa\theta) = \hat{\mathbf{z}} \cdot \nabla \times \nabla \times (\mathbf{j} \times \mathbf{B}), \quad (3.10)$$

$$\nabla_H^2(\partial_t - \nabla^2)g = -\hat{\mathbf{z}} \cdot \nabla \times (\mathbf{u} \times \mathbf{B}), \quad (3.11)$$

$$\nabla_H^2(\partial_t - \nabla^2)h = -\hat{\mathbf{z}} \cdot \nabla \times \nabla \times (\mathbf{u} \times \mathbf{B}), \quad (3.12)$$

where ∇_H^2 denotes the horizontal Laplacian. Our fifth scalar equation is (3.3).

These equations must be supplemented by the equations for the horizontally averaged components:

$$-U_y - E\partial_z^2 U_x = \hat{\mathbf{x}} \cdot (\mathbf{j} \times \mathbf{B}), \quad U_x - E\partial_z^2 U_y = \hat{\mathbf{y}} \cdot (\mathbf{j} \times \mathbf{B}), \quad (3.13, 14)$$

$$(\partial_t - \partial_z^2)b_x = \hat{\mathbf{x}} \cdot \langle \nabla \times (\mathbf{u} \times \mathbf{B}) \rangle, \quad (\partial_t - \partial_z^2)b_y = \hat{\mathbf{y}} \cdot \langle \nabla \times (\mathbf{u} \times \mathbf{B}) \rangle, \quad (3.15, 16)$$

where $\langle \dots \rangle$ denotes the horizontal average

$$\langle f \rangle = \frac{k_x k_y}{4\pi^2} \int f dx dy. \quad (3.17)$$

The choice of boundary conditions is rather important. They divide into two classes, those that are required for (3.3), (3.9)–(3.16) to form a well-posed system, which we call “the large scale boundary conditions,” and the additional boundary conditions that must be imposed when the order of the diffusion operators is increased due either to the addition of hyperdiffusion or to the use of subgrid scale filtering. We call these additional boundary conditions “the subgrid scale boundary conditions.” We may wish to add terms to the large scale boundary conditions as a result of the filtering, but any such terms should vanish in the limit in which the filter size tends to zero.

For the large scale velocity conditions we take the no-slip velocity boundary conditions,

$$e = f = f' = U_x = U_y = 0 \quad \text{on} \quad z = \pm \frac{1}{2}, \quad (3.18)$$

where $f' = \partial_z f = \partial f / \partial z$. As shown in Jones and Roberts (2000) the use of perfectly conducting boundary conditions can prevent the formation of a mean field $(b_x, b_y, 0)$. We therefore adopt magnetically insulating boundary conditions. The magnetic field outside the fluid layer is then current-free; $j_z = 0$ implies $g = 0$ in the insulators, and $j_x = j_y = 0$ implies $\nabla^2 h = 0$ there. Zero current in the insulators implies that the mean field there $(b_x, b_y, 0)$, must be constant in z . We could take this constant to be non-zero, but this would correspond to the problem of dynamo action in an imposed external field, which we do not consider here. We therefore assume the field tends to zero at large distance from the layer, so that in the insulators $b_x = b_y = 0$. Continuity of tangential field then requires

$$g = b_x = b_y = 0 \quad \text{on} \quad z = \pm \frac{1}{2}, \quad (3.19)$$

and since each component of the potential field h must decay exponentially away from the layer, continuity of h and its vertical derivative implies that

$$h'_{lm} = \mp a h_{lm} \quad \text{on} \quad z = \pm \frac{1}{2}, \quad (3.20)$$

where h_{lm} denotes the Fourier component of h with wavenumber lk_x in the x -direction and mk_y in the y -direction, and which therefore has a total wavenumber of $a = (l^2 k_x^2 + m^2 k_y^2)^{1/2}$. The temperature condition is

$$\theta = 0 \quad \text{on} \quad z = \pm \frac{1}{2}. \quad (3.21)$$

The equations are solved by the pseudo-spectral method, a brief description being given in Jones and Roberts (2000). The fields are expanded in Fourier series in x and y , and Chebyshev series in z . The Fourier series are truncated at the N_x and N_y levels, so only harmonics in the range $(-N_x, N_x)$ and $(-N_y, N_y)$ are retained, and $N_z + N_{bc}$ Chebyshev components are retained for each variable, where N_{bc} depends on the number of boundary conditions. We have not imposed any particular symmetry on the solutions. Because the Boussinesq approximation has been adopted and the boundary conditions on the horizontal surfaces are the same, many symmetric solutions exist, but we have explored only the asymmetric solutions which arise naturally if a random initial seed field is used to start the calculation.

With a typical truncation of $N_x = N_y = N_z = 32$, there are of the order of 130,000 modes in the system, so we are effectively solving a coupled system of time-dependent ordinary differential equations of this order. Nevertheless, the linear parts of this system (the terms on the left-hand side of equations (3.3), (3.9)–(3.16)) can be solved implicitly using a Crank-Nicolson scheme, because the x and y dependence can be separated out. The nonlinear terms, however, are treated explicitly, because the large order of the system makes fully implicit methods very expensive. The numerical performance of this code has been discussed in Rotvig and Jones (2002). The spatial resolution provided by the Fourier-Chebyshev expansions has a rather abrupt threshold which means that, if the truncation level is too small for the parameters of the run, the solution suffers explosive numerical instability however small the timestep, but once the resolution is high enough to maintain stability, the accuracy increases very rapidly as the number of modes is increased. In practice, this means that there is a fairly sharp optimum number of modes to use for any given set of parameter values. This optimum value increases as either the Ekman number is lowered or the Rayleigh number is increased. The principal practical difficulty is the very small timestep required for numerical stability. The exact reasons for this are not well understood, though some light can be shed by considering the MAC wave dispersion relation (see below). The numerical method used for the timestepping is a simple Euler predictor-corrector method, and the timestep is varied automatically as the integration proceeds so that the difference between the predictor and corrector solutions is maintained at a set level. This maintains accuracy and stability. If the solution is started with a small seed magnetic field, the timestep chosen by the code during the kinematic build-up phase is not far from that expected from the CFL condition, i.e., the timestep has to be small enough for the product of the maximum velocity (here the local magnetic Reynolds number) and the timestep to be less than the mesh size, which we can approximate as $1/\max(N_x, N_y, N_z)$. However, when the magnetic field starts to saturate, the timestep drops significantly: for the range of parameters considered below, the reduction can be a factor ten or more. As E is lowered, the timestep drops further (see Rotvig and Jones, 2002, for details), and this effectively gives a barrier below which we cannot go in reducing E . Similar problems are experienced in spherical codes as E is lowered.

In view of this, it might seem a good idea to re-introduce inertia. However, the waves resulting from the balance between the inertial and Coriolis forces are the inertial waves, which have a natural timescale of days. The waves associated with buoyancy are internal gravity waves,

though these usually manifest themselves in their unstable form as growing convective modes. The magnetic waves are Alfvén waves, which typically have a timescale of a few years. (These are the analogue of the torsional oscillations that arise in the corresponding spherical system; because of the degeneracy of the plane layer, they have a greater significance here.) The magnetic fields grow in dynamo calculations on the eddy turnover timescale, which is the 10^5 year diffusion timescale d^2/η divided by $R_m \sim 300$. To establish that fields are not decaying through diffusion, it is necessary to run codes for a substantial fraction of a diffusion time, which in practice is at least 0.1 of the diffusion time, i.e., at least 10,000 years. It is clear that if waves on the timescale of days are going to appear in the simulation the timestep needs to be a small fraction of a day, which implies that a minimum of 10^8 timesteps will be needed per run. Even with high speed parallel processors, this is a severe restriction for a fully three-dimensional Navier-Stokes solver. We can overcome this difficulty either by slowing down the inertial waves and the Alfvén waves, by altering the rotation rate relative to the eddy turnover time (this is done in those codes that retain inertia by artificially taking $\nu = \kappa = \eta$), or by removing the inertial term entirely. The $\nu = \kappa = \eta$ approach has the disadvantage that waves, that in reality are on a different timescale from the dynamo modes, are now on the same timescale as the convective turnover in the simulation, and so may be strongly excited by the convection. This in turn can affect the generation mechanism, as evidenced by recent models of Busse (2002) in which inertial effects strongly influence the dynamo. For this reason we here exclude the inertial terms $D\mathbf{u}/Dt$ *ab initio*.

4. The hyperdiffusive and the LANS α -models

With the addition of isotropic hyperdiffusion to all three diffusing quantities, (3.3) and (3.9)–(3.16) become

$$\nabla_H^2 [\partial_z f + E \nabla^2 (1 - \alpha_V^2 \nabla^2) e] = \hat{\mathbf{z}} \cdot \nabla \times (\mathbf{j} \times \mathbf{B}), \quad (4.1)$$

$$\nabla_H^2 [\partial_z e - E \nabla^4 (1 - \alpha_V^2 \nabla^2) f + q Ra \theta] = \hat{\mathbf{z}} \cdot \nabla \times [\nabla \times (\mathbf{j} \times \mathbf{B})], \quad (4.2)$$

$$\nabla_H^2 [\partial_t - \nabla^2 (1 - \alpha_B^2 \nabla^2)] g = -\hat{\mathbf{z}} \cdot \nabla \times (\mathbf{u} \times \mathbf{B}), \quad (4.3)$$

$$\nabla_H^2 [\partial_t - \nabla^2 (1 - \alpha_B^2 \nabla^2)] h = -\hat{\mathbf{z}} \cdot \nabla \times [\nabla \times (\mathbf{u} \times \mathbf{B})], \quad (4.4)$$

$$\partial_t \theta - q \nabla^2 (1 - \alpha_T^2 \nabla^2) \theta + \nabla_H^2 f = -\mathbf{u} \cdot \nabla \theta, \quad (4.5)$$

$$-U_y - E \partial_z^2 (1 - \alpha_V^2 \partial_z^2) U_x = \hat{\mathbf{x}} \cdot (\mathbf{j} \times \mathbf{B}), \quad (4.6)$$

$$U_x - E \partial_z^2 (1 - \alpha_V^2 \partial_z^2) U_y = \hat{\mathbf{y}} \cdot (\mathbf{j} \times \mathbf{B}), \quad (4.7)$$

$$[\partial_t - \partial_z^2 (1 - \alpha_B^2 \partial_z^2)] b_x = \hat{\mathbf{x}} \cdot (\nabla \times (\mathbf{u} \times \mathbf{B})), \quad (4.8)$$

$$[\partial_t - \partial_z^2 (1 - \alpha_B^2 \partial_z^2)] b_y = \hat{\mathbf{y}} \cdot (\nabla \times (\mathbf{u} \times \mathbf{B})). \quad (4.9)$$

Here α_V , α_B and α_T are the length scales associated with the three forms of hyperdiffusion. In principle, they can all be different, but in the numerical runs they are generally taken to have the same value. There is some numerical evidence to suggest that the temperature distribution is comparatively smooth for the parameter regime tested here (see, e.g., Fig. 12 of Rotvig and Jones, 2002) and so we have also done some runs with $\alpha_T = 0$. If the very small q regime could be accessed, we might expect high Péclet number solutions to emerge, in which case small length scales will reappear in the temperature field, so this issue should be left open.

The LANS α -model has recently been extended to the MHD case, and equations closely related to the hyperdiffusive equations have been proposed by Holm (2003) in what we shall call “Holm-MHD theory”. Following the principles of Section 2, we introduce filtered (barred variables) and unfiltered (unbarred) variables with the relations

$$\begin{aligned} e &= (1 - \alpha^2 \nabla^2) \bar{e}, & f &= (1 - \alpha^2 \nabla^2) \bar{f}, \\ g &= (1 - \alpha^2 \nabla^2) \bar{g}, & h &= (1 - \alpha^2 \nabla^2) \bar{h}, & \theta &= (1 - \alpha^2 \nabla^2) \bar{\theta}, \end{aligned} \quad (4.10)$$

$$\begin{aligned} U_x &= (1 - \alpha^2 \partial_z^2) \bar{U}_x, & U_y &= (1 - \alpha^2 \partial_z^2) \bar{U}_y, \\ b_x &= (1 - \alpha^2 \partial_z^2) \bar{b}_x, & b_y &= (1 - \alpha^2 \partial_z^2) \bar{b}_y. \end{aligned} \quad (4.11)$$

The Holm-MHD equations then take the form

$$\nabla_H^2 (\partial_z \bar{f} + E \nabla^2 e) = \hat{\mathbf{z}} \cdot \nabla \times (\mathbf{j} \times \bar{\mathbf{B}}), \quad (4.12)$$

$$\nabla_H^2 (\partial_z \bar{e} - E \nabla^4 f + q Ra \bar{\theta}) = \hat{\mathbf{z}} \cdot \nabla \times [\nabla \times (\mathbf{j} \times \bar{\mathbf{B}})], \quad (4.13)$$

$$\nabla_H^2 (\partial_t \bar{g} - \nabla^2 g) = -\hat{\mathbf{z}} \cdot \nabla \times (\bar{\mathbf{u}} \times \bar{\mathbf{B}}), \quad (4.14)$$

$$\nabla_H^2 (\partial_t \bar{h} - \nabla^2 h) = -\hat{\mathbf{z}} \cdot \nabla \times [\nabla \times (\bar{\mathbf{u}} \times \bar{\mathbf{B}})], \quad (4.15)$$

$$(\partial_t - q \nabla^2) \bar{\theta} + \nabla_H^2 \bar{f} = -\bar{\mathbf{u}} \cdot \nabla \bar{\theta}, \quad (4.16)$$

$$-\bar{U}_y - E \partial_z^2 U_x = \hat{\mathbf{x}} \cdot (\mathbf{j} \times \bar{\mathbf{B}}), \quad \bar{U}_x - E \partial_z^2 U_y = \hat{\mathbf{y}} \cdot (\mathbf{j} \times \bar{\mathbf{B}}), \quad (4.17, 18)$$

$$\partial_t \bar{b}_x - \partial_z^2 b_x = \hat{\mathbf{x}} \cdot (\nabla \times (\bar{\mathbf{u}} \times \bar{\mathbf{B}})), \quad \partial_t \bar{b}_y - \partial_z^2 b_y = \hat{\mathbf{y}} \cdot (\nabla \times (\bar{\mathbf{u}} \times \bar{\mathbf{B}})). \quad (4.19, 20)$$

Equations (4.12)–(4.20) are remarkably similar to the hyperdiffusive system (4.1)–(4.9), provided the hyperdiffusive variables are interpreted as the filtered variables of the Holm-MHD theory. The only difference is in the expression for the Lorentz force (in (4.12), (4.13), (4.17) and (4.18)), where the unfiltered current \mathbf{j} and the filtered $\bar{\mathbf{B}}$ appear in the Holm equations, whereas in the hyperdiffusive system both a filtered current $\bar{\mathbf{j}}$ and a filtered $\bar{\mathbf{B}}$ are used (see also Subramanian, 2003).

5. Subgrid scale boundary conditions

The anisotropic hyperdiffusion models of Glatzmaier and Roberts and Kuang and Bloxham only enhanced the diffusion on spherical surfaces and not in the radial direction. This is convenient if a numerical method is used where the ϑ and ϕ dependence is represented by a spherical harmonic expansion. Thus they replaced the ∇^2 operator by a formula such as

$$(1 + \alpha\ell^3)\nabla^2 \quad (5.1)$$

where ℓ is the degree of the spherical harmonic of the expansion function. In some runs, the standard diffusion is used for harmonics of lower degree, hyperdiffusivity only being applied to the higher harmonics. Using a formula of the type (5.1) avoids the necessity of imposing extra boundary conditions at the inner core boundary and the core-mantle boundary. However, if we wish to avoid an unphysical anisotropy and instead adopt an isotropic hyperdiffusion, to avoid the criticisms of Grote *et al.* (2000), we must add boundary conditions in the radial direction. In our plane layer model this corresponds to additional boundary conditions at $z = \pm 1/2$.

For the LANS α theory, it is helpful to think of the boundary conditions in terms of the variables \mathbf{u} and $\bar{\mathbf{u}}$ related by (2.13). In the no-slip case, the most natural approach would be to set both \mathbf{u} and $\bar{\mathbf{u}}$ to zero at the boundaries, but unfortunately this is not possible. We have three large scale boundary conditions, $\bar{f} = \bar{f}' = \bar{e} = 0$, to apply at each boundary, and we need two further subgrid scale conditions, because (4.10) and (4.11) involve second order operators. However, $e = f = f' = 0$ implies $\bar{f}'' = \bar{f}''' = \bar{e}'' = 0$ which is one too many conditions, so that the resulting problem is not well posed. For the hyperdiffusive equations, it is more natural to force the barred variables to satisfy the large scale boundary conditions.

We next look at the energy equation, and attempt to choose the boundary conditions through the requirement that the surface integrals which appear should vanish. The energy equation from the Holm-MHD equations is formulated by taking the scalar product of the momentum equation with $\bar{\mathbf{u}}$. The viscous term then gives

$$\int \bar{\mathbf{u}} \cdot \nabla^2 \mathbf{u} d^3x = \oint_S [\mathbf{ds} \cdot \nabla \mathbf{u}] \cdot \bar{\mathbf{u}} + \alpha_V^2 \oint_S [\mathbf{ds} \cdot \nabla \bar{\mathbf{u}}] \cdot \nabla^2 \bar{\mathbf{u}} - \int [|\nabla \bar{\mathbf{u}}|^2 + \alpha_V^2 |\nabla^2 \bar{\mathbf{u}}|^2] d^3x. \quad (5.2)$$

The first surface term vanishes if the large scale boundary conditions $\bar{\mathbf{u}} = \mathbf{0}$ are adopted. To rid ourselves of the second term we could impose $\partial_z \bar{\mathbf{u}} = \mathbf{0}$. The z component of this is already zero ($\bar{u}_z = \nabla_H^2 \bar{f}$, and $\bar{f}' = 0$), so the full set of boundary conditions is

$$\bar{f} = \bar{f}' = \bar{e} = \bar{e}' = \bar{f}'' = 0, \quad \text{on} \quad z = \pm \frac{1}{2}, \quad (5.3)$$

the first three being the large scale boundary conditions and the second two the subgrid scale boundary conditions. For the mean flow, we use

$$\bar{U}_x = \bar{U}'_x = \bar{U}_y = \bar{U}'_y = 0 \quad \text{on} \quad z = \pm \frac{1}{2}. \quad (5.4)$$

We have adopted (5.3) and (5.4) in most of the numerical simulations (see Section 7 below). We found that they are numerically satisfactory, in that no new numerical instabilities are introduced by this choice of subgrid scale boundary conditions, and in Section 6 below we find that they do not lead to substantial changes at the onset of convection. This choice is not the only possible one though, because $\nabla^2 \bar{\mathbf{u}}_H$ vanishes if $\bar{e}'' = \bar{f}''' = 0$, and this also gives a well-posed system. An alternative set of mechanical boundary conditions is discussed in Appendix B, which also makes the viscous dissipation positive definite whatever the flow in the layer, a desirable property for a hyperdiffusive system.

We now turn to the magnetic boundary conditions. There is here no problem with applying (3.19) and (3.20) to both the filtered and unfiltered fields. In terms of the filtered variables we then have (for the Fourier component (l, m))

$$\bar{g} = \bar{g}'' = 0, \quad \bar{h}'_{lm} = \mp a \bar{h}_{lm}, \quad \bar{h}'''_{lm} = \mp a \bar{h}''_{lm} \quad \text{on} \quad z = \pm \frac{1}{2}. \quad (5.5)$$

The mean field components can be treated similarly to give

$$\bar{b}_x = \bar{b}_y = \bar{b}''_x = \bar{b}''_y = 0 \quad \text{on} \quad z = \pm \frac{1}{2}. \quad (5.6)$$

These are the magnetic boundary conditions we have used for the numerical simulations in Section 7.

The magnetic energy integrals have integrated surface terms but, unlike the kinetic energy integrals, they do not vanish because the magnetic field is non-zero outside the layer, so that magnetic energy flows through the boundaries. For the hyperdiffusive system (4.1)–(4.9) we would like to ensure that the ohmic dissipation is positive:

$$\int \bar{\mathbf{j}} \cdot (1 - \alpha_B^2 \nabla^2) \bar{\mathbf{j}} d^3x = \int \bar{\mathbf{j}} \cdot \bar{\mathbf{j}} d^3x - \alpha_B^2 \oint_S \bar{\mathbf{j}} \cdot \partial_z \bar{\mathbf{j}} dS + \alpha_B^2 \int |\partial_z \bar{\mathbf{j}}|^2 d^3x > 0. \quad (5.7)$$

The surface integral contains parts involving \bar{b}_x , \bar{b}_y , \bar{g} and \bar{h} , but the parts involving \bar{b}_x , \bar{b}_y and \bar{g} vanish if (5.5) and (5.6) are employed, while the part involving \bar{h} is negative if (5.5) is used. We conclude that (5.5) and (5.6) are suitable boundary conditions for the hyperdiffusive system, but note that, as for the case of the velocity boundary conditions, the choice made through requiring the surface terms to have appropriate properties does not lead to a unique set of boundary conditions. An alternative is discussed in Appendix A.

Our choice of boundary conditions seems physically reasonable and is apparently numerically stable, but a clearer physical motivation for them is desirable. In the case of the Holm-MHD equations this would require an analysis of the expected form of the turbulent fluctuations near the boundaries. See Appendix B for a further discussion.

In cases where a thermal hyperdiffusion is used an additional boundary condition on the temperature is required. This was taken to be $\bar{\theta}' = 0$, so the full thermal boundary conditions used were

$$\bar{\theta} = \bar{\theta}' = 0 \quad \text{on} \quad z = \pm \frac{1}{2}. \quad (5.8)$$

Additional boundary conditions on θ are not required for the Holm-MHD equations.

6. Linear theory

Although the dynamo problem is fully nonlinear, there are two related linear problems that can give some insight into the behavior of the nonlinear system. The nonlinear system is fully three-dimensional, so that it is not practical to carry out detailed explorations of parameter space, but insights from linear theory are crucial for interpreting the output from the limited number of fully nonlinear runs it is possible to perform.

The first linear problem is to examine the effect of hyperdiffusion on the nonmagnetic onset of rotating convection. General features such as the boundary layer structure may persist into the nonlinear regime. The second linear problem is the magnetoconvection problem with a uniform imposed magnetic field, which gives rise to the MAC wave dispersion relation.

The linearized equations for the onset of convection are

$$2\hat{\mathbf{z}} \times \mathbf{u} = -\nabla p + E\nabla^2(1 - \alpha_V^2\nabla^2)\mathbf{u} - Ra\theta\hat{\mathbf{z}}, \quad (6.1)$$

$$\partial_t\theta = u_z + \nabla^2(1 - \alpha_T^2\nabla^2)\theta. \quad (6.2)$$

Here we have used d^2/κ as the unit of time. These are supplemented by the no-slip boundary conditions (3.18) and the constant temperature boundary condition (3.21). The main results from the standard linear theory are given in Fearn *et al.* (1988) and here we just summarize the differences introduced by the hyperdiffusive terms, which we assume are small. We can best understand the behavior if we take the limit $E \rightarrow 0$. In the classical problem, the horizontal wavenumber at marginal stability is then $O(E^{1/3})$ and there is an Ekman layer of thickness $O(E^{1/2})$. The dissipation term $E\nabla^2\mathbf{u}$ in the interior is $O(E^{1/3}\mathbf{u})$, while the dissipation in the Ekman layer is slightly smaller at $O(E^{1/2}\mathbf{u})$ so the interior dissipation dominates. Stress-free boundaries allow a solution with $\theta = \theta_0 \cos \pi z$, which gives marginal stability at

$$Ra = \frac{\pi^2[1 + \alpha_T^2(\pi^2 + k^2)]}{E(\pi^2 + k^2)[1 + \alpha_V^2(\pi^2 + k^2)]} + \frac{E(\pi^2 + k^2)^3[1 + \alpha_V^2(\pi^2 + k^2)][1 + \alpha_T^2(\pi^2 + k^2)]}{k^2}. \quad (6.3)$$

The critical horizontal wavenumber k greatly exceeds π and is therefore determined by

$$\frac{k^6(1 + \alpha_V^2 k^2)^2[2 + 3k^2(\alpha_V^2 + \alpha_T^2) + 4k^4\alpha_V^2\alpha_T^2]}{1 + 2k^2\alpha_V^2 + k^4\alpha_V^2\alpha_T^2} = \frac{E^2}{\pi^2}. \quad (6.4)$$

It is now clear that, if

$$\alpha_V \ll \left(\frac{E^2}{2\pi^2}\right)^{1/6} \quad \text{and} \quad \alpha_T \ll \left(\frac{E^2}{2\pi^2}\right)^{1/6}, \quad (6.5)$$

the hyperdiffusive terms make only a small change to the critical wavenumber for convection. The situation is not greatly changed if no-slip boundaries are used. One may reasonably hope that, provided criteria

(6.5) are satisfied, the behavior of the convection in the interior will not be dramatically altered by the addition of hyperdiffusion. Of course, the smaller length scales generated by nonlinear interactions, which can include for example sharp plume boundaries or regions of rapidly varying magnetic field, will be affected. The hope is that this modification will represent to some extent the turbulent processes at work in the fluid.

We also need to examine the effect of hyperdiffusion on the boundary layers. To do this, we return to (6.3), but now replace π^2 by $-k_z^2$, where $1/k_z$ is the boundary layer thickness, which is much smaller (at low E) than $1/k$. The equation for k_z becomes

$$E^2 k_z^4 (1 - \alpha_V^2 k_z^2)^2 = -1. \quad (6.6)$$

The classical limit is

$$\alpha_V \ll E^{1/2}, \quad (6.7)$$

and then the smallest roots for k_z (which produce the thickest boundary layer) are the classical Ekman layer values $k_z = (\pm 1 \pm i)/\sqrt{2E}$. The solutions corresponding to two of these roots are rejected from the boundary layer solution because they grow as we move into the interior, and the remaining two are determined by the boundary conditions and the matching to the interior flow. Since (6.6) is of eighth order there are four more solutions, of which two can be rejected as they do not decay as we move into the interior. The remaining two are determined by the additional subgrid scale boundary conditions. This leads to an interior boundary layer of thickness α_V lying within the Ekman boundary layer itself of thickness $O(E^{1/2})$. This could cause numerical resolution problems; the need to resolve the Ekman boundary layer is already a serious constraint on the numerical code, and we do not wish to exacerbate this. If there were no magnetic field, we could choose our subgrid scale boundary conditions to eliminate this interior layer. However, in the presence of a magnetic field having the expected $O(1)$ Elsasser number, the Ekman layer becomes an Ekman-Hartmann layer whose thickness varies depending on the local field strength near the boundary as the computation progresses. It is not a practical proposition to choose boundary conditions that always eliminate the α_V boundary layer in a dynamo code.

In practice we therefore choose α_V so that it is of order $E^{1/2}$. The structure of the Ekman boundary layer is affected, but its thickness

remains of $O(E^{1/2})$. The eight roots of (6.6) are then

$$k_z = \pm \left[\frac{1 \pm \sqrt{1 \pm 4i\alpha_V^2 E^{-1}}}{2\alpha_V^2} \right]^{1/2} \quad (6.8)$$

four of which decay as we move into the interior and hence are the basis of the boundary layer solution. This change in the boundary layer structure may be a cause for concern, but we should remember that an $O(E^{1/2})$ boundary layer does not affect the onset of convection at leading order, so we do not expect the interior convection to be strongly affected. The most significant change from the alteration of the Ekman boundary layer is probably to the geostrophic flows induced by the magnetic field. Boundary layer friction can play a role in the dynamics of these motions, and it is probable that this is connected with the stabilization that hyperdiffusion produces.

To see the effect of the hyperdiffusion on the marginal stability problem for the typical values of hyperdiffusion used, we have constructed Table 1 using a linear stability code with no-slip boundaries, and the subgrid scale boundary conditions in (5.3).

The other relevant linear problem is the magnetoconvection problem, that is convection in the presence of an imposed magnetic field \mathbf{B}_0 . We return to the magnetic diffusion time-scale, so the system studied is

$$\partial_z \mathbf{u} = (\mathbf{B}_0 \cdot \nabla) \mathbf{j} + E \nabla^2 (1 - \alpha_V^2 \nabla^2) \zeta + \nabla \times (q Ra \theta \hat{\mathbf{z}}), \quad (6.9)$$

$$\partial_t \mathbf{b} = (\mathbf{B}_0 \cdot \nabla) \mathbf{u} + \nabla^2 (1 - \alpha_B^2 \nabla^2) \mathbf{b}, \quad (6.10)$$

$$\partial_t \theta = u_z + q \nabla^2 (1 - \alpha_T^2 \nabla^2) \theta, \quad (6.11)$$

where $\zeta = \nabla \times \mathbf{u}$. We seek disturbances of the form $\exp i(\mathbf{k} \cdot \mathbf{x} - \omega t)$, and write

$$k_V^2 = k^2 (1 + \alpha_V^2 k^2), \quad k_B^2 = k^2 (1 + \alpha_B^2 k^2), \quad k_T^2 = k^2 (1 + \alpha_T^2 k^2). \quad (6.12)$$

The dispersion relation is

$$\left(\frac{(\mathbf{B}_0 \cdot \mathbf{k})^2}{k_B^2 - i\omega} + E k_V^2 \right) \left(\frac{(\mathbf{B}_0 \cdot \mathbf{k})^2}{k_B^2 - i\omega} + E k_V^2 - \frac{q R a k_{\perp}^2}{k^2 (q k_T^2 - i\omega)} \right) + \frac{k_z^2}{k^2} = 0, \quad (6.13)$$

where k_z is the z -component of the wavevector, and k_{\perp} is the magnitude of the horizontal component. In the absence of all diffusion this reduces

Table 1. The minimum Rayleigh number Ra and the corresponding horizontal wavenumber k are tabulated as a function of Ekman number E and the hyperdiffusivities α_V^2 and α_T^2

E	α_V^2	α_T^2	minimizing k	Ra
0.001	0	0	10.820	71.085
0.001	0.0001	0.0001	10.710	72.725
0.001	0.0002	0.0002	10.590	73.417
0.001	0.0005	0.0005	10.277	75.087
0.001	0.001	0.001	9.861	77.514
0.0005	0	0	13.870	88.224
0.0005	0.0001	0.0001	13.607	89.876
0.0005	0.0002	0.0002	13.367	90.663
0.0005	0.0002	0.0005	13.075	95.924
0.0005	0.0005	0.0002	13.056	87.948
0.0005	0.0005	0.0005	12.785	92.851
0.0005	0.001	0.001	12.070	96.246
0.0002	0	0	19.248	119.888
0.0002	0.0001	0.0001	18.542	121.520
0.0002	0.0002	0.0002	17.997	122.789
0.0002	0.0005	0.0005	16.797	126.809
0.0002	0.001	0.001	15.455	133.005

to the MAC wave dispersion relation (Braginsky, 1964)

$$\omega^2 = \frac{k^2(\mathbf{B}_0 \cdot \mathbf{k})^2}{k_z^2} \left((\mathbf{B}_0 \cdot \mathbf{k})^2 - \frac{qRak_{\perp}^2}{k^2} \right) \quad (6.14)$$

which describes slow waves in the Earth's core. When the Elsasser number $\Lambda = B_0^2/2\Omega\mu\rho\eta$ is order one, the timescale of these waves is the magnetic diffusion time for the core. For those modes where the Rayleigh number is large enough to make negative the term between large brackets in (6.14), the waves become growing convective modes. Although the timescale for these modes is generally on the dynamo timescale, modes with k_z small or zero have much higher frequencies, and this causes numerical difficulties (Walker *et al.*, 1998). These are the geostrophic modes where there is no longer a balance between Coriolis force and Lorentz force. Either Taylor's constraint is satisfied, which in our plane layer means the z -averaged Lorentz force over every vertical line must vanish, or any residual z -averaged component has to be balanced by viscous forces. This can lead to large accelerations in the velocity field.

If we set $k_z = 0$ in (6.13), the toroidal and poloidal components separate, and we obtain

$$\omega = -i \left(k_B^2 + \frac{(\mathbf{B}_0 \cdot \mathbf{k})^2}{Ek_V^2} \right), \quad \omega = -i \left(k_B^2 + \frac{k^2(\mathbf{B}_0 \cdot \mathbf{k})^2 - qRa k_\perp^2}{k^2 E k_V^2} \right). \quad (6.15)$$

The first of these describes damped modes, and the second either damped or growing convective modes. The fastest growing modes are those with “plate-like” motions in which \mathbf{k} is perpendicular to both the rotation axis and the field, corresponding to motions in the plane containing the rotation and magnetic field vectors (Braginsky and Meyl, 1990). In a nonlinear system, the growth of these modes will be controlled by nonlinearity, but the appearance of E in the denominator indicates that numerical difficulties are likely at small E . The fact that the first solution in (6.15) (which belongs to the toroidal e component of the velocity) is strongly damped might give rise to the naïve hope that these modes would not matter in a numerical simulation. In reality, strongly damped modes make the system stiff, and the timestep may be expected to drop to a value for which this fast decay on a timescale E/Λ has to be resolved.

7. Numerical simulations

To test how isotropic hyperdiffusion and the Holm-MHD model behave in numerical simulations, we have performed a series of runs with our plane layer convective dynamo model. Numerical details of the implementation can be found in Jones and Roberts (2000) and Rotvig and Jones (2002). The aim here is simply to test the codes: the implications for dynamo theory are discussed in the previous papers. We have concentrated on three sets of parameter values. Benchmark (i) is an “easy” parameter set with $E = 10^{-3}$, $q = 1$, $Ra = 1000$ and a box size such that the $\ell = m = 1$ horizontal wavenumber is $k_x = k_y = 2\pi\sqrt{2}$ (Run G of Jones and Roberts, 2000). This Rayleigh number is about 14 times the minimum plane layer value (see Table 1), and the restriction on the choice of horizontal wavelengths imposed by the box size only increases the critical value slightly. A cubical periodic box has wavenumber 2π ; the reason for choosing a slightly smaller box is that at these wavenumbers the dynamo is not subcritical, i.e., the magnetic field grows from a small random seed field. A resolution of only 16^3 , i.e., a truncation level of $N_x = N_y = N_z = 16$, is just adequate for this case, but all runs were performed at 32^3 resolution. The second benchmark (ii) is $E = 5 \times 10^{-4}$,

$q = 1$ and $Ra = 2500$, with $\alpha = \beta = 2\pi$. This is also a self-excited dynamo, but the higher Ra and lower E make this a somewhat more demanding benchmark. This was run mainly at 32^3 , though it was also thoroughly tested at $N_x = N_y = N_z = 32 \times 32 \times 48$, as the z -resolution is slightly more critical than the x, y resolution for this run. It was also checked against a longer multiprocessor run at $48 \times 48 \times 64$. Our final benchmark (iii) is $E = 2 \times 10^{-4}$, $Ra = 5000$, $q = 0.5$, again with the cubical box $\alpha = \beta = 2\pi$. The higher Ra and lower q in particular makes benchmark (iii) considerably more demanding. It has not yet been investigated as thoroughly as the others; in the absence of hyperdiffusion a truncation of $32 \times 32 \times 64$ is an adequate resolution to avoid explosive numerical instability.

All three benchmarks are kinematic dynamos, in the sense that, when started from a small random seed, the magnetic field grows and then saturates. As in previous work, it is notable that because the convective flow is highly time-dependent, particularly for the higher Rayleigh number benchmarks, field growth is an erratic process, comparatively short bursts of generation being followed by longer periods of consolidation. This means that the onset of dynamo action is much more easily studied than the saturated regime. For example, the timestep for the saturated range of benchmark (ii) is typically twenty times less than for the kinematic phase.

As outlined above, a major constraint is the small timestep necessary in dynamo codes, so in Table 2 we give the typical timestep in the saturated regime needed to maintain numerical stability. The timestepping procedure used is a simple predictor-corrector scheme. At each step, the maximum relative change of the variables is monitored, and if it is

Table 2. The typical timestep taken in the saturated regime for various runs

Benchmark	Hyperdiffusion	Typical timestep
(i)	None	8×10^{-6}
(ii)	None	5×10^{-7}
(iii)	None	2×10^{-7}
(i)	Holm-MHD $\alpha^2 = 0.0002$	10^{-5}
(ii)	Holm-MHD $\alpha^2 = 0.0002$	6×10^{-7}
(ii)	All 3, $\alpha^2 = 0.0005$	4×10^{-6}
(iii)	All 3, $\alpha^2 = 0.0005$	2×10^{-6}
(iii)	All 3, $\alpha^2 = 0.0010$	4×10^{-6}
(ii)	\mathbf{u}, \mathbf{B} only, $\alpha^2 = 0.0005$	4×10^{-6}

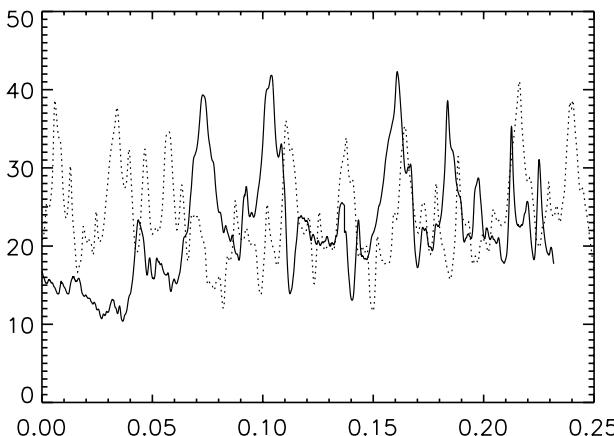


Figure 1. Benchmark (ii), $Ra = 2500$, $q = 1$, $E = 0.0005$. Solid line, plot of magnetic energy against time with no hyperdiffusion, average 22.24. Dotted line, plot of magnetic energy against time with hyperdiffusion $\alpha_V^2 = \alpha_B^2 = \alpha_T^2 = 0.0005$, i.e., equal hyperdiffusion applied to all three diffusing quantities. Average 23.86.

more than 10^{-3} in any Chebyshev-Fourier component, the timestep is halved until the relative change is below the threshold. If the relative change is less than 2×10^{-5} the timestep is doubled. In this way, the code chooses its own timestep. Any attempt to enforce a larger step than that selected by the code rapidly results in numerical instability. We here remark that the linear terms are dealt with implicitly, and the nonlinear terms (the Lorentz force, the nonlinear induction and the temperature advection) are treated explicitly, as usual in pseudo-spectral codes.

In Fig. 1 we compare the time-dependence of the magnetic energy for benchmark (ii) both without hyperdiffusivity and with hyperdiffusivity. In this case all three hyperdiffusions were used, and all set equal to $\alpha^2 = 0.0005$. The equations used here are (4.1)–(4.9), with boundary conditions (5.3)–(5.6). The equivalent plots for the kinetic energy are shown in Fig. 2. Note that, even if the comparison runs are started from the same data file, they will soon begin to diverge because there is chaotic time-dependence with sensitivity to initial conditions. We can therefore only compare statistical quantities such as the mean magnetic field energy. It is clear from Figs. 1 and 2 that the hyperdiffusion is not affecting either the kinetic energy or the magnetic energy very greatly, and indeed it would require a very long run to establish any significant difference between the plots. Note also that it is not only the average energy that is similar, but also the amplitude of the peaks and troughs and the frequency of their occurrence.

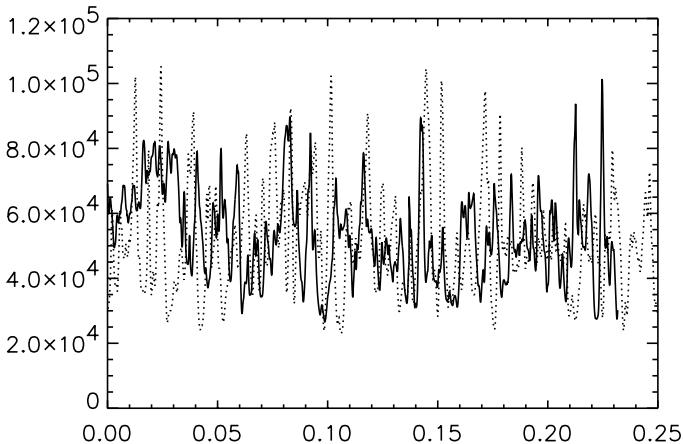


Figure 2. Benchmark (ii), $Ra = 2500$, $q = 1$, $E = 0.0005$. Solid line, plot of kinetic energy against time with no hyperdiffusion, average 53,141. Dotted line, plot of kinetic energy against time with hyperdiffusion $\alpha_V^2 = \alpha_B^2 = \alpha_T^2 = 0.0005$, i.e., equal hyperdiffusion applied to all three diffusing quantities. Average 51,604.

In Figs. 3 and 4 we present the same plots, but for the case where there is magnetic and viscous hyperdiffusion, but no thermal hyperdiffusion. The timestep and general stability of the code has similar characteristics to the case where all three hyperdiffusions are present, but now there is a significant difference between the hyperdiffusive and non-hyperdiffusive solutions.

The hyperdiffusive solutions have significantly smaller average magnetic and kinetic energies than the non-hyperdiffusive solutions and this persists if the run is extended. This result seems surprising at first, because one might expect that removing a hyperdiffusion and replacing it by regular diffusion would increase the energies, not decrease them. However, by removing the temperature hyperdiffusion while retaining the magnetic field hyperdiffusion we have in effect decreased q . It is well known that lowering q makes dynamo action considerably more difficult, so it is less surprising that the magnetic energy is reduced by the removal of the thermal hyperdiffusion. The fall in the kinetic energy is more unexpected, because the linear results in Table 1 suggest that increasing the thermal hyperdiffusion raises the critical Rayleigh number, and so we would expect removing thermal hyperdiffusion would increase the kinetic energy. To investigate this effect we switched off the Lorentz force, and noted that the kinetic energy was still larger with thermal hyperdiffusion than without it, so this is not primarily a magnetic effect. Contour plots of the temperature and velocity suggest

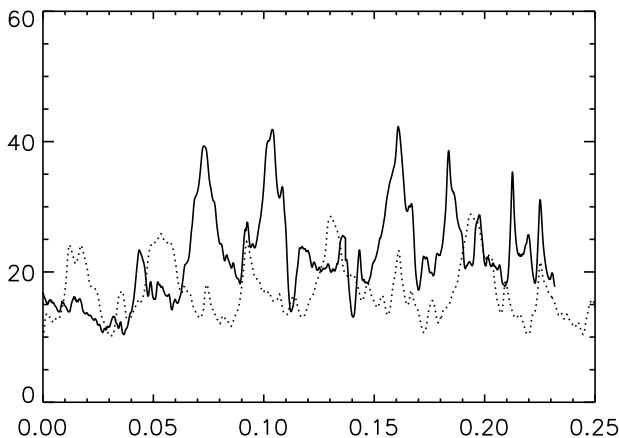


Figure 3. Benchmark (ii), $Ra = 2500$, $q = 1$, $E = 0.0005$. Solid line, plot of magnetic energy against time with no hyperdiffusion, average 22.24. Dotted line, plot of magnetic energy against time with hyperdiffusion $\alpha_V^2 = \alpha_B^2 = 0.0005$, $\alpha_T^2 = 0$, so hyperdiffusion applied to the momentum and induction equations but not the temperature equation. Average 17.46.

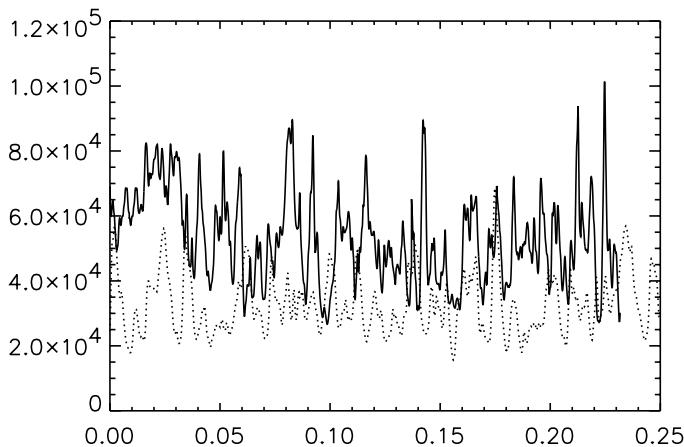


Figure 4. Benchmark (ii), $Ra = 2500$, $q = 1$, $E = 0.0005$. Solid line, plot of kinetic energy against time with no hyperdiffusion, average 53,141. Dotted line, plot of kinetic energy against time with hyperdiffusion $\alpha_V^2 = \alpha_B^2 = 0.0005$, $\alpha_T^2 = 0$, so hyperdiffusion applied to the momentum and induction equations but not the temperature equation. Average 32,266.

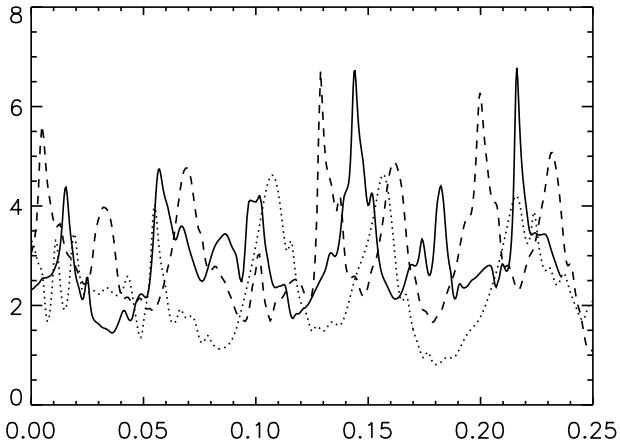


Figure 5. Benchmark (i), $Ra = 1000$, $q = 1$, $E = 0.0010$, $k_x = k_y = 2\sqrt{2}$. Solid line, plot of magnetic energy against time with no hyperdiffusion, average 2.9692. Dotted line, plot of magnetic energy against time using the Holm model with $\alpha^2 = 0.0002$ and boundary conditions (5.3)–(5.6), average 2.4200. Dashed line, plot of magnetic energy against time using the Holm model with $\alpha^2 = 0.0002$ and boundary conditions in Appendices A and B. Average 2.9193.

that it is the effect of hyperdiffusion on the thermal boundary layer that is responsible. With thermal hyperdiffusion in force, the thermal boundary layers close to $z = \pm 1/2$ are thicker, and this allows convection to be more vigorous; recall that in benchmark (ii) the Rayleigh number is well above critical. This highlights a difficulty with using hyperdiffusion when boundary layers play an important role in the solution.

In Figs. 5 and 6 we give the magnetic and kinetic energy plots for benchmark (i) and the Holm-MHD equations. From Table 2, we see that the stability of the system is reduced by introducing the unbarred \mathbf{j} into the Lorentz force, so that a comparison with benchmark (ii) would be time consuming, though preliminary runs suggest that the main features are similar for both benchmark cases. The modification of the Lorentz force term does not appear to have a great effect, and the general behavior of the non-hyperdiffusive system is maintained. The Holm-MHD equations have only hyperdiffusion in magnetic field and momentum, so we might expect the magnetic and kinetic energies to be reduced, as in the comparison between Figs. 1 and 2 and Figs. 3 and 4. Figures 5 and 6 show that this is indeed the case, and this effect seems to be stronger than any effect due to the modification of the Lorentz force term. We have also run this case with the boundary conditions given

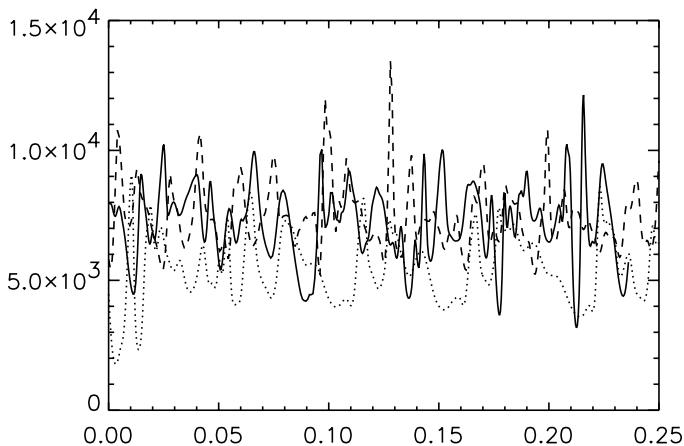


Figure 6. Benchmark (i), $Ra = 1000$, $q = 1$, $E = 0.0010$, $k_x = k_y = 2\sqrt{2}$. Solid line, plot of kinetic energy against time with no hyperdiffusion, average 7205.6. Dotted line, plot of kinetic energy against time using the Holm model with $\alpha^2 = 0.0002$ and boundary conditions (5.3)–(5.6), average 5410.6. Dashed line, plot of kinetic energy against time using the Holm model with $\alpha^2 = 0.0002$ and boundary conditions in Appendices A and B. Average 7506.5.

in Appendix A (A.5)–(A.10), and Appendix B (B.9)–(B.10). These also gave stable numerical integrations, and the change in boundary conditions did not significantly affect the timestep.

Changing the magnetic boundary conditions from (5.5), (5.6) to (A.5)–(A.10) appears not to make a great difference, but the Appendix B mechanical conditions affect the boundary layers and this increases the kinetic energy significantly and the magnetic energy somewhat. Interestingly, this compensates the drops in these quantities found when using (5.3), (5.4), so that the Appendix B boundary conditions give the best fit to the non-hyperdiffusive results when the Holm-MHD equations are used.

Finally, in Figs. 7 and 8 we have examined the behavior of hyperdiffusive runs (with all three hyperdiffusivities) for benchmark (iii), the most demanding case. We plan to do integrations of this case with no hyperdiffusion, but this will require a great many processor hours to establish the behavior definitively. By comparison, the hyperdiffusive runs can be performed quite quickly. By comparing the two cases $\alpha^2 = 0.0005$ and $\alpha^2 = 0.001$, we can extrapolate to the case of zero hyperdiffusion, and it will be an interesting test of the usefulness (or otherwise) of hyperdiffusivity to see if this predicted behavior does in fact occur when the non-hyperdiffusive equations are integrated.

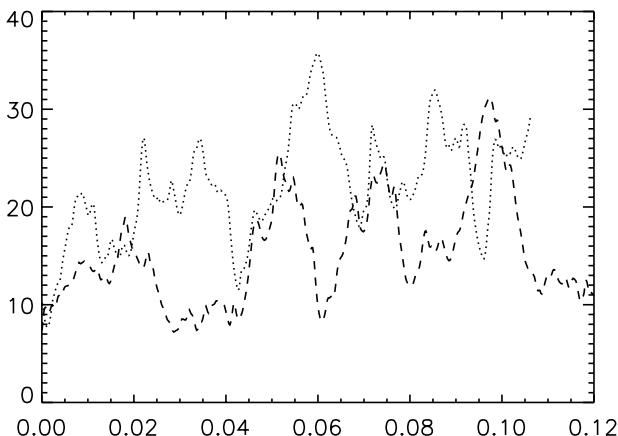


Figure 7. Benchmark (iii), $Ra = 5000, q = 0.5, E = 0.0002$. Dotted line, plot of magnetic energy against time with hyperdiffusion $\alpha_V^2 = \alpha_B^2 = 0.0005, \alpha_T^2 = 0.0005$, average 22.11. Dashed line: plot of magnetic energy against time with hyperdiffusion $\alpha_V^2 = \alpha_B^2 = \alpha_T^2 = 0.001$, average 17.59.

Another encouraging feature of the cases where all three hyperdiffusions are used is that the ratio of poloidal to toroidal components is not greatly affected. An important observation from the anisotropic hyperdiffusive simulations of Grote *et al.* (2000) was that hyperdiffusion strongly suppressed the poloidal components of the magnetic field in comparison with the toroidal components, thus changing the whole nature of the large scale magnetic field. We find no corresponding effect in these plane layer simulations.

8. Conclusions

Convection-driven dynamo models have great potential in helping us to understand the dynamics of planetary interiors, and this impacts on many areas of scientific activity. There is therefore a strong motivation for developing numerical schemes that will allow us to integrate the equations without requiring overwhelming computer resources. The use of hyperdiffusion, even when small coefficients are put in front of the hyperdiffusive terms, gives a remarkable amount of numerical stabilization, so it is a very tempting way forward. It is also true that all numerical geodynamo simulations use some form of turbulent dissipation, and it is by no means obvious that merely increasing the numerical

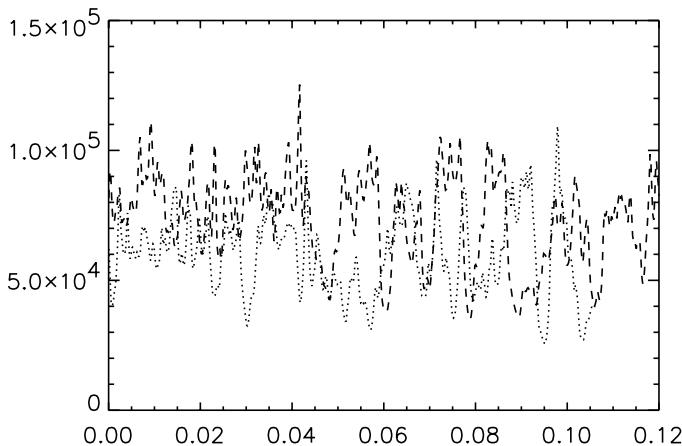


Figure 8. Benchmark (iii), $Ra = 5000$, $q = 0.5$, $E = 0.0002$. Dotted line, plot of kinetic energy against time with hyperdiffusion $\alpha_V^2 = \alpha_B^2 = 0.0005$, $\alpha_T^2 = 0.0005$, average 59,672. Dashed line: plot of kinetic energy against time with hyperdiffusion $\alpha_V^2 = \alpha_B^2 = 0.001$, $\alpha_T^2 = 0.001$, average 69,952.

value of the regular diffusion terms is the best physical representation of these terms.

Generally, the isotropic hyperdiffusion appears to give quite similar behavior to the non-hyperdiffusive benchmarks, though more work needs to be done before this can be asserted positively. Interestingly, the best results appear to be given when all the diffusive operators are treated in the same way. It appears that upsetting the ratios between the various diffusive processes has more effect than adding in hyperdiffusion to all diffusing quantities. This study has also shed some more light on the nature of the numerical instabilities that make the dynamo problem so hard to solve. Of particular note is the fact that when the magnetic field is strong, the timestep decreases by a factor of at least ten in the absence of hyperdiffusion. It is possible that this is due in part to the geostrophic modes having a rather singular behavior, in which case there may be alternative ways of stabilizing the system other than by invoking hyperdiffusion.

The issue of the boundary conditions to be used is not yet fully resolved, though we have found a set of boundary conditions which seem practical to use and yet do not radically change the nature of the solutions. The boundary conditions given in Appendices A and B are a possible alternative, and seem to give better results for the Holm-MHD equations. While our choice of boundary conditions was motivated by what we believe are sensible physical assumptions, we do not yet have

a definite philosophy leading to a unique choice for the subgrid scale boundary conditions. This lack of a proper theoretical basis is a cause for concern, especially if we wish in the future to extend these ideas to different geometries. We must also not lose sight of the fact that the boundary layers play an important role in any strongly nonlinear convecting system, and hyperdiffusion can potentially have a big effect on the nature of these boundary layers. This alone means that hyperdiffusive solutions will need to be carefully interpreted before being applied to studies of the geomagnetic field.

It will also be of interest to see whether subgrid scale models such as the Holm-MHD model portray the small scale behavior more faithfully than crude “eddy diffusivities” can do. In this study, the effects of the boundary layers have overwhelmed the effects of the modified nonlinear terms but, if the LANS α method can be adjusted to eliminate this difficulty, it may prove to be a fruitful research tool.

Acknowledgments

We thank the US National Science Foundation for support under grant EAR0222334, during the tenure of which this paper was written, and we also thank Darryl Holm for useful discussions.

References

- Andrews, D. G. and McIntyre, M. E., “An exact theory of nonlinear waves on a Lagrangian-mean flow,” *J. Fluid Mech.* **89**, 609–646 (1978a).
- Andrews, D. G. and McIntyre, M. E., “On wave-action and its relatives,” *J. Fluid Mech.* **89**, 647–664. Corrigendum: *ibid.*, **95**, 796 (1978b).
- Braginsky, S. I., “Magnetohydrodynamics of the Earth’s core,” *Geomagn. Aeron.* **4**, 698–712 (1964).
- Braginsky, S. I. and Meytis, V. P., “Local turbulence in the Earth’s core,” *Geophys. Astrophys. Fluid Dynam.* **55**, 71–87 (1990).
- Buffett, B. A., “A comparison of subgrid-scale models for large-eddy simulations of convection in the Earth’s core,” *Geophys. J. Int.* **153**, 753–765 (2003).
- Busse, F. H., “Convective flows in rapidly rotating spheres and their dynamo action,” *Phys. Fluids* **14**, 1301–1314 (2002).
- Christensen, U., Olson, P. and Glatzmaier, G. A., “Numerical modelling of the geodynamo: a systematic parameter study,” *Geophys. J. Int.* **138**, 393–409 (1999).
- Dormy, E., Valet, J.-P. and Courtillot, V., “Numerical models of the geodynamo and observational constraints,” *Geochem. Geophys. Geosyst.*, **1**, art. no. 2000GC000062 (2000).

- Eckart, C., "Some transformations of the hydrodynamic equations," *Phys. Fluids* **6**, 1037–1041 (1963).
- Fearn, D. R., Roberts, P. H. and Soward, A. M., "Convection, stability and the dynamo," in *Energy, Stability and Convection* (Eds. G. P. Galdi and B. Straughan), pp. 60–324, Longmans (1988).
- Frick, P., Reshetnyak, M. and Sokoloff, D., "Combined grid-shell approach for convection in a rotating spherical layer," *Europhys. Lett.* **59**, 212–217 (2002).
- Gjaja, I. and Holm, D. D., "Self-consistent Hamiltonian dynamics of wave mean-flow interaction for a rotating stratified incompressible fluid," *Physica D* **98**, 343–378 (1996).
- Glatzmaier, G. A. and Roberts, P. H., "A 3-dimensional self-consistent computer simulation of a geomagnetic field reversal," *Nature* **377**, 203–209 (1995).
- Glatzmaier, G. A. and Roberts, P. H., "Simulating the geodynamo," *Contemp. Phys.* **38**, 269–288 (1997).
- Grote, F. H., Busse, F. H. and Tilgner, A., "Effects of hyperdiffusivities on dynamo simulations," *Geophys. Res. Lett.* **27**, 2001–2004 (2000).
- Holm, D. D., "Averaged Lagrangians, and the mean effects of fluctuations in ideal fluid dynamics," *Physica D* **170**, 253–286 (2002).
- Jones, C. A. and Roberts, P. H., "Convection driven dynamos in a rotating plane layer," *J. Fluid Mech.* **404**, 311–343 (2000).
- Kono, M. and Roberts, P. H., "Recent geodynamo simulations and observations of the geomagnetic field," *Rev. Geophys.* **40**, art. no. 1013 (2002).
- Kuang, W. and Bloxham, J., "Numerical modeling of magnetohydrodynamic convection in a rapidly rotating spherical shell: weak and strong field dynamo action," *J. Comput. Phys.* **153**, 51–81 (1999).
- Montgomery, D. C. and Pouquet, A., "An alternative interpretation of the Holm alpha model," *Phys. Fluids* **14**, 3365–3366 (2002).
- Rotvig, J. and Jones, C. A., "Rotating convection-driven dynamos at low Ekman number," *Phys. Rev. E* **66**, art. no. 056308, 1–15 (2002).
- Soward, A. M., "A kinematic theory of large magnetic Reynolds number dynamos," *Phil. Trans. R. Soc. Lond. A* **272**, 431–462 (1972).
- Starchenko, S. and Jones, C. A., "Typical velocities and magnetic field strengths in planetary interiors," *Icarus* **157**, 426–435 (2002).
- Subramanian, K., "Hyperdiffusion in nonlinear large- and small-scale turbulent dynamos," *Phys. Rev. Lett.* **90**, art. no. 90.245003 (2003).
- Taylor, J. B. "The magneto-hydrodynamics of a rotating fluid and the Earth's dynamo problem," *Proc. R. Soc. Lond. A* **274**, 274–283 (1963).
- Walker, M. R., Barenghi, C. F. and Jones, C. A., "A note on dynamo action at asymptotically small Ekman number," *Geophys. Astrophys. Fluid Dynam.* **88**, 261–275 (1998).
- Zhang, K. and Jones, C. A., "The effect of hyperviscosity on geodynamo models," *Geophys. Res. Lett.* **24**, 2869–2872 (1997).
- Zhang, K., Jones, C. A. and Sarson, G. R., "The dynamical effects of hyperviscosity on numerical geodynamo models," *Studia geoph. et geod.* **42**, 247–253 (1998).

Appendix A: Other magnetic boundary conditions

This appendix seeks an alternative to the boundary conditions (5.5) and (5.6) for the Holm-MHD equations, one that is better motivated from a physical point of view than (5.5) and (5.6). The basic idea is that the subgrid scale magnetic fields will penetrate from the conducting fluid into the surrounding insulator in the same way as the large scale fields do, and that therefore it is necessary to consider both filtered and unfiltered magnetic fields in the insulator. We shall distinguish these from the corresponding fields in the conductor by an added circumflex.

As in (4.10), we have

$$\hat{g} = (1 - \alpha^2 \nabla^2) \bar{g}, \quad \hat{h} = (1 - \alpha^2 \nabla^2) \bar{h}. \quad (\text{A.1})$$

In the insulator, the unfiltered field is a source-free potential field, so that

$$\hat{g} = 0, \quad \nabla^2 \hat{h} = 0, \quad (\text{A.2})$$

and, by (A.1),

$$0 = (1 - \alpha^2 \nabla^2) \bar{g}, \quad 0 = \nabla^2 (1 - \alpha^2 \nabla^2) \bar{h}. \quad (\text{A.3})$$

Since there are no sources of field at infinity, the relevant solutions of (A.3) are those that vanish with distance from the conductor. It follows that, for the (l, m) components in $z \geq \frac{1}{2}$,

$$\hat{\bar{g}}_{lm} = g_0 e^{-nz}, \quad \hat{\bar{h}}_{lm} = h_0 e^{-nz} + h_1 e^{-az}, \quad (\text{A.4})$$

where, as before, $a = (l^2 k_x^2 + m^2 k_y^2)^{1/2}$ is the total wave number and $n = (a^2 + \alpha^{-2})^{1/2}$. (In (A.4), g_0 , h_0 and h_1 are constants.)

The conditions (3.19) and (3.20) on g and h follow from requiring that $\hat{g} = g$, $\hat{h} = h$ and $\hat{h}' = h'$ on $z = 1/2$. Similarly, \bar{g} , \bar{h} and \bar{h}' must be continuous there, so that (A.4) requires

$$\partial_z \bar{g}_{lm} = \mp n \bar{g}_{lm}, \quad (\pm \partial_z + a) \bar{h}_{lm} = -\alpha^2 (n - a) \nabla^2 \bar{h}_{lm} \quad \text{on } z = \pm \frac{1}{2}. \quad (\text{A.5, A.6})$$

(We have taken the opportunity here of including the conditions at $z = -\frac{1}{2}$ also.)

Conditions (A.5) are supplemented by those obtained from (3.19), (3.20) and (4.10):

$$(1 - \alpha^2 \nabla^2) \bar{g} = 0, \quad (1 - \alpha^2 \nabla^2) \bar{h}'_{lm} = \mp a (1 - \alpha^2 \nabla^2) \bar{h}_{lm}, \quad \text{on } z = \pm \frac{1}{2}. \quad (\text{A.7, A.8})$$

Conditions (A.6)–(A.8) are the alternatives to (5.5) sought. One may also argue from (4.11) that the mean filtered field in the insulator must satisfy conditions similar to (A.5):

$$\alpha \partial_z \bar{b}_x = \mp \bar{b}_x, \quad \alpha \partial_z \bar{b}_y = \mp \bar{b}_y, \quad \text{on } z = \pm \frac{1}{2}. \quad (\text{A.9, A.10})$$

Appendix B: α -layers

The relationship between \mathbf{u} and $\bar{\mathbf{u}}$ in the Camassa-Holm approach is

$$\mathbf{u} = (1 - \Delta) \bar{\mathbf{u}}, \quad (\text{B.1})$$

where

$$\Delta = \nabla_i \overline{\xi_i \xi_j} \nabla_j. \quad (\text{B.2})$$

Their Taylor ansatz allowed them to assume that the correlation in the displacement ξ is homogeneous and isotropic:

$$\overline{\xi_i \xi_j} = \alpha_0^2 \delta_{ij}, \quad \text{where} \quad \alpha_0 = \text{constant} \quad (\text{B.3})$$

(see (2.10)). Then $\Delta = \alpha_0^2 \nabla^2$ and (B.1) reduces to (2.13).

At a fixed no-slip boundary, ξ is zero, and it is smaller than α_0 in an “ α -layer” adjacent to the boundary. Evidently $\overline{\xi_i \xi_j}$ increases with distance, z , from the boundary to become $\alpha_0^2 \delta_{ij}$ as $z \rightarrow \infty$ on the boundary layer scale. This means that, as with every other situation in which there are reasons for believing that $\overline{\xi_i \xi_j}$ is a function of position, $\nabla \cdot \mathbf{u}$ and $\nabla \cdot \bar{\mathbf{u}}$ cannot simultaneously be zero. The reason for this may be traced back to (2.8) which in general does not allow both $\bar{\mathbf{u}}$ and \mathbf{u}' to be divergenceless. It seems clear that the unfiltered velocity must satisfy $\nabla \cdot \mathbf{u} = 0$ in the Boussinesq approximation, so that $\nabla \cdot \bar{\mathbf{u}} \neq 0$ in the α -layer, an unwelcome complication; it is not even possible to write $\bar{\mathbf{u}}$ in the form (3.5) in the α -layer.

The thickness of the α -layer is of order α_0 and, assuming that this is small compared with all other relevant length scales, the α -layer is a type of surface layer abutting a locally flat boundary, $z = 0$; moreover to leading order $\overline{\xi_i \xi_j}$ depends on z alone. In the generic case, $\xi_H = O(z)$ and $\xi_z = O(z^2)$ as $z \rightarrow 0$, so that $\overline{\xi_H i \xi_H j} = O(z^2)$, $\overline{\xi_H i \xi_z} = O(z^3)$ and $\overline{\xi_z^2} = O(z^4)$. (Here the suffix H refers to “horizontal”, x -, y -components.)

The Camassa-Holm equations corresponding to (3.1) and (3.5) are, in the non-magnetic, non-buoyant case,

$$\hat{\mathbf{z}} \times \bar{\mathbf{u}} = -\nabla p + E \nabla^2 \mathbf{u}, \quad \nabla \cdot \mathbf{u} = 0, \quad (\text{B.4})$$

which, after the substitution (B.1), are

$$\hat{\mathbf{z}} \times \bar{\mathbf{u}} + \nabla p - E \nabla^2 \bar{\mathbf{u}} = -E \nabla^2 \Delta \bar{\mathbf{u}}, \quad \nabla \cdot \bar{\mathbf{u}} = \nabla \cdot \Delta \bar{\mathbf{u}}. \quad (\text{B.5}, \text{B.6})$$

The α -layer structure is determined by the $O(\partial_z^3 \alpha_z^2 \partial_z \bar{\mathbf{u}})$ part of the right-hand side of (B.5), which shows that a singular solution exists that is of order $\exp(1/z)$ for $z \rightarrow 0$. The exclusion of this singularity places an implicit condition on each component of $\bar{\mathbf{u}}_H$. When the resulting two conditions are in force, the solution may be obtained conceptually by iteration. It is expanded as

$$\bar{\mathbf{u}} = \sum_{n=0}^{\infty} \alpha_0^{2n} \bar{\mathbf{u}}^{(n)}, \quad (\text{B.7})$$

substituted into (B.5), and like powers of α_0^2 are equated. The resulting equations are each solved subject to no-slip boundary conditions.

Because the singular solution has been eliminated, the procedure just adumbrated succeeds both in the mainstream (i.e., outside the α -layer) and within the α -layer itself. Because $\nabla \cdot \bar{\mathbf{u}} \neq 0$ in the α -layer however, this is an inconvenient procedure to adopt in numerical work. Instead, the focus is best placed on the mainstream solution for which $\xi_i \xi_j$ is constant. This requires boundary conditions to be applied at the “edge of the boundary layer” where $z = O(\alpha_0)$ or, in the asymptotic sense, where $z \rightarrow 0$ on the mainstream scale but $z \rightarrow \infty$ on the boundary layer scale. Because $\alpha \neq 0$ at the edge of the boundary layer, five explicit conditions are required to replace the two implicit conditions and the three explicit conditions at $z = 0$. Three explicit conditions at the edge of the boundary layer follow, via (B.1), from the no-slip conditions on \mathbf{u} at $z = 0$; they are

$$(1 + \alpha_0^2 a^2) \bar{e} - \alpha_0^2 \bar{e}'' = (1 + \alpha_0^2 a^2) \bar{f} - \alpha_0^2 \bar{f}'' = (1 + \alpha_0^2 a^2) \bar{f}' - \alpha_0^2 \bar{f}''' = 0, \quad (\text{B.8})$$

where a is the total wavenumber defined in Appendix A.

The two further conditions would ideally express in mathematical terms the absence of the singular parts of $\bar{\mathbf{u}}_H$ in the α -layer, but this idea is not easily implemented, and we have not completed the required analysis. We must therefore, as in Section 5, postulate two *ad hoc* conditions at the edge of the boundary layer. Even if it transpires that these are not equivalent to the two implicit conditions, the resulting error will be mainly confined to a spurious boundary layer of thickness α_0 at the edge of the mainstream and will not be serious elsewhere. The simplest choice is $\bar{\mathbf{u}}_H = \mathbf{0}$, i.e., $\bar{e} = \bar{f}' = 0$. By (B.8), the complete set of

five conditions to apply at the edge of the boundary layer is then

$$\bar{e} = \bar{e}'' = (1 + \alpha_0^2 a^2) \bar{f} - \alpha_0^2 \bar{f}'' = \bar{f}' = \bar{f}''' = 0. \quad (\text{B.9})$$

For the x, y independent parts of the velocity field, the corresponding choice is

$$\overline{U}_x = \overline{U}_x'' = \overline{U}_y = \overline{U}_y'' = 0. \quad (\text{B.10})$$

11 Planetary and stellar dynamos: challenges for next generation models

Gary A. Glatzmaier

*Earth Sciences Department, University of California,
Santa Cruz, CA 95064, USA*

Three-dimensional (3D) global dynamo models numerically simulate convection and magnetic field generation in the fluid interiors of planets and stars. Geodynamo models, for example, currently produce large-scale magnetic fields that, outside the core, resemble the Earth's surface field to first order. However, no 3D global model has yet been able to afford the spatial resolution required to simulate a broad spectrum of turbulence, which surely exists in planetary and stellar interiors. A critical question is: How realistic are the laminar flows in the interiors of the models that generate these simulated fields? We show, via a series of two-dimensional (2D) magnetoconvection calculations, how the structure and time dependence of even the large-scales change dramatically when the solution becomes strongly turbulent. These exploratory results highlight several issues that will need to be addressed when designing the next generation of global dynamo models.

1. Introduction

For two decades now, 3D computer simulations have been used to study how magnetic fields are generated by convection in the deep interiors of stars and planets. Models of the geodynamo have become particularly popular in recent years. They have produced dipole dominated fields outside the core and westward drift of the non-dipolar field, similar to that observed on the Earth's surface. Many of these models show an eastward rotation of the solid inner core relative to the surface as inferred from some seismic analyses. Some models also occasionally produce spontaneous dipole reversals on a time scale not unlike that

seen for the paleomagnetic record. These results from magnetohydrodynamic (MHD) simulations have given some people the impression that the “geodynamo problem” is solved.

However, the spectrum of turbulence in the Earth’s core spans many orders of magnitude, from the global scale to the scale at which molecular viscosity becomes important. Models of the geodynamo attempt to resolve as much of the large-scale end of this spectrum as can be afforded with current computational resources. The energy in a simulation that cascades down to the truncated scale is diffused away by an artificially large scalar viscosity. That is, the transport and mixing by the rest of the spectrum, the “subgrid scale” turbulence, is crudely modeled as a diffusion process that is isotropic and homogeneous. This “turbulent viscosity” used in current geodynamo simulations is roughly ten orders of magnitude greater than the molecular viscosity of the Earth’s fluid core. Even so, viscous forces in these simulations are relatively small, typically four orders of magnitude smaller than the Coriolis and Lorentz forces. Therefore the argument has been made by those of us in the geodynamo community that maybe viscosity in the current laminar simulations is asymptotically small enough for our models to capture the correct large-scale dynamics of the geodynamo. This seems to be true for “large eddy simulations” of the global circulations in the Earth’s atmosphere and ocean. Also, measurements of the depths of turbulent boundary layers in the atmosphere and ocean (e.g., Hunkins, 1966) show that they are orders of magnitude deeper than a laminar molecular Ekman boundary depth would be. Indeed, the current geodynamo models have produced relatively realistic energy spectra and time dependencies of the surface magnetic fields, which seems to justify the use of low spatial resolution and greatly enhanced viscosity.

After briefly reviewing some results from the Glatzmaier-Roberts geodynamo simulations, we question if we are getting the right answer for the wrong reason. Unlike the Earth’s atmosphere and ocean, the dynamics of the fluid core can only be crudely inferred from afar, for example, via magnetic measurements at the surface of the Earth where only the very large length and timescales of the core’s field are detected. Are the current geodynamo models correctly simulating the internal dynamics of the core to first order, or are these models just demonstrating that dipole dominated fields can be produced by many different laminar flow structures in model cores? We try to get some insight on this issue by comparing 2D magnetoconvection simulations in the parameter regime of current 3D geodynamo simulations with those in a much more realistic parameter regime that we would hope to someday be able to reach in 3D. The effects of compressibility are of

particular interest, not only for convection that spans a large density stratification, as exists in stars and giant gaseous planets, but also for the small density stratification within the Earth's core. This exercise highlights some of the real challenges facing the next generation of stellar and planetary dynamo models.

2. Current 3D geodynamo simulations

The geomagnetic field, which has likely existed for most of the age of the Earth, is maintained by convection in the electrically conducting fluid outer core. As the core cools, iron in the iron-alloy fluid preferentially plates onto the solid inner core releasing latent heat and leaving behind a higher concentration of light elements. These two main sources of buoyancy, thermal and compositional, at the inner core boundary (ICB) drive convection in the fluid outer core. Additional sources exist due to secular cooling of the fluid core and possibly internal heating due to radioactive decay of potassium-40.

The trajectories of the buoyant fluid parcels, viewed in the rotating frame of the Earth, are significantly influenced by Coriolis forces because the fluid viscosity is very small, possibly as small as that for water, but with a density ten times greater. The twisting and shearing flow of this electrically conducting fluid against existing magnetic field generates electric currents with new magnetic field. However, the details of how this maintains the intensity, structure and time dependencies of the observed geomagnetic field, including occasional dipole reversals, requires the solution of a coupled set of nonlinear equations that describe the 3D evolution of the thermodynamic variables, the fluid velocity and the magnetic field.

The first dynamically consistent 3D global simulations of a convective dynamo were designed to study the solar dynamo (Gilman and Miller, 1981 (Boussinesq); Glatzmaier, 1984 (anelastic)). Zhang and Busse (1988) developed a 3D model to study the onset of dynamo action within the Boussinesq approximation. Nearly a decade later a variety of geodynamo models, employing various approximations and run in different parameter regimes, were developed by several groups (e.g., Kageyama *et al.*, 1995; Jones *et al.*, 1995; Glatzmaier and Roberts, 1995a; Kuang and Bloxham, 1997; Sarson *et al.*, 1997; Kida *et al.*, 1997; Busse *et al.*, 1998; Christensen *et al.*, 1998; Sakuraba and Kono, 1999; Katayama *et al.*, 1999; Hollerbach, 2000). Several review articles describe and compare these models (e.g., Hollerbach, 1996; Glatzmaier and Roberts, 1997; Fearn, 1998; Busse, 2000; Roberts and

Glatzmaier, 2000; Dormy *et al.*, 2000; Christensen *et al.*, 2001; Busse, 2002; Glatzmaier, 2002; Kono and Roberts, 2002).

Since none of these models can afford to run in a realistic parameter regime for the Earth's core, they all make approximations (Glatzmaier, 2002). Some features of the simulated fields are robust, like the dipole dominated field well outside the core. Other features, like the structure and time dependence of the temperature, flow and field inside the core, depend on the chosen boundary conditions, parameter space and numerical resolution. Here we briefly describe our anelastic geodynamo model.

The anelastic approximation, like the Boussinesq approximation, filters out acoustic waves; effectively the sound speed is infinite so the pressure distribution at each numerical timestep is in equilibrium throughout the fluid core. This approximation is valid when the fluid velocity is small relative to the local sound speed and the thermodynamic perturbations are small relative to their reference state values; these conditions are well satisfied for the Earth's liquid core and for the deep interiors of stars and giant planets. The main advantage of these approximations over the fully compressible set of equations is that the numerical timestep can be typically a million times larger. The additional advantage of the anelastic approximation over the Boussinesq approximation is that the effects of compressibility in a density stratified fluid are represented; that is, rising fluid expands and sinking fluid contracts. This is certainly important for models of giant planets and stars. Most models of the geodynamo, however, employ the simpler Boussinesq approximation because the density at the ICB is only about 20% greater than at the core-mantle boundary (CMB). We show here, though, that even a 20% density stratification has a major effect on the flow structure, and consequently also on the field generation, when the turbulent viscosity is significantly reduced below current model values.

The following anelastic MHD equations (Glatzmaier, 1984; Braginsky and Roberts, 1995) describe the 3D, time-dependent perturbations relative to a no-flow, non-magnetic, radially dependent reference state (the “barred” variables). This reference state of our fluid outer core is fitted to the *Preliminary Reference Earth Model* (PREM) and is constrained to be a hydrostatic, adiabatic, well-mixed binary alloy:

$$\nabla \cdot (\bar{\rho} \mathbf{v}) = 0, \quad \nabla \cdot \mathbf{B} = 0, \quad \nabla^2 U = 4\pi G \rho, \quad (1,2,3)$$

$$\rho = \left[\left(\frac{\partial \bar{\rho}}{\partial S} \right)_{\xi,p} S + \left(\frac{\partial \bar{\rho}}{\partial \xi} \right)_{S,p} \xi + \left(\frac{\partial \bar{\rho}}{\partial p} \right)_{S,\xi} p \right] = \left[C \bar{\rho} + \left(\frac{\partial \bar{\rho}}{\partial p} \right)_{S,\xi} p \right], \quad (4)$$

$$\bar{\rho} \frac{\partial \mathbf{v}}{\partial t} = -\nabla \cdot (\bar{\rho} \mathbf{v} \mathbf{v}) - \bar{\rho} \nabla (p/\bar{\rho} + U) - C \bar{\rho} \bar{g} \hat{\mathbf{r}} + 2\bar{\rho} \mathbf{v} \times \Omega + \nabla \cdot \left\{ 2\bar{\rho} \bar{v} \left[\vec{\mathbf{e}} - \frac{1}{3}(\nabla \cdot \mathbf{v}) \vec{\delta} \right] \right\} + \frac{1}{\mu} (\nabla \times \mathbf{B}) \times \mathbf{B}, \quad (5)$$

$$\frac{\partial \mathbf{B}}{\partial t} = \nabla \times (\mathbf{v} \times \mathbf{B}) - \nabla \times (\bar{\eta} \nabla \times \mathbf{B}), \quad (6)$$

$$\begin{aligned} \bar{\rho} \frac{\partial S}{\partial t} &= -\nabla \cdot (\bar{\rho} S \mathbf{v}) + \nabla \cdot (\bar{\rho} \bar{\kappa}^S \nabla S) + \frac{1}{\bar{T} r^2} \frac{d}{dr} \left(r^2 c_p \bar{\rho} \bar{\kappa}^T \frac{d \bar{T}}{dr} \right) \\ &\quad + \frac{\bar{\eta}}{\mu \bar{T}} |\nabla \times \mathbf{B}|^2 + \frac{\bar{g}}{\bar{T}} \left[\bar{\kappa}^S \left(\frac{\partial \bar{\rho}}{\partial S} \right)_{\xi, p} \frac{\partial S}{\partial r} + \bar{\kappa}^\xi \left(\frac{\partial \bar{\rho}}{\partial \xi} \right)_{S, p} \frac{\partial \xi}{\partial r} \right], \end{aligned} \quad (7)$$

$$\bar{\rho} \frac{\partial \xi}{\partial t} = -\nabla \cdot (\bar{\rho} \xi \mathbf{v}) + \nabla \cdot (\bar{\rho} \bar{\kappa}^\xi \nabla \xi). \quad (8)$$

These equations are solved to obtain the 3D time-dependent fluid flow \mathbf{v} , magnetic field \mathbf{B} , and perturbations in density ρ , pressure p , specific entropy S , light constituent mass fraction ξ , and gravitational potential U , relative to the reference state in a frame of reference rotating with the Earth's angular velocity Ω . Here, $\vec{\mathbf{e}}$ is the rate of strain tensor, μ is the magnetic permeability of free space, \bar{v} , $\bar{\kappa}^S$ and $\bar{\kappa}^\xi$ are the turbulent viscous, thermal and compositional diffusivities, $\bar{\kappa}^T$ and $\bar{\eta}$ are the actual thermal and magnetic diffusivities and $\vec{\delta}$ is the unit tensor. We are using the co-density (C) formulation of Braginsky and Roberts (1995), which nicely, without any approximations other than being anelastic, combines the pressure gradient, the gravitational potential perturbation and the pressure contribution to buoyancy all into a gradient of the reduced pressure, $(p/\bar{\rho} + U)$, leaving the actual driving force to be only the buoyancy due to entropy and compositional perturbations. These anelastic equations describe (1) mass conservation, (2) magnetic flux conservation, (3) the perturbation of the gravitational potential, (4) the equation of state, (5) momentum conservation, (6) magnetic induction, (7) heat and (8) composition.

Our thermal boundary conditions at the ICB constrain the local flux of latent heat to be proportional to the local flux of light constituent and to the local cooling rate (Glatzmaier and Roberts, 1996a). Consequently, neither the temperature nor the heat flux at the ICB is specified; they are part of the solution, varying with location and time. We prescribe the heat flux at the CMB, which controls the cooling rate of the core and therefore the production rate of buoyancy sources at the ICB and so ultimately the intensity of the magnetic field. For our geodynamo simulations we set the total heat flow out of the core to 7.2 TW; of this,

5 TW is due to heat flow conducted down the adiabat. There is no flux of light constituent through the CMB in our model.

The ICB and CMB in the model are impermeable and either non-slip or viscously stress free. The solid inner core and solid mantle rotate in reaction to magnetic torques, viscous torques if non-slip boundary conditions are applied, and gravitational torques if the gravitational forces between the mantle and the inner core are included (Buffett and Glatzmaier, 2000). The total angular momentum of the inner core, fluid core and mantle remains zero in the rotating frame, but the angular momentum of each of these is time dependent.

The magnetic field is generated in the model's fluid outer core and diffuses into its conducting solid inner core. A thin conducting layer exists in the model above the fluid core, at the base of the mantle, allowing electric currents and therefore magnetic torque between the core and mantle. Above this layer the mantle is assumed to be an insulator and so the external magnetic field is a potential field, albeit time dependent.

Equations (1)–(8) are solved using a spectral method (spherical harmonic and Chebyshev polynomial expansions) that treats the linear terms implicitly and the nonlinear terms explicitly (Glatzmaier, 1984; Glatzmaier and Roberts, 1996a). Simulations that span hundreds of thousands of years, tens of millions of numerical timesteps, are run at very low spatial resolution (33 radial, 32 latitudinal, 64 longitudinal levels) with hyperdiffusion to damp the small scales more than the large scales. Simulations at higher resolution (289 radial, 384 latitudinal, 384 longitudinal levels), without hyperdiffusion, are now run on massively parallel computers to resolve more of the energy spectrum; but, of course, the higher the spatial resolution the less time can be simulated for a given amount of computing resources.

A typical snapshot of our simulated magnetic field is illustrated in Fig. 1 with a set of field lines ([see color insert](#) following page 234). Helical fluid flow twists toroidal (east-west) field into poloidal (meridional) field. The poloidal field extends beyond the core where it is significantly weaker and dominantly dipolar, not unlike the geomagnetic field. Within the fluid core, poloidal magnetic field is sheared into toroidal field; this occurs mainly on the cylinder tangent to the inner core equator and on the ICB.

The fluid flow in this region has a “thermal wind” component that, near the inner core, is predominantly eastward relative to the mantle. Magnetic field, which permeates both this flow and the inner core, drags the inner core in the direction of the flow (Glatzmaier and Roberts, 1996b; Aurnou and Olson, 2000). The original super-rotation rate of the inner core predicted by a geodynamo simulation (Glatzmaier and

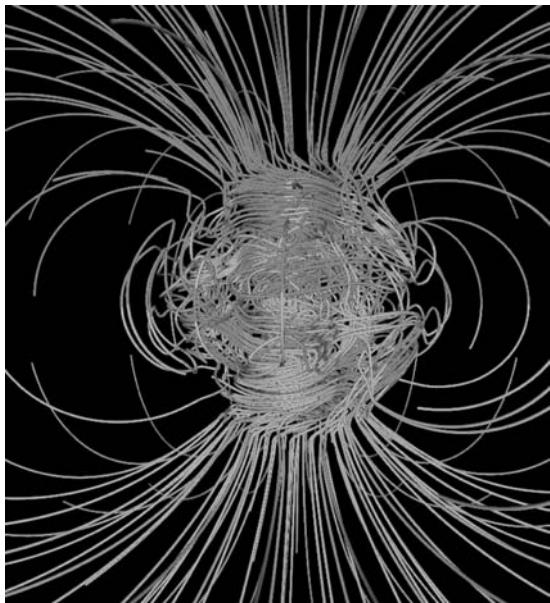


Figure 1. Snapshot of the 3D magnetic field structure for a Glatzmaier-Roberts geodynamo simulation, illustrated with a set of magnetic lines of force, which are blue where directed inward and gold where directed outward. (See color insert.) The axis of rotation is vertical and centered in the image. The lines are drawn out to two Earth radii. The field is a smooth, dipole-dominated, potential field outside the core. (Reprinted with permission from the Annual Review of Earth and Planetary Sciences, Volume 30 ©2002 by Annual Reviews, www.annualreviews.org.)

Roberts, 1995a) was about 2° longitude per year relative to the surface. In a higher resolution simulation the magnetic torque tends to be balanced by gravitational torque between the inner core and mantle (Buffett and Glatzmaier, 2000), which opposes rotational displacement of the inner core relative to the mantle. The model includes a simple parameterization for the deformation of the surface layer of the inner core relative to the bulk of the inner core which rotates below it. Figure 2 shows the smaller resulting time-dependent angular rotation rate of the inner core during a short interval of a simulation. Most of the time the inner core is rotating slightly faster than the mantle; but the rate is highly variable. The average inner core super-rotation rate produced by this model, however, depends on the assumed value of the very poorly constrained viscosity of the surface layer of the “solid” inner core. The super-rotation rate of the Earth’s inner core has been inferred from several seismic analyses (e.g., Song and Richards, 1996; Su *et al.*, 1996;

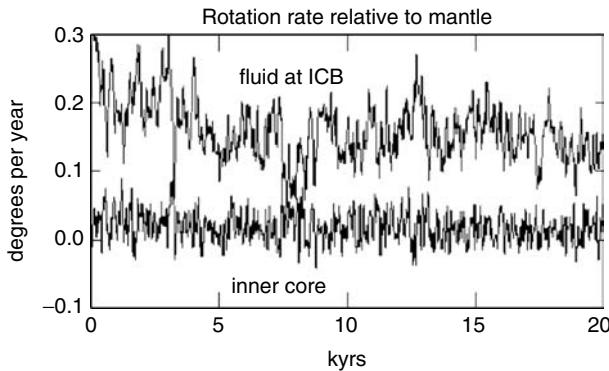
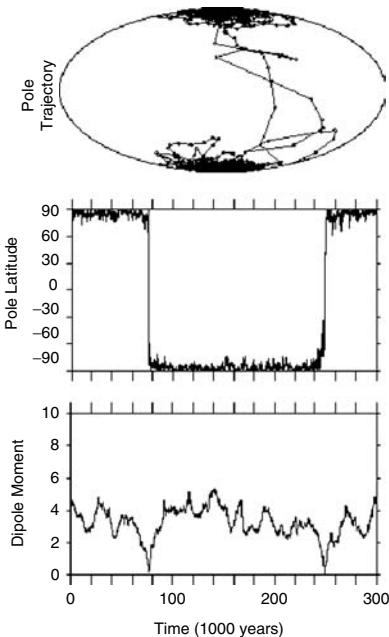


Figure 2. The angular velocity of the inner core and the fluid just above the ICB for a 20,000-year interval of a geodynamo simulation. Positive is eastward relative to the mantle. (Reproduced with permission of American Geophysical Union, from Buffett, B. A. and Glatzmaier, G. A., “Gravitational braking of inner-core rotation in geodynamo simulations,” *Geophys. Res. Lett.* **27**, 3125–3128 (2000), copyright by the American Geophysical Union.)

Xu and Song, 2003), but is still uncertain (e.g., Creager, 1997; Souriau, 1998; Laske and Masters, 1999). Currently there is a spread in the values, from the initial estimates of 1° to 3° east per year (relative to the Earth’s surface) to some that are zero to within an uncertainty of 0.2° per year.

On a much larger timescale, geodynamo simulations produce spontaneous magnetic dipole reversals (e.g., Glatzmaier and Roberts, 1995b; Sarson and Jones, 1999; Kageyama *et al.*, 1999; Glatzmaier *et al.*, 1999; Kutzner and Christensen, 2002). The time between reversals is measured in hundreds of thousands of years; whereas the time to complete a reversal is typically a few thousand years, less than a magnetic diffusion time. These times are not unlike those seen in the paleomagnetic record. On an even longer timescale the frequency of reversals seen in the paleomagnetic record varies. The frequency of reversals in geodynamo simulations has been found to depend on the pattern of heat flux imposed over the CMB (Glatzmaier *et al.*, 1999) and on the vigor of the convection (Kutzner and Christensen, 2002). The dipole reversals seen in these simulations are spontaneous and, as in the paleomagnetic reversal record, are not periodic. Dynamo simulations with less convective driving can produce magnetic fields that continuously reverse as a periodic “dynamo wave”, as seen in the early solar dynamo simulations (Gilman, 1983) and more recently in geodynamo simulations (e.g., Kida *et al.*, 1997; Kida and Kitauchi, 1998; Grote *et al.*, 1999, 2000; Simitev and Busse, 2002).

Figure 3. 300,000 years of a geodynamo simulation illustrating the evolution of the magnetic dipole (outside the core) in terms of the south magnetic pole trajectory on an Aitoff-Hammer projection of the entire surface, its latitude and its dipole moment (in units of 10^{22} Am 2) versus time (in units of 1000 years). The tic marks on the time axis are at intervals of 20,000 years, one dipole magnetic diffusion time. (Reproduced with permission of *Nature*, from figure 1, Glatzmaier, G. A., Coe, R. S., Hongre, L. and Roberts, P. H., "The role of the Earth's mantle in controlling the frequency of geomagnetic reversals," *Nature* **401**, 885–890 (1999), copyright by Nature, www.nature.com.)



A simulation with two spontaneous reversals is illustrated in Fig. 3. The geographic location of the south magnetic pole, outside the core, its latitude and the dipole moment are plotted during 300,000 years of a 500,000-year simulation. The average numerical time-step in this simulation is 15 days. As seen in the paleomagnetic record, the intensity of the field typically decreases by at least an order of magnitude during a reversal. Hydromagnetic instabilities are continually trying to reverse the field polarity but only after many attempts does one succeed. One simulated reversal is portrayed in Fig. 4 with four snapshots spanning about 15,000 years. The radial component of the field is shown at both the CMB and the surface of the model Earth. In addition, the longitudinally averaged poloidal and toroidal parts of the field inside the core are plotted at these times. Although when viewed at the model's surface the reversal appears complete by the third snapshot, another three thousand years are required for the original field polarity to decay out of the inner core and the new polarity to diffuse in. More often however a new polarity is not able to survive and the original polarity recovers. This is likely what happens during "events" (Lund *et al.*, 1998) when the paleomagnetic field reverses and then reverses back, all within about ten thousand years (Gubbins, 1999, Glatzmaier *et al.*, 1999).

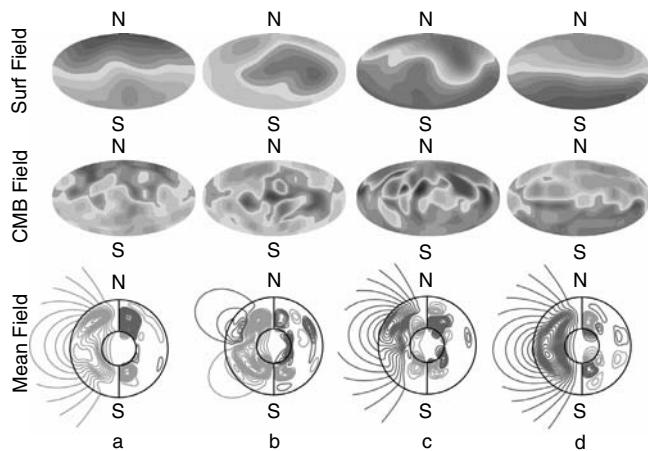


Figure 4. A sequence of snapshots of the longitudinally averaged magnetic field through the interior of the core and of the radial component of the field at the core-mantle boundary and at what would be the surface of the Earth, displayed at roughly 3000-year intervals spanning a dipole reversal from a geodynamo simulation. (See color insert.) In the plots of the average field, the small circle represents the inner core boundary and the large circle is the core-mantle boundary. The poloidal field is shown as magnetic field lines on the left-hand sides of these plots (blue is clockwise and red is counterclockwise). The toroidal field direction and intensity are represented as contours (not magnetic field lines) on the right-hand sides (red is eastward and blue is westward). Aitoff-Hammer projections of the entire core-mantle boundary and surface are used to display the radial field (with the two different surfaces displayed as the same size). Reds represent outward directed field and blues inward field; the surface field, which is typically an order of magnitude weaker, was multiplied by 10 to enhance the color contrast. (Reproduced with permission of *Nature*, from figure 2, Glatzmaier, G. A., Coe, R. S., Hongre, L. and Roberts, P. H., “The role of the Earth’s mantle in controlling the frequency of geomagnetic reversals,” *Nature* **401**, 885–890 (1999), copyright by Nature, www.nature.com.)

Other geodynamo models produce somewhat similar results even though the models differ in several respects. For example, different thermal, velocity and magnetic boundary conditions are sometimes imposed, different parameter regimes are chosen, the Boussinesq instead of the anelastic approximation is used, compositional buoyancy is neglected, the inner core may be treated as an insulator instead of a conductor. In addition, some employ hyperdiffusion, as we have, to be able to lower the diffusion on the large spatial scales (Glatzmaier, 2002); whereas others use scale-independent diffusivities. However, no model of the geodynamo has yet been run with a turbulent viscous diffusivity as small as even the Earth’s magnetic diffusivity, $2 \text{ m}^2/\text{s}$. That is,

the Ekman number, $Ek = \nu/2\Omega D^2$ (D is the depth of the fluid core), for all geodynamo simulations has been at least four orders of magnitude larger than the Earth's magnetic Ekman number, $Em = \eta/2\Omega D^2$. Based on rough estimates of the flow velocities and length scales, the turbulent Ek for geodynamo simulations should be no larger than the Earth's Em , 10^{-9} . Although all these models produce dipole dominated fields at the surface of the model Earth, the detailed structure and time dependencies of the flow and field inside the model cores differ (Dormy *et al.*, 2000; Glatzmaier, 2002) as one would expect for different model specifications.

3. Laminar approximations of turbulent flows

How well do the dynamo mechanisms simulated in the models represent the actual magnetohydrodynamics of the Earth's core? This is the fundamental question. Many of us have suggested that the models do capture enough of the energy spectrum and that Ekman numbers of order 10^{-4} and 10^{-5} are asymptotically small enough that the large (global) scales of the simulations should be fairly realistic. As computing resources improve, we will be able to resolve more of the spectrum, reduce the turbulent diffusivities and produce turbulent simulations that will begin to answer this fundamental question about our current geodynamo simulations.

Since the goal is to reduce diffusion, it is helpful to characterize our simulations with nondimensional numbers that do not involve diffusivities. As recognized a quarter of a century ago (Gilman, 1977), a nondimensional number that conveniently estimates buoyancy effects relative to Coriolis effects, without involving the presumably smaller effects of thermal and viscous diffusivities, is the convective Rossby number

$$Ro_c = \left(\frac{g\alpha\Delta T}{D} \right)^{1/2} / 2\Omega.$$

This number is also proportional to the ratio of the rotation period to free fall time across D for a gravitational acceleration, g , that is reduced by the relative density perturbation, $\alpha\Delta T$. Here α is the thermal expansion coefficient and ΔT is the superadiabatic temperature drop

across D . The $\alpha\Delta T$ term could be replaced with

$$-\left[\left(\frac{\partial \bar{\rho}}{\partial S}\right)_{\xi,p} \Delta S + \left(\frac{\partial \bar{\rho}}{\partial \xi}\right)_{S,p} \Delta \xi\right] / \bar{\rho}$$

at, for example, mid-depth for anelastic models.

Gilman (1977) found that when this number is significantly larger than unity a differential rotation profile, with the equatorial region rotating more slowly than higher latitudes, is maintained in a convecting fluid; the opposite occurs when the convective Rossby number is significantly less than unity. That is, when buoyancy dominates, fluid tends to conserve its angular momentum, resulting in a lower angular velocity in the equatorial region; whereas when the Coriolis force dominates, the tilted convection cells in the spherical shell produce a convergence of the nonlinear flux of angular momentum in the equatorial region, which maintains a higher angular velocity there.

The convective Rossby number is still used in the solar physics and computational fluid dynamics communities (e.g., Glatzmaier and Gilman, 1982; Cabot and Pollack, 1992; Julien *et al.*, 1996; Miesch *et al.*, 2000). It equals $(Ra/Pr)^{1/2}Ek$, where Ra is the classic Rayleigh number and Pr is the Prandtl number, ν/κ . Ro_c is the square root of what some authors in the geodynamics community (e.g., Christensen, 2002) have recently been calling a modified Rayleigh number. When ΔT is the *subadiabatic* temperature difference across D for a stable region this number is the *Rossby Radius of Deformation* divided by the depth of the stable layer (e.g., Pedlosky, 1987) and equals the buoyancy (*Brunt-Väisälä*) frequency divided by the Coriolis frequency, 2Ω . When this number is squared it is called the Burger number (Pedlosky, 1987).

One can likewise estimate magnetic effects relative to rotational effects without involving diffusivities by constructing a magnetic Rossby number as the ratio of an Alfvén frequency to the Coriolis frequency

$$Ro_m = \left(\frac{B_o}{(\rho\mu)^{1/2}D} \right) / 2\Omega,$$

where B_o is a typical magnetic field intensity and $\rho\mu$ is taken at mid-depth. This number is also proportional to the ratio of the rotation period to the Alfvén travel time across D . It equals $(Q/Pr q)^{1/2}Ek$, where Q is the Chandrasekhar number and q is the Roberts number, κ/η .

When the diffusivities are small enough these nondimensional Rossby numbers provide an estimate for the relative significance of the buoyancy, Coriolis and Lorentz forces in dynamo simulations.

However, how small do the diffusivities need to be? Because of their relatively large values in current dynamo models, the fluid flow in dynamo simulations has been laminar. Mildly turbulent convection, however, has recently been simulated by Miesch *et al.* (2000) with a 3D anelastic model to study solar convection. Their results show that, as the viscous and thermal diffusivities are reduced and the flow becomes turbulent, the convergence of the nonlinear flux of angular momentum changes. As a result, the “constant-on-cylinders” profile of differential rotation that was originally seen in the low-resolution simulations of the solar dynamo (Gilman and Miller, 1981; Glatzmaier, 1984) becomes more like the “constant-on-radii” profile inferred from helioseismology (Thompson *et al.*, 1996). The original solar dynamo simulations are characterized by large laminar convection cells, not unlike those seen in current geodynamo simulations. Turbulent convection simulations, however, demonstrate how the dynamics of small-scale plumes, after they detach from the boundary layers, is dominated by vortex-vortex interactions in the bulk of the fluid (e.g., Julien *et al.*, 1996; Brummell *et al.*, 1996, 1998). Both the laminar and mildly turbulent solar simulations produce a surface zonal wind profile that matches the solar surface observations; however, the original solar dynamo simulations, which could only afford low spatial resolution at the time, apparently got the right surface zonal wind profile for the wrong reason. In light of this, how robust are the current geodynamo simulations, all of which are very laminar?

4. 2D Turbulent magnetoconvection simulations

To gain some insight to what the answer might be to this question, we examine a series of 2D simulations of magnetoconvection within the anelastic approximation. Although the 2D constraint precludes realistic global flows and a dynamo mechanism, it does allow us to run at much greater spatial resolution and therefore at much smaller viscous and thermal diffusivities. That is, we are able to run in a more realistic parameter regime, very high Rayleigh numbers and very low Ekman numbers, that we would like to be able to reach someday in 3D.

Two-dimensional magnetoconvection has been studied via numerical simulation for many years (e.g., Weiss, 1981a,b). These studies have helped us understand how thermal convection of an electrically conducting fluid is affected by a magnetic field and vice versa (Proctor, 2004, Chapter 8 in this volume). However, most have been conducted

for relatively low Rayleigh numbers and so have been studies of laminar flows. Turbulent, high Rayleigh number simulations of non-magnetic convection have been conducted (e.g., Werne, 1993; Vincent and Yuen, 2000), but these, like the low Rayleigh number magnetocouvection simulations, have employed the Boussinesq approximation and therefore have not explored the effects of density stratification. In addition, for the Boussinesq approximation in 2D, Coriolis forces are completely balanced by part of the pressure field and so have no effect on the dynamics.

There are some 2D convection studies, however, that have included the effects of density stratification. Fully compressible 2D simulations of laminar convection have been conducted by Hurlburt *et al.* (1986); and 2D turbulent convection has been simulated with anelastic models by Rogers *et al.* (2003) and Evonuk and Glatzmaier (2004). Here we present additional 2D anelastic simulations of strongly turbulent magnetocouvection in a rotating, density-stratified fluid in a parameter regime that has been suggested as a grand challenge for the next generation of 3D dynamo models.

4.1. The model

We specify a box of fluid with length L in the horizontal direction (x), depth D in the vertical direction (z) and no flow, field or variations in the y direction. There is a constant gravitational acceleration, $-g\hat{\mathbf{z}}$, and a rotation rate, $\Omega\hat{\mathbf{y}}$, of the rotating frame of reference; the constant centrifugal acceleration is absorbed in g . This box can be considered a small part of the equatorial plane, with D being small relative to the distance to the axis of rotation. The top and bottom boundaries are impermeable, stress free and at constant entropy with a specified drop in specific entropy, ΔS , across the depth D . The magnetic field at the top and bottom boundaries is constrained to have no horizontal component, i.e., these boundaries are magnetically stress free. The total magnetic flux (per unit length in the y direction) through any constant z -level is B_oL , where B_o is the intensity of the externally applied, uniform, background field in the z direction. The side boundaries are periodic.

We employ a 2D version of the anelastic equations (1)–(7) and choose a reference state that is adiabatic and hydrostatic with no flow relative to the rotating frame and no magnetic field. We prescribe it as a polytrope such that the reference state pressure, \bar{p} , and density, $\bar{\rho}$, are

$$\bar{p}(z) = p_o \left(1 - \frac{z}{z_o}\right)^{n+1}$$

and

$$\bar{\rho}(z) = \rho_o \left(1 - \frac{z}{z_o}\right)^n,$$

where n is the polytropic index and z_o is a constant ($> D$) that determines the density stratification. It is related to the number of density scale heights, N_ρ , across D by

$$N_\rho = -n \ln \left(1 - \frac{D}{z_o}\right).$$

In the Boussinesq limit, $z_o \rightarrow \infty$, $N_\rho \rightarrow 0$; in the large stratification limit, $z_o \rightarrow D$, $N_\rho \rightarrow \infty$. The reference state specific entropy is a constant and the reference state (adiabatic) temperature gradient is

$$\frac{d\bar{T}}{dr} = -g\bar{\alpha}\bar{T}/c_p.$$

We assume a constant specific heat capacity at constant pressure, c_p , and a thermal expansion coefficient, $\bar{\alpha} = \bar{T}^{-1}$, so

$$\bar{T}(z) = T_o \left(1 - \frac{z}{z_o}\right).$$

Here, p_o , ρ_o and T_o are the values at the bottom boundary, $z = 0$. The local inverse density and temperature scale height functions are

$$h_\rho = \frac{d}{dz}(\ln \bar{\rho}) = -\frac{n}{(z_o - z)} \quad \text{and} \quad h_T = \frac{d}{dz}(\ln \bar{T}) = -\frac{1}{(z_o - z)}$$

respectively.

To solve the system of perturbation equations (1)–(7), we define a stream function and vector potential and use the method that has been employed in many previous 2D Boussinesq studies (e.g., Weiss, 1981a,b), but with the following modification for our anelastic model (Rogers *et al.*, 2003; Evonuk and Glatzmaier, 2004). Our streamfunction, Ψ , is defined such that

$$\bar{\rho}\mathbf{v} = \nabla \times \Psi \hat{\mathbf{y}},$$

which ensures mass conservation (1). The magnetic vector potential, A , is defined such that

$$\mathbf{B} = \nabla \times A \hat{\mathbf{y}},$$

which ensures magnetic flux conservation (2).

Taking the curl of the momentum equation (5) gives the vorticity equation

$$\frac{\partial \omega}{\partial t} = -\mathbf{v} \cdot \nabla \omega + (2\Omega + \omega)h_\rho v_z - \frac{g}{c_p} \frac{\partial S}{\partial x} + v \nabla^2 \omega + \frac{1}{\bar{\rho}} (\mathbf{B} \cdot \nabla J - h_\rho J B_z), \quad (9)$$

where ω is the vorticity, $\omega = \nabla \times \mathbf{v}$, and J is the electric current density, $\mathbf{J} = \mu^{-1} \nabla \times \mathbf{B}$. In this 2D problem, $\omega = \omega \hat{\mathbf{y}}$ and $\mathbf{J} = J \hat{\mathbf{y}}$, like $\Psi \hat{\mathbf{y}}$ and $A \hat{\mathbf{y}}$, are all in the y -direction and their amplitudes are related by

$$\omega = -\frac{1}{\bar{\rho}} \left(\nabla^2 \psi - h_\rho \frac{\partial \Psi}{\partial z} \right) \quad \text{and} \quad J = -\frac{1}{\mu} \nabla^2 A.$$

We have assumed a constant turbulent viscous diffusivity, v , and have dropped the additional viscous diffusion terms that would appear due to $\bar{\rho}(z)$ if we were using a molecular viscosity.

Uncurling the magnetic induction equation (6) gives us

$$\frac{\partial A}{\partial t} = -\mathbf{v} \cdot \nabla A + \eta \nabla^2 A, \quad (10)$$

where we have assumed a constant magnetic diffusivity, η .

In this simple model we neglect viscous and Joule heating, assume a constant turbulent thermal diffusivity, κ , and do not consider composition. Therefore the heat equation (7) becomes

$$\frac{\partial S}{\partial t} = -\mathbf{v} \cdot \nabla S + \kappa \left[\nabla^2 S + (h_\rho + h_T) \frac{\partial S}{\partial z} \right], \quad (11)$$

where

$$\left(\frac{\partial \bar{\rho}}{\partial S} \right)_p = \frac{\bar{\rho} \bar{T} h_T}{g}.$$

We solve this set of equations using Fourier expansions in the horizontal direction and a finite difference scheme on a Chebyshev grid in the vertical direction. The nonlinear terms are computed with a spectral transform method. The system is evolved in time using an explicit Adams-Bashforth method for the nonlinear terms and an implicit method for the linear terms. No hyperdiffusion is employed.

The numerical timestep is limited by a Courant condition based on the fluid and Alfvén velocities and the grid resolution. We use 2001 Chebyshev levels in the vertical direction, which places eight levels within what would be the Ekman boundary layer (if this were 3D) for $Ek = 10^{-9}$. Variables are expanded in 1333 complex Fourier modes to represent their horizontal structures, using 4001 horizontal grid

points to compute alias-free nonlinear terms. We set the aspect ratio, L/D , to 2.

4.2. The results

We begin by describing a 2D simulation that is time dependent but laminar. We choose a polytropic index of unity and a relatively small density stratification, like that of the Earth's outer fluid core, $N_\rho = 0.2$, for which the density at the bottom boundary is about 20% greater than that at the top. The Rayleigh number

$$Ra = \frac{g \Delta S D^3}{c_p v \kappa}$$

is set to 3×10^6 and the Ekman number, Ek , is set to 10^{-4} . All diffusivities are constant and equal, making Pr and q both unity. Consequently, for this case $Ro_c = 1.7 \times 10^{-1}$. We specify the background magnetic field such that $Ro_m = 10^{-2}$. The resulting simulation has a Reynolds number

$$Re = v_{max} D / v$$

of order 10^3 and so the classic Rossby number

$$Ro = v_{max} / 2\Omega D = Ek \cdot Re$$

is of order 10^{-1} . Here, v_{max} is the time-averaged maximum fluid velocity. The induced magnetic field has an average maximum intensity, B_{max} , about an order of magnitude greater than the background field intensity, B_o ; the magnetic energy is comparable to the kinetic energy. [Figure 5a](#) is a snapshot of the entropy perturbation for this Case A (see [color insert](#) following page 234). Notice how the thermal plumes extend from one boundary to the other, similar to convection cells. The evolution of the case resembles the flow seen in a lava lamp. This large-scale laminar convection is typical of the style of convection simulated with current 3D geodynamo models.

Now we examine what happens if we stop the rotation, turn off the background magnetic field and reduce the viscous and thermal diffusivities each by a factor of 577, which increases Ra to 10^{12} . The average Re increases to about 10^6 , indicating turbulence. A snapshot of this Case B is shown in Fig. 5b. As expected, this case has much more energy in the small spatial scales compared with Case A. Plumes become detached from the boundaries and interact turbulently with others in the bulk of the convection zone. Several intense, small-scale vortices are generated by the shear flows. However, as we have been suggesting in the

Entropy

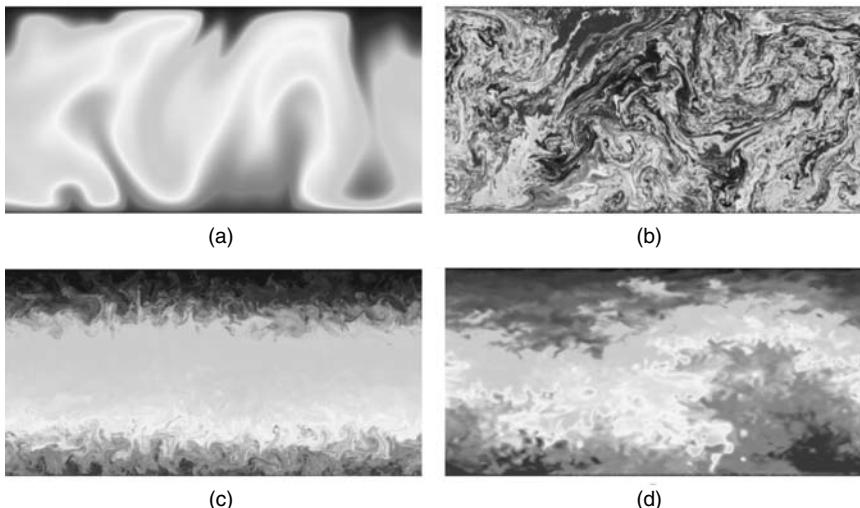


Figure 5. Snapshots of the entropy for four different 2D anelastic magnetoconvection calculations. (See color insert.) Reds are high entropy; blues are low. Gravity is downward, the bottom boundary is hot and the top is cold. (a) Rotating and magnetic, $Ra = 3 \times 10^6$, $Ek = 10^{-4}$, $Pr = 1$, $Q = 10^4$, $q = 1$, $N_\rho = 0.2$ and $n = 1$, (b) nonrotating and nonmagnetic, $Ra = 10^{12}$, $Pr = 1$, $N_\rho = 0.2$ and $n = 1$, (c) rotating but nonmagnetic, $Ra = 3 \times 10^{12}$, $Ek = 10^{-9}$, $Pr = 1$, $N_\rho = 0.2$ and $n = 1$, (d) rotating and magnetic, $Ra = 10^{11}$, $Ek = 10^{-9}$, $Pr = 1$, $Q = 10^8$, $q = 1$, $N_\rho = 2.5$ and $n = 1.5$.

geodynamics community for years, this small-scale turbulence, which is resolved in Case B, is crudely represented as enhanced diffusivities in Case A. Indeed, the large-scale structures and motions in Case B are qualitatively not unlike those of Case A. Of course part of the reason these large-scale structures are prominent here is the inverse cascade of the kinetic energy to small wave numbers in 2D flows.

Consider what happens when we restore rotation, now setting Ek to 10^{-9} and Ra to 3×10^{12} . We keep Pr at unity, so Ro_c is 1.7×10^{-3} . This simulation produces a Re of about 4×10^5 and Ro of about 4×10^{-4} , indicating very strong Coriolis effects. A typical snapshot of this turbulent rotating Case C (Figure 5c) shows a very different flow structure compared with Cases A and B. Deep turbulent boundary layers develop. The kinetic energy of buoyant plumes rising from the bottom boundary is converted into rotational kinetic energy. That is, as plumes rise from the bottom boundary the Coriolis forces resulting from the expansion generate negative vorticity, i.e., vortices with counter-clockwise rotation.

Likewise, positive vorticity (clockwise rotating vortices) is generated in plumes sinking from the top boundary. This vortex generation due to compressibility was described in the early anelastic studies (Glatzmaier and Gilman, 1981) and is represented here by the $2\Omega h_\rho v_z$ term in the vorticity equation (9). The effect is relatively insignificant in Case A and in the current 3D anelastic geodynamo simulations because viscosity in these calculations is relatively large, i.e., the Ekman number is not small enough. It is completely absent in Boussinesq simulations at any value of Ek .

When we add a background magnetic field to Case C (not shown in Fig. 5), with $Ro_m = 10^{-6}$ and $q = 1$, the induced magnetic energy is about half the kinetic energy. The Lorentz forces somewhat weaken the strong rotational constraint, resulting in more small scale plume transport through the convection zone. However, high amplitude Alfvén waves now exit in the form of local horizontal oscillations.

Accounting for the effects of compressibility becomes even more important when studying turbulent convection in the strongly density-stratified interiors of giant planets and stars. Consider Case D, which has Ek also at 10^{-9} and Pr and q still at unity. Ra is set to 10^{11} ($Ro_c = 3.3 \times 10^{-4}$) and Ro_m is set to 10^{-5} . We choose a greater density stratification, one with a density ratio of 12 ($N_\rho = 2.5$) and now assume a perfect gas ($n = 1.5$). The Re in this case becomes 10^6 ($Ro = 10^{-3}$) and the induced magnetic energy is comparable to the kinetic energy. A snapshot of this Case D is seen in Fig. 5d. Both the spatial structure and the time dependency are more complicated compared with Case C. A movie of this case shows high-frequency high-amplitude Alfvén waves manifested as local horizontal oscillations of the entropy and magnetic field structures. These oscillations have their largest amplitude and frequency in the upper, low-density region, as would be expected. Kinetic energy of this wave motion greatly exceeds that of the convection but is comparable to the magnetic energy. As a result, the numerical timestep needs to be a couple orders of magnitude smaller than what it would be without a magnetic field. This effect is not seen in current 3D anelastic simulations of stellar and giant planet dynamos or in 3D geodynamo simulations, which have less density stratification but greater magnetic energy relative to kinetic energy. It is the relatively large viscosity used in those simulations that damps Alfvén waves, i.e., again the Ekman number has been too large.

Figure 6 compares the mean entropy profile in the vertical direction for these four cases. Since we specified a constant κ for all of these cases, the mean entropy at mid-depth is greater than the average of the boundary values, which it would be for Boussinesq simulations, because of the

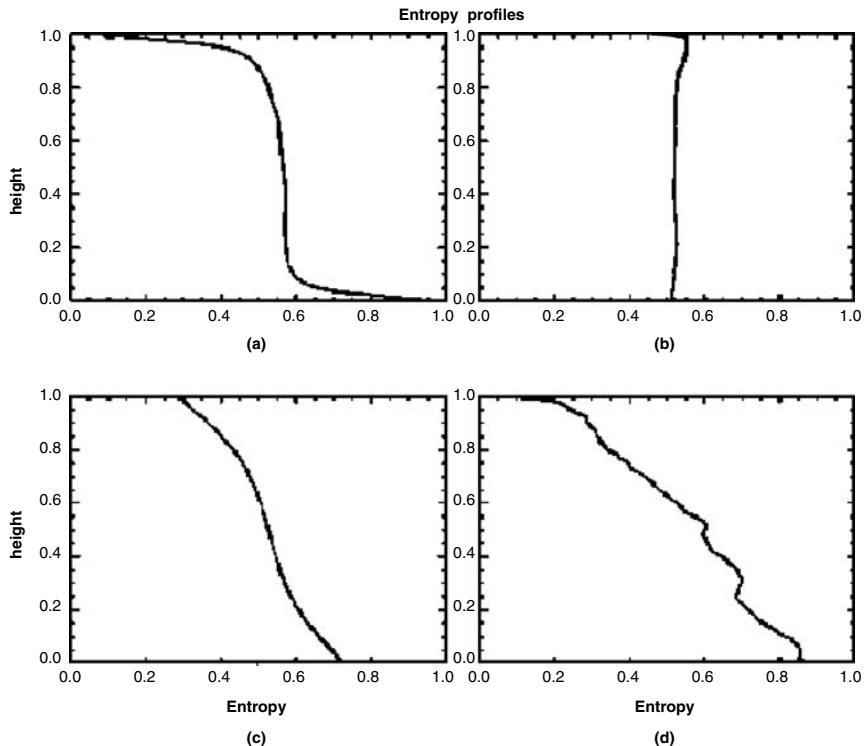


Figure 6. Time and horizontally averaged entropy (S) vs. height (z) for the four cases listed in Fig. 5. In all cases, entropy is 1 at $z = 0$ and 0 at $z = 1$.

density stratification. That is, since $\bar{\rho} \bar{T}$ at the bottom boundary exceeds that at the top boundary and the diffusive heat flux ($-\kappa \bar{\rho} \bar{T} \partial S / \partial z$) is on average the same at both boundaries, the radial gradient of the entropy is less steep at the bottom. Since the boundary layer depths are comparable, the drop in entropy across the bottom boundary is less than that across the top boundary (Evonuk and Glatzmaier, 2004).

However, what is dramatic about the plots of Fig. 6 is the structure of the mean entropy profile within the interior. The profiles for cases A and B are similar to what has traditionally been seen in non-rotating Boussinesq simulations; that is, the interior is nearly adiabatic and all the driving occurs within the shallow boundary layers. The profiles for the other cases clearly show that the combination of rotation and density stratification requires superadiabaticity in the interior to help drive both convective and diffusive heat flux. The rotational effect is

discussed by Julien *et al.* (1996) for 3D Boussinesq convection; basically, plume-plume interactions cause enhanced lateral mixing and therefore less efficient convective heat transfer in the vertical direction. The expansion of rising plumes due to a large density stratification and the lateral oscillations due to an intense magnetic field also enhance lateral mixing to the extent of producing a series of alternating stable (subadiabatic) and unstable (superadiabatic) layers (Fig. 6d). The enhanced lateral mixing is apparent when comparing the “washed out” look of Fig. 5d with the “fine structure” look of Fig. 5c.

For the 2D cases discussed so far, we have made all three diffusivities equal and constant. However, Prandtl and Roberts numbers less than unity may likely be more realistic for the fluid interiors of planets and stars. For our final case we choose a non-magnetic simulation with $Ra = 2 \times 10^{12}$, $Ek = 10^{-9}$ and $Pr = 0.1$ ($Ro_c = 4.5 \times 10^{-3}$). We also further increase the density stratification by setting $N_\rho = 5$ and $n = 1.5$, which makes the density at the bottom 148 times greater than at the top. Here the Re becomes roughly 3×10^6 ($Ro = 3 \times 10^{-3}$). The initial mean entropy, which linearly decreases by ΔS from the bottom boundary to the top, drives vigorous turbulent convection that generates many strong vortices by both shear flows and the density stratification effect. Fairly quickly, though, the mean entropy profile evolves into one with a series of stable and unstable layers, similar to that of Case D but with a deep stable layer just below the shallow top boundary layer. This significantly reduces the vigor of the convection, changes its style and causes the largest flows to exist in the upper low-density region in the form of a horizontal oscillation with a prograde drift. Several small, intense, warm core vortices, however, continue to exist and wander through the high-density region (Fig. 7; see color insert following page 234). The prograde drift is partially due to the prograde phase propagation described by Glatzmaier and Gilman (1981) as Rossby waves driven by the density stratification and may have an effect on the zonal winds in the outer, low-conductivity layers of giant planets.

5. Discussion

Our understanding of how magnetic fields are generated by convection in planets and stars has improved during the past decade with the development of magnetohydrodynamic dynamo models. We would like to believe that the relatively realistic structures and time dependencies of the simulated geomagnetic fields (outside the core), including the occasional spontaneous reversals, is not just a coincidence. However, no

Entropy

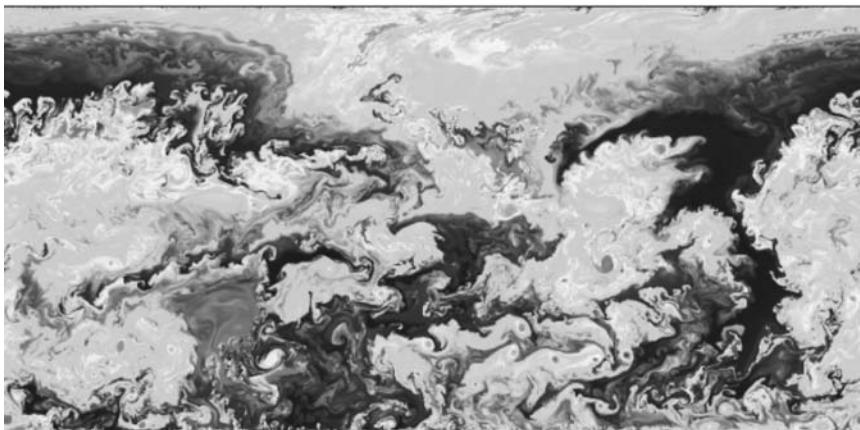


Figure 7. A snapshot of the entropy for a non-magnetic case with $Ra = 2 \times 10^{12}$, $Ek = 10^{-9}$, $Pr = 0.1$, $N_p = 5$ and $n = 1.5$. ([See color insert.](#))

dynamo model has yet simulated strongly turbulent convection, which is characteristic of stellar and planetary dynamos. We have assumed that the enhanced eddy diffusivities in our laminar simulations are adequately, albeit crudely, accounting for the mixing by the unresolved turbulence. Although asymptotic scaling relations for fluid flow and heat flow suggest that some of the current solutions are converging with increasing convective Rossby number (Christensen 2002), these simulations are still Boussinesq and laminar.

Our 2D anelastic simulations demonstrate that turbulent rotating magnetoconvection is significantly different than the corresponding laminar simulations obtained with much larger diffusivities. An important effect is the generation of negative vorticity in rising fluid as it expands and rotates due to the resulting Coriolis forces; likewise, sinking fluid generates positive vorticity. This effect is absent within the Boussinesq approximation, which most dynamo models employ. The effect is present in anelastic geodynamo simulations but damped by the large viscosity ($Ek > 10^{-5}$) used in these 3D laminar simulations. Our 2D anelastic simulations with much smaller diffusivities ($Ek = 10^{-9}$) show how significant this effect is even for the small density stratification of the Earth's fluid core. Our 2D models, however, have not included vortex stretching due to the curvature of spherical boundaries in 3D; this would have a larger effect than the density stratification for

laminar convection if, as for the Earth's core, $N_\rho < 1$. But for turbulent rotating convection, eddies and plumes in the interior of the flow have little connection with the boundaries. That is, it is unlikely that thin Taylor columns span from the southern hemisphere to the northern hemisphere without developing instabilities that destroy this 2D structure (Christensen, 2002). However, vortex generation due to flow through a density stratification is present even for 3D turbulence and is likely more significant than the boundary curvature effect, even when turbulent convection spans only a fraction of a density scale height.

When the magnetic field is strong, there is another important effect: large-amplitude high-frequency Alfvén waves that have comparable or greater kinetic energy than the convection. Again, the large diffusivities of current dynamo simulations damp these waves before they develop large amplitudes. However, in our 2D small-diffusion simulations these horizontal oscillations have significantly larger amplitude than the convective flow, which forces the numerical timestep to be extremely small for codes that treat nonlinear advection explicitly. The spatial structure and frequency spectrum of these waves in the Earth's core will be exciting to investigate; however, when simulating the long term evolution of the geodynamo, at high spatial resolution and low diffusion, a fully implicit numerical timestepping scheme may be required to effectively filter out these waves.

Unlike laminar simulations, turbulent convection in a rotating density-stratified fluid is not driven only by the shallow strongly superadiabatic thermal boundary layers; the conversion of kinetic energy of vertical flows to rotational kinetic energy of vortices requires a substantially superadiabatic mean entropy profile in the interior. When the density stratification is large and the Ekman number is small, the enhanced lateral mixing by vortices creates a staircase of alternating stable and unstable layers within the convection zone. The associated heterogeneity and anisotropy of the turbulent dynamics would likely have a significant influence on the details of the dynamo mechanism in 3D.

Although 2D dynamics certainly differ from 3D dynamics, our 2D anelastic simulations do suggest that current 3D laminar dynamo simulations are missing critical dynamical phenomena. It is therefore important to strive for much greater spatial resolution in 3D models in order to significantly reduce the eddy diffusion coefficients and simulate turbulence. This will require significant improvements in computational resources and numerical methods. In addition, subgrid scale models are needed to better represent the transport of heat, composition, momentum and possibly also magnetic field by the part of the turbulence spectrum that remains unresolved. We look forward to the

next generation of turbulent dynamo models to make new discoveries about the dynamics deep within planets and stars.

Acknowledgments

The computing resources were provided by the NSF (MCA99S012S) at the Pittsburgh Supercomputing Center and at UCSC via an NSF MRI grant (AST-0079757) and by the DOE at NERSC (DE-FC02-01ER41176). Support for this research was provided by grants from NSF (EAR-9902969, EAR-0221941), NASA (NAG5-11220) and from Los Alamos National Laboratory IGPP (1225A) and UCDRD (9948).

References

- Aurnou, J. M. and Olson, P. L., “Control of inner core rotation by electromagnetic, gravitational and mechanical torques,” *Phys. Earth Planet. Inter.* **117**, 111–121 (2000).
- Braginsky, S. I. and Roberts, P. H., “Equations governing convection in Earth’s core and the geodynamo,” *Geophys. Astrophys. Fluid Dynam.* **79**, 1–97 (1995).
- Brummell, N. H., Hurlburt, N. E. and Toomre, J., “Turbulent compressible convection with rotation. I. Flow structure and evolution,” *Astrophys. J.* **473**, 494–513 (1996).
- Brummell, N. H., Hurlburt, N. E. and Toomre, J., “Turbulent compressible convection with rotation. I. Mean flows and differential rotation,” *Astrophys. J.* **493**, 955–969 (1998).
- Buffett, B. A. and Glatzmaier, G. A., “Gravitational braking of inner-core rotation in geodynamo simulations,” *Geophys. Res. Lett.* **27**, 3125–3128 (2000).
- Busse, F. H., “Homogeneous dynamos in planetary cores and in the laboratory,” *Annu. Rev. Fluid Mech.* **32**, 383–408 (2000).
- Busse, F. H., “Convective flows in rapidly rotating spheres and their dynamo action,” *Phys. Fluids* **14**, 1301–1314 (2002).
- Busse, F. H., Grote, E. and Tilgner, A., “On convection driven dynamos in rotating spherical shells,” *Studia Geoph. Geod.* **42**, 1–6 (1998).
- Cabot, W. and Pollack, J. B., “Direct numerical simulations of turbulent convection: II. Variable gravity and differential rotation,” *Geophys. Astrophys. Fluid Dynam.* **64**, 97–133 (1992).
- Christensen, U., “Zonal flow driven by strongly supercritical convection in rotating spherical shells,” *J. Fluid Mech.* **470**, 125–133 (2002).
- Christensen, U., Olson, P. and Glatzmaier, G. A., “A dynamo model interpretation of geomagnetic field structures,” *Geophys. Res. Lett.* **25**, 1565–1568 (1998).

- Christensen, U. R., Aubert, J., Cardin, P., Dormy, E., Gibbons, S., et al., “A numerical dynamo benchmark,” *Phys. Earth Planet. Inter.* **128** 25–34 (2001).
- Creager, K., “Inner core rotation rate from small scale heterogeneity and time-varying travel times,” *Science* **278** 1284–1288 (1997).
- Dormy, E., Valet, J.-P. and Courtillot, V., “Numerical models of the geodynamo and observational constraints,” *Geochem. Geophys. Geosyst.* **1**, paper 2000GC000062. (2000).
- Evonuk, M. and Glatzmaier, G. A., “2D studies of various approximations used for modeling convection in giant planets,” *Geophys. Astrophys. Fluid Dynam.* **98**, 241–255 (2004).
- Fearn, D. R., “Hydromagnetic flow in planetary cores,” *Rep. Prog. Phys.* **61**, 175–235 (1998).
- Gilman, P. A., “Nonlinear dynamos of Boussinesq convection in a deep rotating spherical shell,” *Geophys. Astrophys. Fluid Dynam.* **8**, 93–135 (1977).
- Gilman, P. A., “Dynamically consistent nonlinear dynamos driven by convection in a rotating spherical shell. II. Dynamos with cycles and strong feedbacks,” *Astrophys. J. Suppl.* **53**, 243–268 (1983).
- Gilman, P. A. and Miller, J., “Dynamically consistent nonlinear dynamos driven by convection in a rotating spherical shell,” *Astrophys. J. Suppl.* **46**, 211–238 (1981).
- Glatzmaier, G. A., “Numerical simulations of stellar convective dynamos. I. The model and the method,” *J. Comput. Phys.* **55**, 461–484 (1984).
- Glatzmaier, G. A., “Geodynamo simulations — How realistic are they?,” *Annu. Rev. Earth Planet. Sci.* **30**, 237–257 (2002).
- Glatzmaier, G. A. and Gilman, P. A., “Compressible convection in a rotating spherical shell. III. Analytic model of compressible vorticity waves,” *Astrophys. J. Suppl.* **45**, 381–388 (1981).
- Glatzmaier, G. A. and Gilman, P. A., “Compressible convection in a rotating spherical shell. V. Induced differential rotation and meridional circulation,” *Astrophys. J.* **256**, 316–330 (1982).
- Glatzmaier, G. A. and Roberts, P. H., “A three-dimensional convective dynamo solution with rotating and finitely conducting inner core and mantle,” *Phys. Earth Planet. Inter.* **91**, 63–75 (1995a).
- Glatzmaier, G. A. and Roberts, P. H., “A three-dimensional self-consistent computer simulation of a geomagnetic field reversal,” *Nature* **377**, 203–209 (1995b).
- Glatzmaier, G. A. and Roberts, P. H., “An anelastic evolutionary geodynamo simulation driven by compositional and thermal convection,” *Physica D* **97**, 81–94 (1996a).
- Glatzmaier, G. A. and Roberts, P. H., “Rotation and magnetism of Earth’s inner core,” *Science* **274**, 1887–1891 (1996b).
- Glatzmaier, G. A. and Roberts, P. H., “Simulating the Geodynamo,” *Contemp. Phys.* **38**, 269–288 (1997).
- Glatzmaier, G. A., Coe, R. S., Hongre, L. and Roberts, P. H., “The role of the Earth’s mantle in controlling the frequency of geomagnetic reversals,” *Nature* **401**, 885–890 (1999).

- Grote, E., Busse, F. H. and Tilgner, A., "Convection-driven quadrupolar dynamos in rotating spherical shells," *Phys. Rev. E* **60**, 5025–5028 (1999).
- Grote, E., Busse, F. H. and Tilgner, A., "Regular and chaotic spherical dynamos," *Phys. Earth Planet. Inter.* **117**, 259–272 (2000).
- Gubbins, D., "The distinction between geomagnetic excursions and reversals," *Geophys. J. Int.* **137**, F1–F3 (1999).
- Hollerbach, R., "On the theory of the geodynamo," *Phys. Earth Planet. Inter.* **98**, 163–185 (1996).
- Hollerbach, R., "A spectral solution of the magneto-convection equations in spherical geometry," *Int. J. Numer. Meth. Fluids* **32**, 773–797 (2000).
- Hunkins, K., "Ekman drift currents in the Arctic Ocean," *Deep-Sea Res.* **13**, 607–620 (1966).
- Hurlburt, N. E., Toomre, J. and Massaguer, J. M., "Nonlinear compressible convection penetrating into stable layers and producing internal gravity waves," *Astrophys. J.* **311**, 563–577 (1986).
- Jones, C. A., Longbottom, A. and Hollerbach, R., "A self-consistent convection driven geodynamo model, using a mean field approximation," *Phys. Earth Planet. Inter.* **92**, 119–141 (1995).
- Julien, K., Legg, S., McWilliams, J. and Werne, J., "Rapidly rotating turbulent Rayleigh-Benard convection," *J. Fluid Mech.* **322**, 243–273 (1996).
- Kageyama, A., Sato, T., Watanabe, K., Horiuchi, R., Hayashi, T., et al., "Computer simulation of a magnetohydrodynamic dynamo. II," *Phys. Plasmas* **2**, 1421–1431 (1995).
- Katayama, J. S., Matsushima, M., and Honkura, Y., "Some characteristics of magnetic field behavior in a model of MHD dynamo thermally driven in a rotating spherical shell," *Phys. Earth Planet. Inter.* **111**, 141–159 (1999).
- Kida, S. and Kitauchi, H., "Thermally driven MHD dynamo in a rotating spherical shell," *Prog. Theor. Phys.* **130**, Suppl. 121–136 (1998).
- Kida, S., Araki, K. and Kitauchi, H., "Periodic reversals of magnetic field generated by thermal convection in a rotating spherical shell," *J. Phys. Soc. Japan* **66**, 2194–2201 (1997).
- Kono, M. and Roberts, P. H., "Recent geodynamo simulations and observations of the geomagnetic field," *Rev. Geophys.* **40**, 4.1–4.35 (2002).
- Kuang, W. and Bloxham, J., "An Earth-like numerical dynamo model," *Nature* **389**, 371–374 (1997).
- Kutzner, C. and Christensen, U. R., "From stable dipolar towards reversing numerical dynamos," *Phys. Earth Planet. Inter.* **131**, 29–45 (2002).
- Laske, G. and Masters, G., "Limits on differential rotation of the inner core from an analysis of the Earth's free oscillations," *Nature* **402**, 66–69 (1999).
- Lund, S. P., Acton, G., Clement, B., Hastedt, M., Okada, M. and Williams, T., "Geomagnetic field excursions occurred often during the last million years," *EOS Trans. AGU* **79**, pp. 178–179 (1998).
- Miesch, M. S., Elliott, J. R., Toomre, J., Clune, T. L., Glatzmaier, G. A. and Gilman, P. A., "Three-dimensional spherical simulations of solar convection. I. Differential rotation and pattern evolution achieved with laminar and turbulent states," *Astrophys. J.* **532**, 593–615 (2000).

- Pedlosky, J., *Geophysical Fluid Dynamics*, New York: Springer-Verlag (1987).
- Proctor, M. R. E., "M magnetoconvection," this volume, Chapter 8.
- Roberts, P. H. and Glatzmaier, G. A., "Geodynamo theory and simulations," *Rev. Mod. Phys.* **72**, 1081–1123 (2000).
- Rogers, T. M., Glatzmaier, G. A. and Woosley, S. E., "Simulations of two-dimensional turbulent convection in a density-stratified fluid," *Phys. Rev. E* **67**, 026315-1-6 (2003).
- Sakuraba, A. and Kono, M., "Effect of the inner core on the numerical solution of the magnetohydrodynamic dynamo," *Phys. Earth Planet. Inter.* **111**, 105–121 (1999).
- Sarson, G. R. and Jones, C. A., "A convection driven dynamo reversal model," *Phys. Earth Planet. Inter.* **111**, 3–20 (1999).
- Sarson, G. R., Jones, C. A. and Longbottom, A. W., "The influence of boundary region heterogeneities on the geodynamo," *Phys. Earth Planet. Inter.* **101**, 13–32 (1997).
- Simitev, R. and Busse, F. H., "Parameter dependences of convection driven spherical dynamos," in: *High Performance Computing in Science and Engineering '02* (Eds. E. Krause and W. Jager), pp. 15–35, Springer (2002).
- Song, X. and Richards, P., "Seismological evidence for differential rotation of the Earth's inner core," *Nature* **382**, 221–224 (1996).
- Souriau, A., "New seismological constraints on differential rotation of the inner from Novaya Zemlya events recorded at DRV Antarctica," *Geophys. J. Int.* **134**, F1–F5 (1998).
- Su, W., Dziewonski, A. M. and Jeanloz, R., "Planet within a planet: Rotation of the inner core of the Earth," *Science* **274**, 1883–1887 (1996).
- Thompson, M. J., Toomre, J., Anderson, E. R. et al., "Differential rotation and dynamics of the solar interior," *Science* **272**, 1300–1305 (1996).
- Vincent, A. P. and Yuen, D. A., "Transition to turbulent thermal convection beyond $\text{Ra} = 10^{10}$ detected in numerical simulations," *Phys. Rev. E* **61**, 5241–5246 (2000).
- Weiss, N. O., "Convection in an imposed magnetic field. Part 1. The development of nonlinear convection," *J. Fluid Mech.* **108**, 247–272 (1981a).
- Weiss, N. O., "Convection in an imposed magnetic field. Part 2. The dynamical regime," *J. Fluid Mech.* **108**, 273–289 (1981b).
- Werne, J., "Structure of hard-turbulence convection in two dimensions: Numerical evidence," *Phys. Rev. E* **48**, 1020–1035 (1993).
- Xu, X. and Song, X., "Evidence for inner core super-rotation from time-dependent differential PKP travel times observed at Beijing Seismic Network," *Geophys. J. Int.* **152**, 509–514 (2003).
- Zhang, K. and Busse, F. H., "Finite amplitude convection and magnetic field generation in a rotating spherical shell," *Geophys. Astrophys. Fluid Dynam.* **41**, 33–53 (1988).

12 Convection in rotating spherical fluid shells and its dynamo states

F.H. Busse and R. Simitev

*Physikalisches Institut, Universität Bayreuth, D-95440
Bayreuth, Germany*

The problem of convection in rotating spherical fluid shells is reviewed with emphasis on recently obtained results. The generation of magnetic fields by convection shows a strong dependence on the Prandtl number P of the fluid. Results for the computationally accessible regime of convection driven dynamos in the parameter space are given and the validity of the magnetostrophic approximation is discussed. Of particular interest are various types of dipole oscillations, reversals and torsional oscillations.

1. Introduction

The exploration of the solar system by space probes and the expanding observations of stellar magnetic fields have led to a growing interest in the problem of the generation of magnetic fields by convection flow in rotating fluid spheres. In addition to the classical examples of geomagnetism and the solar magnetic field a large variety of magnetic fields generated in celestial bodies have become known and the conditions favoring different types of dynamos are receiving increasing attention. The facts that all planetary and stellar dynamos are highly turbulent and that they usually operate in the deep interiors of these celestial bodies pose severe challenges for the understanding of their nature. Nevertheless it seems desirable to explore the large parameter space of convection driven dynamos in rotating spherical fluid shells as far as it is accessible to computational simulations.

In this paper we shall review recent results on convection in rotating spheres and its dynamo action and add some new results on

its Prandtl number dependence. While most simulations of spherical dynamos assume a Prandtl number $P = 1$, this parameter is likely to differ from unity even in highly turbulent systems where the molecular diffusivities are replaced by eddy diffusivities. High Rayleigh number convection experiments (Ahlers and Xu, 2001) and theoretical considerations (Eschrich and Rüdiger, 1983) suggest that the properties of fully turbulent convection still depend on the Prandtl number. Low Prandtl numbers are of special interest for applications because of the metallic constituents of planetary cores and because the radiative heat transport in stars contributes to a reduction of the effective Prandtl number.

In the following we first describe the mathematical formulation of the model in Section 2 and then turn to the problem of convection in the absence of a magnetic field. After a short introduction to the linear problem of the onset of convection in Section 3 we focus on the low Prandtl number regime which has received less attention in the past. While Section 4 deals with columnar convection at Prandtl number of the order $P = 0.1$, equatorially attached convection which becomes predominant at lower values of P is described in Section 5. The Prandtl number P_c separating the two types of convection decreases with increasing the Coriolis number τ , which is the dimensionless measure of the rotation rate of the system. Convection-driven dynamos with $P = 1$ are briefly reviewed in Section 6 before we turn to more recent results obtained for higher Prandtl numbers. Dynamos in this regime appear to require magnetic Prandtl numbers P_m which increase with P . The same trend continues to hold for Prandtl numbers less than unity. This regime thus seems to be best suited for attaining the goal of minimal values of P_m . In Section 7 we shall report on typical properties of low Prandtl number dynamos. Torsional oscillations are of special interest because of their connection with jerks of the geomagnetic field and will be considered in Section 8. Various types of oscillatory dipolar dynamos are discussed in Section 9. The paper closes with a discussion and an outlook on future research in Section 10.

2. Mathematical formulation of the problem

For the description of finite amplitude convection in rotating spherical shells and its dynamo action we follow the standard formulation used in earlier work by the authors (Zhang and Busse, 1989; Busse *et al.*, 1998; Grote *et al.*, 1999, 2000). But we assume that a more general static

state exists with the temperature distribution $T_S = T_0 - \beta d^2 r^2 / 2 + \Delta T \eta r^{-1} (1 - \eta)^{-2}$ where η denotes the ratio of inner to outer radius of the shell and d is its thickness. ΔT is the temperature difference between the boundaries in the special case $\beta = 0$. The gravity field is given by $\mathbf{g} = -\gamma d \mathbf{r}$ where \mathbf{r} is the position vector with respect to the center of the sphere and r is its length measured in units of d . In addition to d , the time d^2/v , the temperature $v^2/\gamma \alpha d^4$ and the magnetic flux density $v(\mu \varrho)^{1/2}/d$ are used as scales for the dimensionless description of the problem where v denotes the kinematic viscosity of the fluid, κ its thermal diffusivity, ϱ its density and μ is its magnetic permeability. We use the Boussinesq approximation in that we assume ρ to be constant except in the gravity term where its temperature dependence given by $\alpha \equiv -(d\varrho/dT)/\varrho = \text{const.}$ is taken into account. Since the velocity field \mathbf{u} as well as the magnetic flux density \mathbf{B} are solenoidal vector fields, the general representation in terms of poloidal and toroidal components can be used,

$$\mathbf{u} = \nabla \times (\nabla v \times \mathbf{r}) + \nabla w \times \mathbf{r}, \quad (1a)$$

$$\mathbf{B} = \nabla \times (\nabla h \times \mathbf{r}) + \nabla g \times \mathbf{r}. \quad (1b)$$

By multiplying the (curl)² and the curl of the Navier-Stokes equations in the rotating system by \mathbf{r} we obtain two equations for v and w

$$[(\nabla^2 - \partial_t) L_2 + \tau \partial_\phi] \nabla^2 v + \tau Q w - L_2 \Theta = -\mathbf{r} \cdot \nabla \times [\nabla \times (\mathbf{u} \cdot \nabla \mathbf{u} - \mathbf{B} \cdot \nabla \mathbf{B})], \quad (2a)$$

$$[(\nabla^2 - \partial_t) L_2 + \tau \partial_\phi] w - \tau Q v = \mathbf{r} \cdot \nabla \times (\mathbf{u} \cdot \nabla \mathbf{u} - \mathbf{B} \cdot \nabla \mathbf{B}), \quad (2b)$$

where ∂_t and ∂_ϕ denote the partial derivatives with respect to time t and with respect to the angle ϕ of a spherical system of coordinates r, θ, ϕ and where the operators L_2 and Q are defined by

$$L_2 \equiv -r^2 \nabla^2 + \partial_r(r^2 \partial_r),$$

$$Q \equiv r \cos \theta \nabla^2 - (L_2 + r \partial_r)(\cos \theta \partial_r - r^{-1} \sin \theta \partial_\theta).$$

The heat equation for the dimensionless deviation Θ from the static temperature distribution can be written in the form

$$\nabla^2 \Theta + [R_i + R_e \eta r^{-3} (1 - \eta)^{-2}] L_2 v = P(\partial_t + \mathbf{u} \cdot \nabla) \Theta \quad (2c)$$

and the equations for h and g are obtained through the multiplication of the equation of magnetic induction and of its curl by \mathbf{r} ,

$$\nabla^2 L_2 h = P_m [\partial_t L_2 h - \mathbf{r} \cdot \nabla \times (\mathbf{u} \times \mathbf{B})], \quad (2d)$$

$$\nabla^2 L_2 g = P_m \{\partial_t L_2 g - \mathbf{r} \cdot \nabla \times [\nabla \times (\mathbf{u} \times \mathbf{B})]\}. \quad (2e)$$

The Rayleigh numbers R_i and R_e , the Coriolis parameter τ , the Prandtl number P and the magnetic Prandtl number P_m are defined by

$$R_i = \frac{\alpha\gamma\beta d^6}{\nu\kappa}, \quad R_e = \frac{\alpha\gamma\Delta T d^4}{\nu\kappa}, \quad \tau = \frac{2\Omega d^2}{\nu}, \quad P = \frac{\nu}{\kappa}, \quad P_m = \frac{\nu}{\lambda}, \quad (3)$$

where λ is the magnetic diffusivity. We assume stress-free boundaries with fixed temperatures:

$$v = \partial_{rr}^2 v = \partial_r(w/r) = \Theta = 0, \quad \text{at} \quad r = r_i \equiv \eta/(1 - \eta) \\ \text{and at} \quad r = r_o = (1 - \eta)^{-1}. \quad (4a)$$

Throughout this chapter the case $\eta = 0.4$ will be considered unless indicated otherwise. For the magnetic field electrically insulating boundaries are used such that the poloidal function h must be matched to the function $h^{(e)}$ which describes the potential fields outside the fluid shell

$$g = h - h^{(e)} = \partial_r(h - h^{(e)}) = 0, \quad \text{at} \quad r = r_i \quad \text{and} \quad r = r_o. \quad (4b)$$

But computations for the case of an inner boundary with no-slip conditions and an electrical conductivity equal to that of the fluid have also been done. The numerical integration of equations (2) together with boundary conditions (4) proceeds with the pseudo-spectral method as described by Tilgner and Busse (1997) which is based on an expansion of all dependent variables in spherical harmonics for the θ , ϕ -dependences, i.e.,

$$v = \sum_{l,m} V_l^m(r, t) P_l^m(\cos \theta) \exp(im\phi) \quad (5)$$

and analogous expressions for the other variables, w , Θ , h and g . P_l^m denotes the associated Legendre functions. For the r -dependence expansions in Chebychev polynomials are used. (For further details see also Busse *et al.*, 1998, or Grote *et al.*, 1999.)

For the computations to be reported in the following a minimum of 33 collocation points in the radial direction and spherical harmonics up to the order 64 have been used. But in many cases the resolution was increased to 49 collocation points and spherical harmonics up to the order 96 or 128.

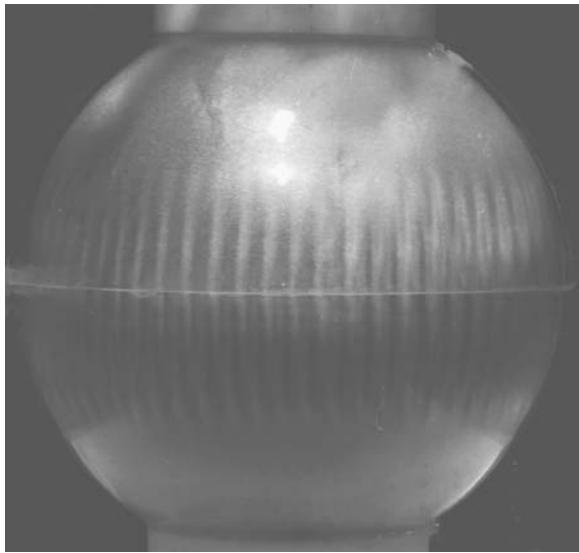


Figure 1. Banana cells in a thin rotating spherical fluid shell cooled from within. The motion is made visible by a suspension of flakes which become aligned with the shear.

3. Convection in rotating spherical shells

For an introduction to the problem of convection in spherical shells we refer to the recent review of Busse (2002a). The model of the rotating cylindrical annulus has been especially useful for the understanding of convection in rotating spheres. A rough idea of the dependence of the critical Rayleigh number R_{ic} for the onset of convection on the parameters of the problem in the case $R_e = 0$ can be gained from the expressions derived from the annulus model (Busse, 1970):

$$R_{ic} = 3 \left(\frac{P\tau}{1+P} \right)^{4/3} \frac{(\tan \theta_m)^{8/3}}{2^{2/3} r_m^{1/3}}, \quad (6a)$$

$$m_c = \left(\frac{P\tau}{1+P} \right)^{1/3} \frac{(r_m \tan \theta_m)^{3/2}}{2^{1/6}} \quad (6b)$$

$$\omega_c = \left(\frac{\tau^2}{(1+P)^2 P} \right)^{1/3} \frac{(\tan \theta_m)^{4/3}}{2^{5/6} r_m^{2/3}}, \quad (6c)$$

where r_m refers to the mean radius of the fluid shell, $r_m = (r_i + r_o)/2$, and θ_m to the corresponding colatitude, $\theta_m = \arcsin(r_m(1 - \eta))$. The

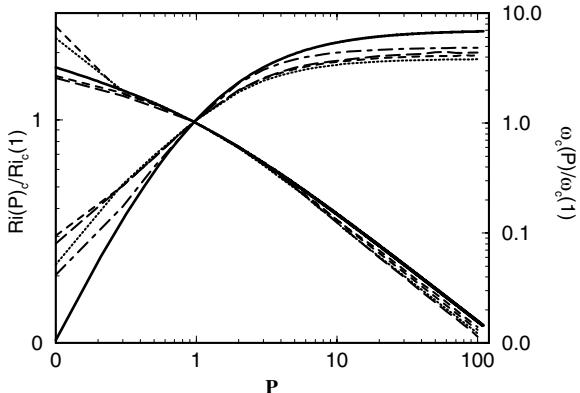


Figure 2. Critical Rayleigh number R_{ic} and frequency ω_c (right ordinate) as a function of the Prandtl number P in the case $\eta = 0.4$ for the Coriolis numbers $\tau = 5 \times 10^3$ (dotted), 10^4 (dashed), 1.5×10^4 (long-dashed) and 10^5 (dot-dashed). The thick solid line corresponds to expression (6a). The reference value $R_{ic}(1)$ equals 80296, 190140, 318395, 3665919 for $\tau = 5 \times 10^3, 10^4, 1.5 \times 10^4, 10^5$, respectively. For expression (6a) $R_{ic}(1)$ is given by $0.67543 \times \tau^{4/3}$.

azimuthal wavenumber of the preferred mode is denoted by m_c and the corresponding angular velocity of the drift of the convection columns in the prograde direction is given by ω_c/m_c . In Fig. 2 the expressions (6a,b) are compared with accurate numerical values which indicate that the general trend is well represented by expressions (6a,b). The same property holds for m_c . For a rigorous asymptotic analysis including the radial dependence we refer to Jones *et al.* (2000). Since we shall continue to restrict the attention to the case $R_e = 0$, unless indicated otherwise, we shall drop the subscript i of R_i .

There is a second mode of convection which becomes preferred at onset for sufficiently low Prandtl numbers. It is characterized by convection cells attached to the equatorial part of the outer boundary not unlike the “banana cells” seen in the narrow gap experiment of Fig. 1. The equatorially attached mode actually represents an inertial wave modified by the effects of viscous dissipation and thermal buoyancy. An analytical description of this type of convection can thus be attained through the introduction of viscous friction and buoyancy as perturbations as has been done by Zhang (1994) for the case of stress-free as well as for no-slip boundaries (Zhang, 1995). According to Ardes *et al.* (1997) equatorially attached convection is preferred at onset for $\tau < \tau_l$ where τ_l increases in proportion to $P^{-1/2}$.

4. Evolution of convection columns at small Prandtl numbers

In general the onset of convection in rotating fluid spheres occurs supercritically. As long as the convection assumes the form of shape-preserving traveling thermal Rossby waves as described by linear theory, its azimuthally averaged properties are time independent. In fact, as seen from a frame of reference drifting together with the convection columns the entire pattern is steady. A differential rotation is generated through the action of the Reynolds stress. The latter is caused by the spiraling cross section of the columns which persists as a dominant feature at moderate Prandtl numbers far into the turbulent regime. The plots of the streamlines $r\partial v/\partial\phi = \text{const.}$ in the equatorial plane shown in Fig. 3 give a good impression of the spiraling nature of the columns.

A true time dependence of convection develops in the form of vacillations after a subsequent bifurcation. First the transition to amplitude vacillations occurs in which case just the amplitude of convection varies periodically in time as exhibited in the left plot of Fig. 3. At a somewhat higher Rayleigh number shape vacillations become noticeable which are characterized by periodic changes in the structure of the columns as shown in the right plot of Fig. 3. The outer part of the columns is stretched out, breaks off and decays. The tendency toward breakup is caused by the fact that the local frequency of propagation varies with distance from the axis according to expression (6c) after θ_m has been replaced by the local colatitude θ .

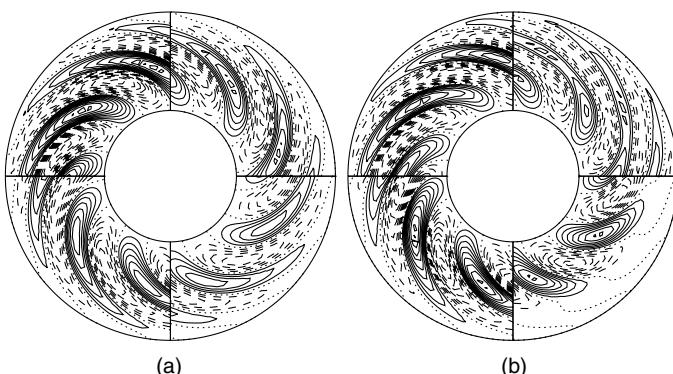


Figure 3. Lines of constant $r\partial v/\partial\phi$ in the equatorial plane in the case $\tau = 3 \times 10^4$, $P = 0.1$ for (a) $R = 3 \times 10^5$ and (b) $R = 3.5 \times 10^5$. The four sections are a quarter of the vacillation period, $t_p = 0.0864$ (left) and $t_p = 0.0124$ (right), apart with time progressing in the clockwise sense.

The two types of vacillations also differ significantly in their frequencies of oscillation. This is evident from the time records of the energy densities of convection which have been plotted in Fig. 4, namely

$$E_p^m = \frac{1}{2} \langle |\nabla \times (\nabla \bar{v} \times \mathbf{r})|^2 \rangle, \quad E_t^m = \frac{1}{2} \langle |\nabla \bar{w} \times \mathbf{r}|^2 \rangle, \quad (7a)$$

$$E_p^f = \frac{1}{2} \langle |\nabla \times (\nabla \check{v} \times \mathbf{r})|^2 \rangle, \quad E_t^f = \frac{1}{2} \langle |\nabla \check{w} \times \mathbf{r}|^2 \rangle, \quad (7b)$$

where \bar{v} refers to the azimuthally averaged component of v and \check{v} is defined by $\check{v} = v - \bar{v}$. As the Rayleigh number is increased further a fairly sudden transition into a chaotic regime occurs where convection has become strongly inhomogeneous in space and in time. A typical sequence of plots is shown in Fig. 5 which covers about one period of the relaxation cycles seen in the time record for $R = 3.8 \times 10^5$ in Fig. 4. In contrast to the more common relaxation oscillations encountered at higher Rayleigh numbers—see for example the time record for $R = 5 \times 10^5$ in Fig. 4—convection does not die off entirely at any time during the cycle. But the interaction between convection and differential rotation appears to be similar. As the amplitude of convection as measured by E_p^f and E_t^f grows, the differential rotation generated by the Reynolds stress grows as well with just a small delay in time. When the differential rotation reaches a critical level the convection columns become disrupted and their amplitude decays. Subsequently the differential rotation decays as well on the timescale of viscous diffusion. It is typical for this type of relaxation cycles that the viscous decay is shorter than the growth time of the differential rotation in contrast to the relaxation oscillations at higher values of R . The regime of relaxation oscillations is interrupted once in a while by a return to the more regular regime of shape vacillations as shown in the inserted enlargement of the time record for $R = 3.8 \times 10^5$ in Fig. 4. But in contrast to the vacillation of the right plot of Fig. 3 the pattern is now strongly modulated. The component with the azimuthal wavenumber $m = 1$ plays a dominant role in this modulation.

As R is further increased the spatio-temporal structure of convection becomes more irregular as can be seen in the time series for $R = 4.5 \times 10^5$ in Fig. 4. But at $R = 5 \times 10^5$ the relaxation oscillations with only intermittent convection become firmly established and they continue to persist up to Rayleigh numbers of the order of 10^6 . They are basically the same phenomenon as found by Grote and Busse (2001) at $P = 1$ and even the period is the same, about 0.1, which corresponds to the viscous decay time of the differential rotation.

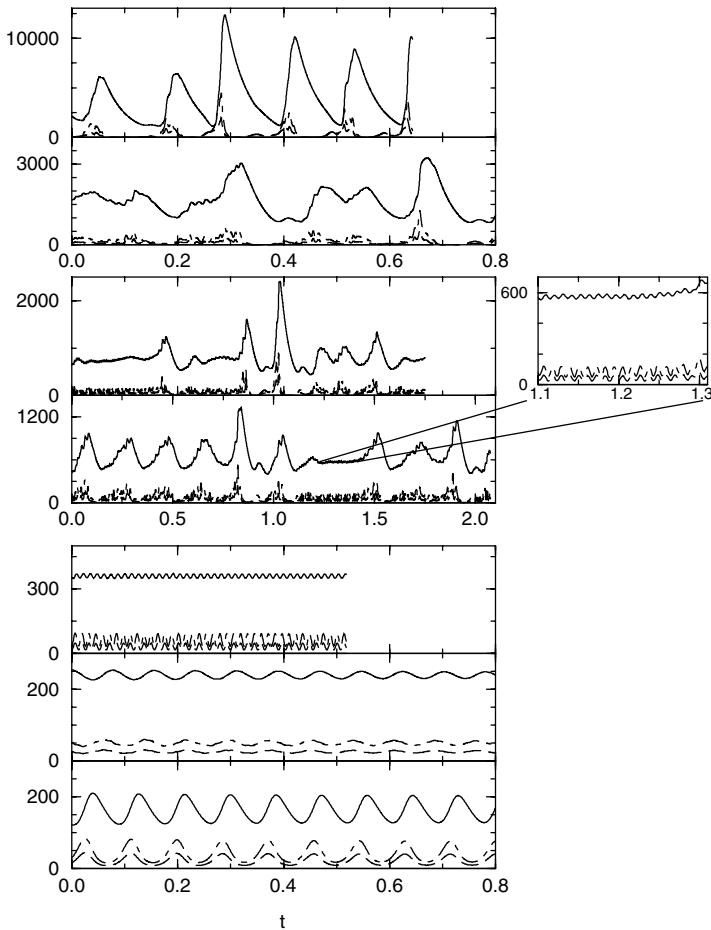


Figure 4. Time series of energy densities of convection in the case $\tau = 3 \cdot 10^4$, $P = 0.1$ for $R = 3 \times 10^5, 3.3 \times 10^5, 3.5 \times 10^5, 3.8 \times 10^5, 4 \times 10^5, 4.5 \times 10^5, 5 \times 10^5$ (from bottom to top). Solid, dot-dashed and dashed lines indicate E_t^m , E_t^f , and E_p^f , respectively. The critical Rayleigh number for onset of convection is $R_c = 222518$. E_p^m is smaller by more than an order of magnitude than the other energy densities and has not been plotted for this reason.

The main difference between convection at $P = 0.1$ and Prandtl numbers of the order unity occurs at the transition between regular and irregular patterns. A typical scenario for $P = 0.5$ is shown in Fig. 6. After convection has set in in the form of eight drifting column pairs, the usual amplitude vacillations occur as the Rayleigh number

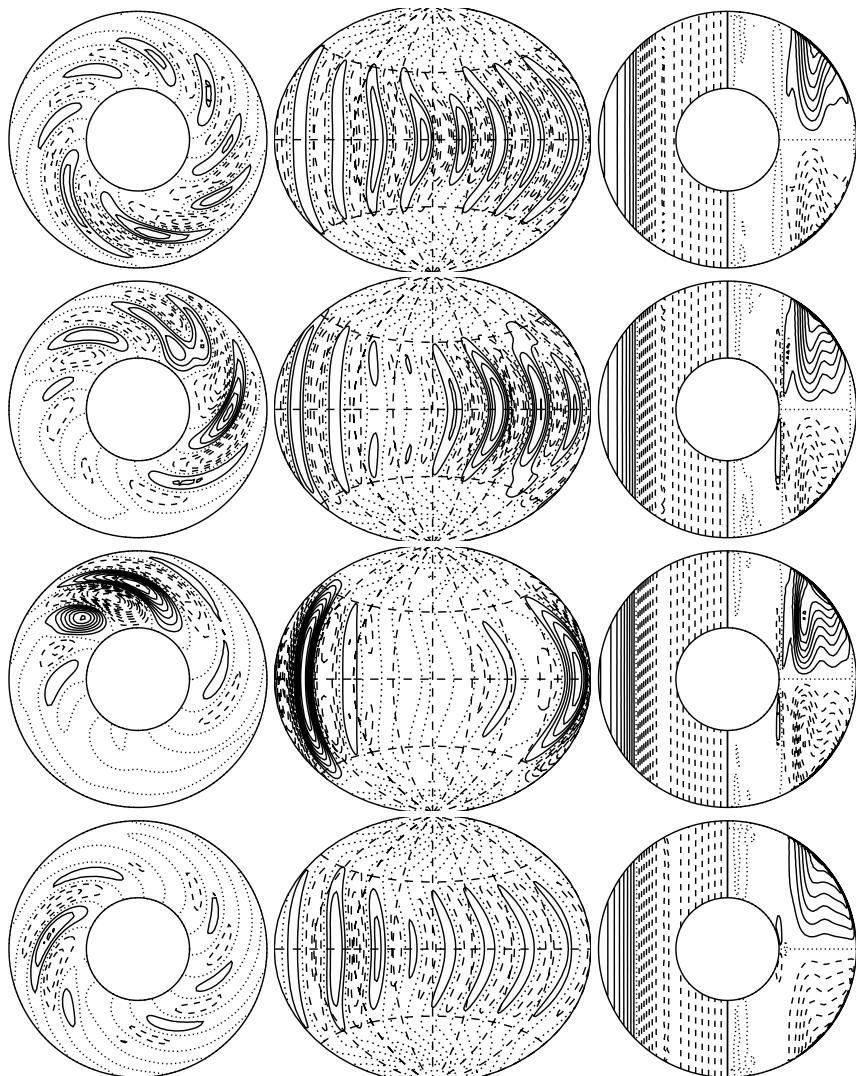


Figure 5. Sequence of plots equidistant in time (from top to bottom with $\Delta t = 0.05$) for $R = 3.8 \times 10^5$, $\tau = 3 \times 10^4$, $P = 0.1$. The left column shows streamlines, $r \partial v / \partial \phi = \text{const.}$, in the equatorial plane and in the middle column lines of constant radial velocity u_r on the mid surface, $r = r_i + 0.5$, are shown. The left halves of the circles of the right column show lines of constant u_ϕ which is the azimuthally averaged azimuthal component of the velocity field. The right halves show streamlines of the axisymmetric meridional circulation.

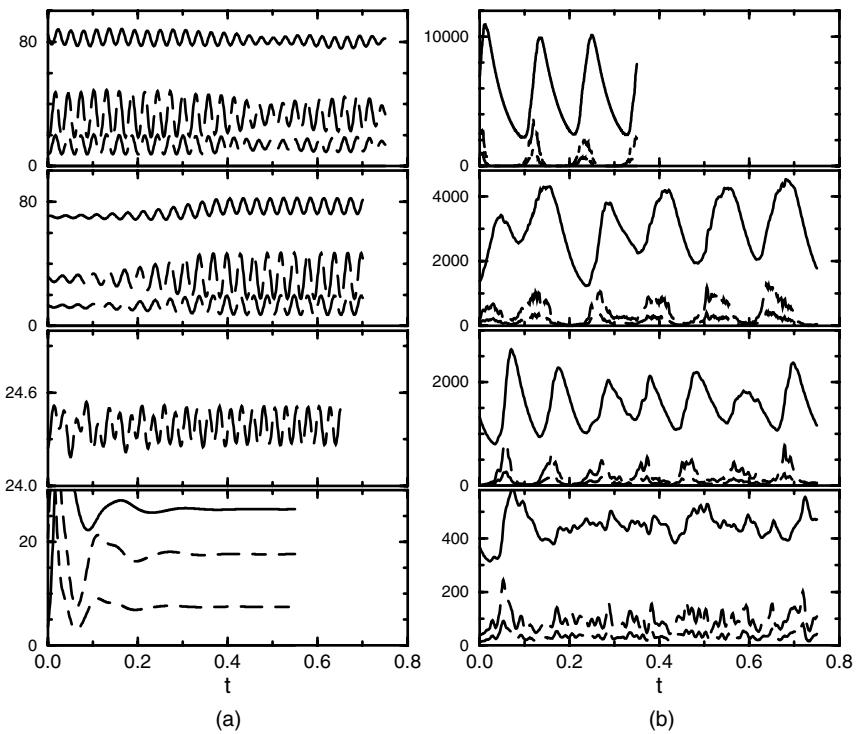


Figure 6. Time series of energy densities of convection in the case $\tau = 1.5 \times 10^4$, $P = 0.5$ for (a) $R = 3 \times 10^5, 3.2 \times 10^5, 3.45 \times 10^5, 3.5 \times 10^5$ (from bottom to top) and (b) $R = 5 \times 10^5, 7 \times 10^5, 8 \times 10^5, 10^6$ (from bottom to top). Solid, dot-dashed and dashed lines indicate E_t^m , E_t^f and E_p^f , respectively. The critical Rayleigh number for onset is $R_c = 215142$.

is increased. The shape vacillations, however, exhibit a modulation with wavenumber $m = 4$ as shown in Fig. 7. Only every second pair of columns gets stretched until the outer part separates, a little earlier for the cyclonic column than for the anticyclonic one. Then the same process is repeated for the other pairs of columns such that the sequence shown in Fig. 7 exhibits only half a period of the oscillation. In addition, of course, the column pattern drifts in the prograde direction. With increasing Rayleigh number the stretching process gets out of phase and an $m = 1$ -modulation becomes noticeable. The time dependence is still periodic as indicated in the record for $R = 3.45 \times 10^5$ in Fig. 6. But the oscillations have become more complex in that some of the separated outer parts become attached to the preceding column pair. By the time when R has reached 3.5×10^5 the modulated vacillations

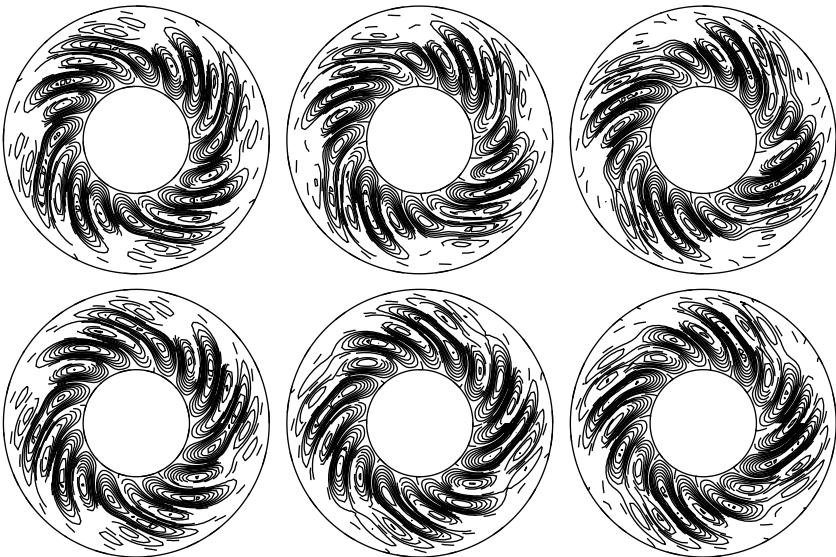


Figure 7. Modulated shape vacillations of convection for $R = 3.2 \times 10^5$, $\tau = 1.5 \times 10^4$, $P = 0.5$. The sequence of plots equidistant in time ($\Delta t = 0.005$), starting at the upper left and continuing clockwise, shows streamlines, $r\partial v/\partial\phi = \text{const.}$, in the equatorial plane. Since the modulation period is about 0.025, the last plot resembles the first plot except for a shift in azimuth.

have become aperiodic and with increasing Rayleigh number convection becomes more and more chaotic. Regularity reappears only in the form of relaxation oscillations as shown in the section for $R = 7 \times 10^5$ of Fig. 6. Remainders of the vacillations can still be seen in this time record. Around $R = 8 \times 10^5$ the relaxation oscillations occur in combination with the phenomenon of localized convection (Grote and Busse, 2001) in that the convection is confined to a meridional sector with an extent of about 90° in longitude as shown in Fig. 8. At $R = 10^6$, however, the relaxation oscillations have become fully established with a sharp rise and a slower decay of the differential rotation and with convection occurring only in an intermittent fashion, but nearly homogeneously in azimuth.

5. Convection at very small Prandtl numbers

Since stellar interiors as well as metallic planetary cores are characterized by rather small Prandtl numbers the onset of convection in rotating fluid spheres in the limit of vanishing P has received special attention.

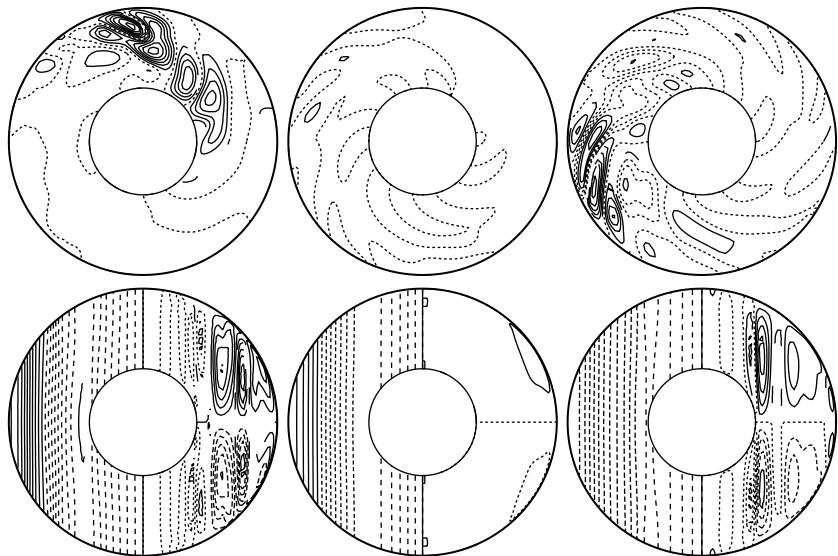


Figure 8. Streamlines, $r\partial v/\partial\phi = \text{const.}$, in the equatorial plane (upper row) and lines of constant u_ϕ (left halves of circles in lower row) and streamlines of the axisymmetric meridional circulation (right halves) are shown for the case $P = 0.5$, $R = 8 \times 10^5$, $\tau = 1.5 \times 10^4$ at the times $t = 0.1 + n \cdot 0.05$, $n = 1, 2, 3$ (left to right) of the corresponding time series in Fig. 6.

Zhang and Busse (1987) found the equatorially attached mode which is quite distinct from the columnar mode discussed in the preceding section. The new mode represents an inertial oscillation which becomes excited when viscous dissipation is exceeded by the energy provided per unit time by thermal buoyancy. The fact that both energies can be regarded as small perturbations has led Zhang (1994, 1995) to solve the problem of onset of convection by an asymptotic analysis. A detailed numerical study together with some analytical approximations can be found in the paper of Ardes *et al.* (1997). The results of these various efforts have turned out to be rather complex since the preferred inertial modes travel in the prograde as well in the retrograde directions depending on the parameters of the problem. Moreover, the azimuthal wavenumber m of convection does not increase monotonically with the Coriolis parameter τ as is usually found for the columnar mode at values of P of the order unity or higher. In addition to the simple “single cell” inertial modes of Zhang and Busse (1987) multicellular modes have been found in the study of Ardes *et al.* (1997) which appear to be closely related to the multicellular modes described by the Airy function in the

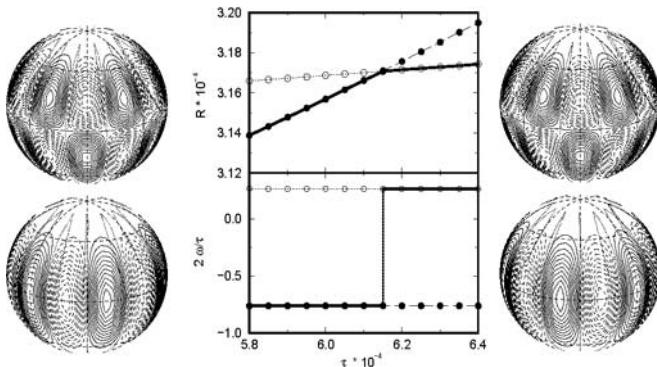


Figure 9. In the middle part: the Rayleigh numbers R of the competing prograde (dotted line, empty circles) and retrograde mode (dashed line, filled circles) as well as the actual critical value (thick solid line) and the corresponding frequencies as a function of τ in the case $P = 0.001$, $\eta = 0.2$, $m = 6$. On the left and right: Contours of constant radial velocity u_r (lower plots) and the streamlines, $w = \text{const.}$ on the spherical surface $r = 0.9$ (upper plots) for $\tau = 58000$ and 64000 .

analysis by Yano (1992) of the analogous problem of convection in the rotating cylindrical annulus. The numerical solutions for this problem obtained by Pino *et al.* (2000) also indicate the onset of multicellular convection in parts of the parameter space.

Here we shall present only some examples of the variety of patterns encountered in a rotating sphere of a low Prandtl number fluid. More details can be found in the recent paper of Simitev and Busse (2003). In Fig. 9 a typical change from the retrograde to the prograde modes as a function of τ has been indicated and the corresponding changes in the patterns are shown. The retrograde and prograde modes differ little in their form. The opposite phase between toroidal and poloidal components of motion is the most characteristic difference as indicated in Fig. 9. The sense of outward spiraling is also opposite for prograde and retrograde modes as is evident from Fig. 10. But in the limit of vanishing Prandtl number or infinite τ the sense of spiraling disappears because the phase of the inertial modes does not vary with distance from the axis.

Whenever τ is sufficiently large the frequency ω is closely approximated by the frequency of the corresponding inertial modes which is given by the analytical expression (Zhang, 1994; Ardes *et al.*, 1997),

$$\omega = \frac{\tau}{m+2} \left\{ 1 \pm \left[1 + \frac{m(m+2)}{2m+3} \right]^{1/2} \right\}. \quad (8)$$



Figure 10. Lines of constant $r \partial v / \partial \phi$ in the equatorial plane corresponding to the values of $R = 15293.7, 17190.8, 31963.2, 763401$, for $\tau = 950, 1500, 8 \times 10^5, 3.5 \times 10^7$, respectively (left to right, first upper row then lower row), in the case $P = 10^{-4}$.

The negative (positive) sign applies for modes drifting in the prograde (retrograde) direction.

6. Convection driven dynamos at moderate Prandtl numbers

Numerous reviews of convection driven dynamos in rotation spherical shells have appeared in recent years (Busse, 2000; Busse *et al.*, 2003; Dormy *et al.*, 2000; Kono and Roberts, 2002 and others) which usually have focused on the case of $P = 1$, although some simulations have been carried out for Prandtl numbers as high as 900 (Glatzmaier and Roberts, 1995). Here we do not wish to provide still another general survey, but instead focus on the Prandtl number dependence and draw attention to a few questions that have not been answered satisfactorily. Most prominent among these questions is the dependence of the average magnetic density energy M on the parameters of the problem. As a convenient definition for M we shall use the sum $M = M_p^m + M_t^m + M_p^f + M_t^f$ where M_p^m etc. are defined in direct analogy to the definitions (7)

for the kinetic energy densities. The most popular rules of thumb are the criterion that the Elsasser number

$$\Lambda = 2MP_m/\tau \quad (9)$$

assumes a value of the order unity and the equipartition criterion that magnetic and kinetic energies are in approximate balance. This latter criterion is satisfied, for example, when the dynamics of a turbulent plasma are dominated by Alfvén waves. But this situation is not realized in the case of planetary dynamos or in the case of the Sun. The condition $\Lambda \approx 1$ determines minimum Rayleigh number for the onset of convection in a horizontal fluid layer heated from below and rotating about a vertical axis when a homogeneous magnetic field is applied parallel to that axis (Chandrasekhar, 1961). Similar conditions apply in the case of a rotating sphere when an azimuthal magnetic field is applied (Fearn, 1979) or in the corresponding annulus model (Busse, 1983) and in some other configurations involving rotation and a homogeneous magnetic field (Eltayeb, 1972). There is a little evidence that the dynamo generated magnetic field plays a dynamical role similar to that of a nearly homogeneous imposed field. It has not been possible so far to find convection at Rayleigh numbers below the critical value in the presence of dynamo action as must be expected if the main function of the Lorentz force is to counteract the Coriolis force. Instead it has been found – at least for P of the order unity or less – that the generated magnetic field inhibits the differential rotation and thereby increases the amplitude of convection and its heat transport (Grote and Busse, 2001; Busse *et al.*, 2003).

Kinetic and magnetic energies have been plotted as a function of R in Fig. 11 in the case $P = P_m = 1$, $\tau = 5 \times 10^3$. The Elsasser number Λ approaches unity for the maximum value of the magnetic energy which is achieved at Rayleigh numbers in the neighborhood of 1.4×10^5 . Flux expulsion and an increasingly filamentary structure of the magnetic field cause a decrease of the energy for higher values of R , while the Ohmic dissipation continues to increase. For the same reason axisymmetric components of the fields decay faster with increasing R than the fluctuating nonaxisymmetric components. Since the critical value R_c is only about 8.03×10^4 , convection in the polar region has set in about $R \approx 5 \times 10^5$ and dipolar and quadrupolar components of the magnetic field contribute about equal amounts to the energy for $R \geq 10^6$.

That convection amplitudes in an intermediate range are most suitable for the generation of a global magnetic field appears to be a general

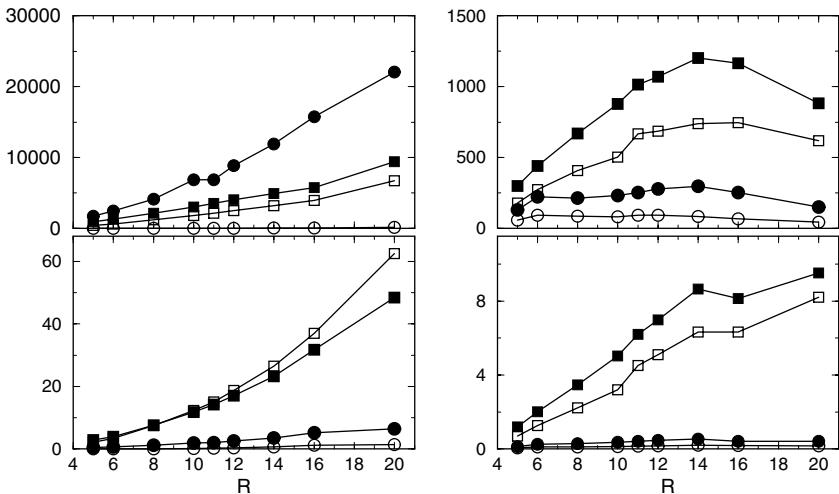


Figure 11. Kinetic (upper left) and magnetic (upper right) energy densities and viscous (lower left) and Ohmic (lower right) dissipation are plotted as a function of R_i for convection driven dynamos in the case $\tau = 5 \times 10^3$, $P = P_m = 1$. Filled (open) symbols indicate toroidal (poloidal) components of the energies and dissipations, circles (squares) indicate axisymmetric (non-axisymmetric) components. The values of R_i at the abscissa should be multiplied by 10^5 . The scales of the ordinates in the two lower plots must also be multiplied by the factor 10^5 .

property of spherical dynamos. In Fig. 12 (see color insert following page 234) convection driven dynamos are indicated in the $R - P - P_m$ -space. In the particular case of $P = 10$, $P_m = 10$, it was found that the magnetic field decays for $R = 3 \cdot 10^5$ and $R = 7 \cdot 10^5$ while sustained dynamo action appears to exist for $R = 5 \cdot 10^5$ and $R = 6 \cdot 10^5$ as shown in Fig. 13. The Elsasser number for these latter cases is 0.12.

The structure of dynamos corresponds to the expectation that quadrupolar dynamos are found for lower values of P_m while dipolar dynamos predominate at higher values of P_m and also at higher Rayleigh numbers. Here, however, a clear distinction can often no longer be made in that components of both symmetries contribute about equally to the magnetic energy. There exists a critical value P_{mc} of the magnetic Prandtl number for dynamo action which appears to increase monotonically with P . At Prandtl numbers of the order unity P_{mc} increases more strongly with P because of the concurrent decrease of the differential rotation. The ω -effect provided by the latter certainly enhances the dynamo action. As can be seen from the logarithmic plot of Fig. 14

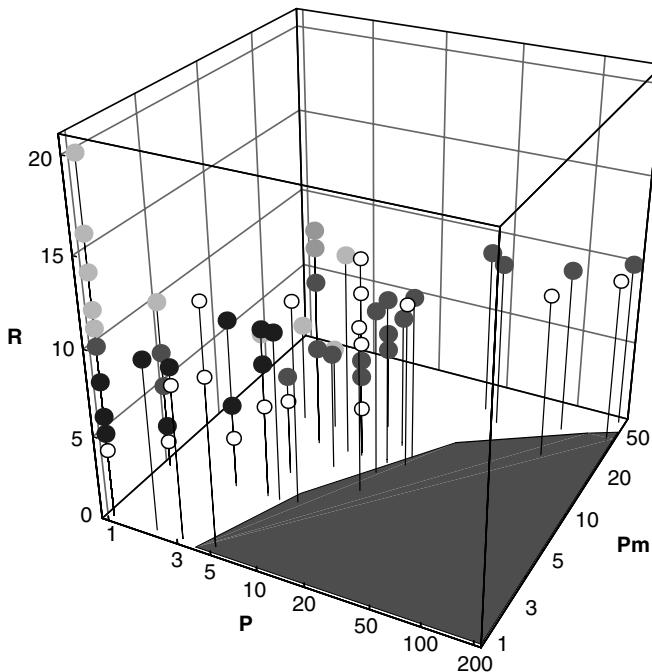


Figure 12. Dynamo solutions indicated by red (dipolar), blue (quadrupolar), green (hemispherical) and yellow (mixed symmetry) balls in the $R - P - P_m$ parameter space. (See color insert.) No dynamo solution could be obtained for values of P, P_m in the shaded region. The values of the ordinate should be multiplied by the factor 10^5 .

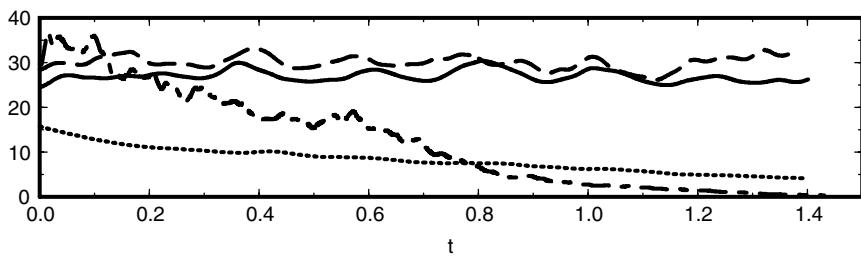


Figure 13. Total magnetic energies in the cases $P = P_m = 10$, $\tau = 5 \times 10^5$ and $R = 3 \times 10^5$ (dotted line), $R = 5 \times 10^5$ (solid line), $R = 6 \times 10^5$ (dashed line), $R = 7 \times 10^5$ (dot-dashed line). The abscissa of the case $R = 3 \times 10^5$ is multiplied by the factor 0.65 and in the case $R = 7 \times 10^5$ by 0.286.

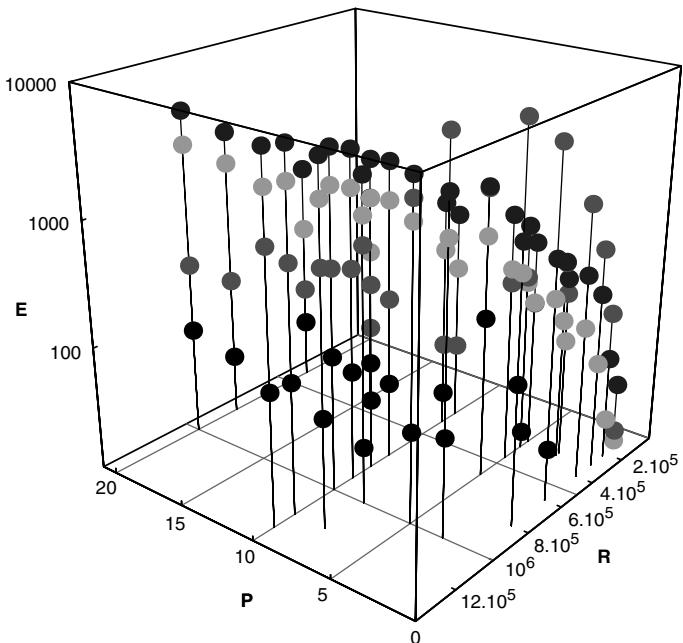
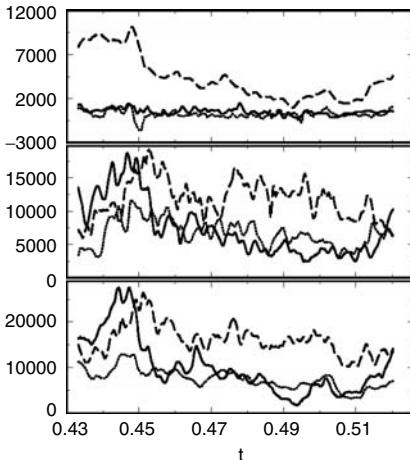


Figure 14. Dependences of energy densities of the axisymmetric toroidal (red), axisymmetric poloidal (black, multiplied by 10^2), non-axisymmetric toroidal (blue) and non-axisymmetric poloidal (green) components of motion on R and P in the case of $\tau = 5 \times 10^3$. (See color insert.) The energy densities have been multiplied by P^2 and thus are measured in terms of the thermal scaling.

(see color insert following page 234) the energy E_t^m quickly decreases in comparison to the energies E_t^f and E_p^f when P is increased.

It is of interest to compare the major sources of the energies of the various components of the magnetic field. To do this the equations (2d,e) have been multiplied by $\nabla^2 h$, $\nabla^2 \check{h}$ and by $-\bar{g}$ and $-\check{g}$, respectively, and averaged over the spherical shell. Some typical time series are shown in Fig. 15 for the case $P = 1$. Only the major contributors have been included. As expected, the shearing of the mean poloidal field by the differential rotation is the main generator of the azimuthal field. The mean poloidal field is generated primarily by the interaction between fluctuating velocity and magnetic field components. The generation of the latter component also occurs mainly through the interaction of fluctuating components. An α -effect in the traditional sense does not operate. This pattern of the dynamo process does not change much as the Prandtl number is increased. Even though the differential rotation decreases, it is still the major energy source for the mean azimuthal field.

Figure 15. The terms $(\check{v}\check{g}\check{h})$ (solid line), $(\check{w}\check{h}\check{h})$ (dotted line) and $(\bar{w}\bar{h}\bar{g})$ (dashed line) in the upper plot and the terms $(\check{v}\check{g}\check{h})$ (solid), $(\check{v}\check{g}\check{h})$ (dashed), and $(\check{v}\check{h}\check{h})$ (dotted) in the middle plot and the terms $(\check{w}\check{g}\check{g})$ (solid), $(\check{w}\check{g}\check{g})$ (dashed) and $(\check{w}\check{h}\check{g})$ (dotted) in the lower plot are displayed as a function of time in the case $\tau = 5 \times 10^3$, $R = 14 \times 10^5$, $P = P_m = 1$. The notation $(\check{v}\check{g}\check{h})$ indicates the contribution made by terms involving \check{v} and \check{g} when the equation (2d) is multiplied by $\nabla^2 \check{h}$ and averaged over the fluid shell.



But the latter weakens significantly in comparison to the mean poloidal field with increasing P as is also evident from Fig. 16.

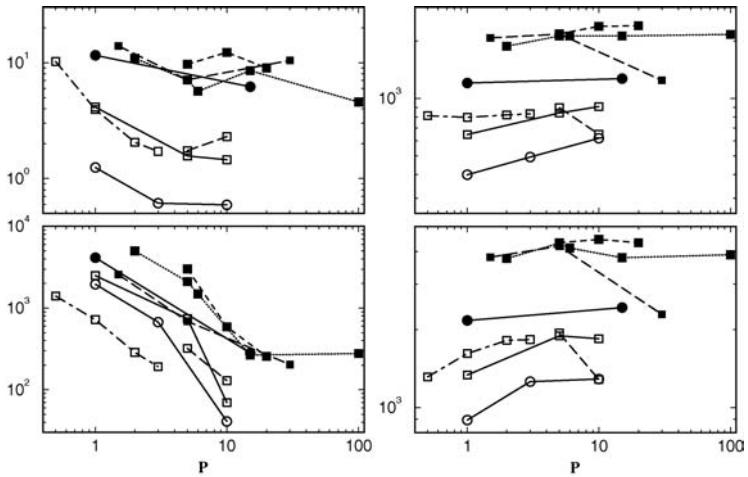
The fact that convection driven dynamos in rotating systems depend on a rather large number of parameters has led to numerous attempts to eliminate one or more parameters through reductions of the basic equations. The most important among these reductions is the magnetostrophic approximation in which the acceleration of fluid particles is neglected in comparison to the Coriolis force and the Lorentz force. This approximation is most easily obtained when the thermal timescale d^2/κ is used instead of the viscous timescale and $\sqrt{\rho\mu\kappa\nu}/d$ is used as scale of the magnetic field. The basic dimensionless equations of motion, equation of induction and the heat equation can then be written in the form

$$P^{-1}(\partial_t \mathbf{u} + \mathbf{u} \cdot \nabla \mathbf{u}) + \tau \mathbf{k} \times \mathbf{u} = -\nabla \pi + \Theta \mathbf{r} + \nabla^2 \mathbf{u} + (\nabla \times \mathbf{B}) \times \mathbf{B}, \quad (10a)$$

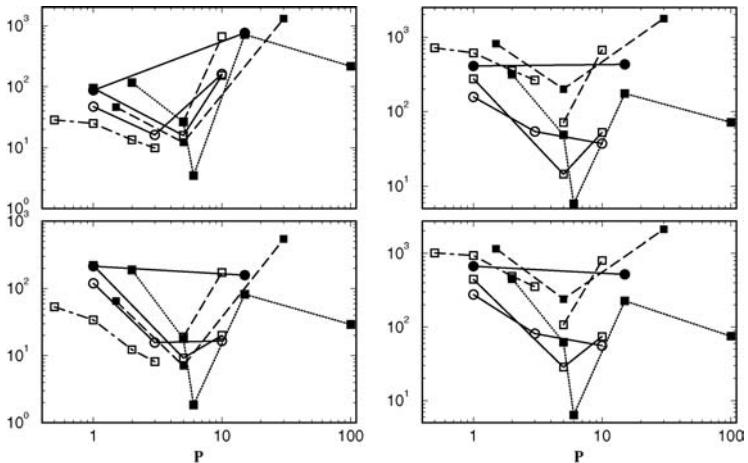
$$\frac{\kappa}{\lambda}(\partial_t \mathbf{B} + \mathbf{u} \cdot \nabla \mathbf{B} - \mathbf{B} \cdot \nabla \mathbf{u}) = \nabla^2 \mathbf{B}, \quad (10b)$$

$$\partial_t \Theta + \mathbf{u} \cdot \nabla \Theta = R \mathbf{u} \cdot \mathbf{r} + \nabla^2 \Theta, \quad (10c)$$

where \mathbf{k} is the unit vector parallel to the axis of rotation. From the form of (10) it is clear that the magnetostrophic approximation should certainly be valid in the limit $P \rightarrow \infty$. For values of P of the order unity or less the question of the validity of the magnetostrophic approximation is less obvious since the generation of differential rotation by the Reynolds stresses of convection is not represented in that approximation. Also, inertial oscillations which play a role in low Prandtl



(a)



(b)

Figure 16. (a) Kinetic energy densities E_p^m (upper left panel), E_p^f (upper right panel), E_t^m (lower left panel) and E_t^f (lower right panel) all multiplied by P^2 as a function of the Prandtl number P in the case $\tau = 5 \times 10^3$. The dynamos corresponding to fixed ratios $\kappa/\lambda = 1, 2, 5, 0.5$ are indicated by solid, long-dashed, dash-dotted and dotted lines, respectively, and cases of pure convection by a short-dashed line. The values of the Rayleigh number $R = 5 \times 10^5, 6 \times 10^5, 8 \times 10^5, 10^6$ are denoted by empty circles and squares and full circles and squares, respectively. (b) Same as Fig. 16a but for the magnetic energy densities M_p^m (upper left panel), M_p^f (upper right panel), M_t^m (lower left panel) and M_t^f (lower right panel) all multiplied by P .

number convection cannot be represented in the magnetostrophic approximation. In Figs. 16a and b the energy densities have been plotted for fixed values of κ/λ . It can be seen that the kinetic energies tend to become independent of P with increasing P in accordance with the magnetostrophic assumption. The energy E_t^m representing the differential rotation is the only exception as expected. No indication of an approach toward the validity of the magnetostrophic approximation is found, however, when the magnetic energy densities are considered. As has already been mentioned the dynamo process is rather sensitive to the presence of the differential rotation and much higher values of P may be needed before the magnetostrophic regime is approached.

Results obtained on the basis of the magnetostrophic approximation are by definition independent of P . In particular the ratio between the magnetic energy and kinetic energy will be proportional to P (Glatzmaier and Roberts, 1995) as is evident from the different scales used for the velocity and for $\mathbf{B}/\sqrt{\rho\mu}$ in (10a). The ratio between Ohmic and viscous dissipation would be independent of P and would depend only on κ/λ . This latter parameter seems to be even more important than the magnetic Prandtl number for convection driven dynamos. At least for Prandtl number of unity or less $\lambda \lesssim \kappa$ appears to be a condition for the dynamo action as is indicated by a comparison of dynamos obtained for $P = 1$ (see Fig. 1 of Grote *et al.*, 2000) and for $P = 0.1$ (see Fig. 17).

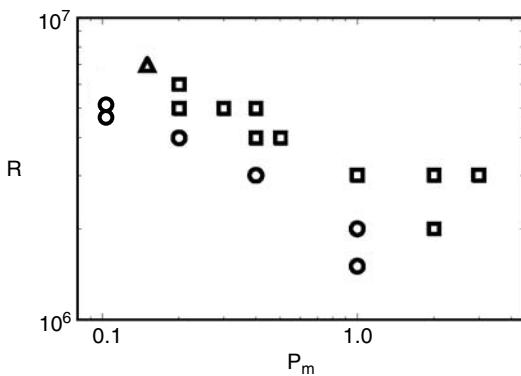


Figure 17. Convection driven dynamos of different types as a function of the Rayleigh number R and the magnetic Prandtl number P_m in the case $\tau = 10^5$ and $P = 0.1$. Dipolar and hemispherical dynamos are indicated by squares and triangles, respectively, while the circles correspond to decaying dynamos. The critical Rayleigh number for onset of convection at these parameters is $R_c = 1.035 \times 10^6$.

Before leaving the magnetostrophic approximation we wish to mention that it is also obtained when the limit $P_m \rightarrow \infty$ is approached. This property is easily seen when d^2/λ is used as timescale and the magnetic flux density \mathbf{B} is scaled with $\sqrt{\rho\mu\lambda v}/d$. Instead with P^{-1} the acceleration term in the equation of motion (10a) will then be multiplied by P_m^{-1} . Dynamo solution should become independent for $P_m \rightarrow \infty$ and the ratio of magnetic to kinetic energy would grow in proportion to P_m .

7. Convection driven dynamos in low Prandtl number fluids

A systematic exploration of the onset of dynamos in low Prandtl number spherical shells has been done in the case $P = 0.1$. The results are shown in Fig. 17 and indicate the usual behavior that the Rayleigh number must be increased for decreasing P_m in order to reach a critical magnetic Reynolds number of the order 100 for dynamo action. The property that for larger values P_m dipolar dynamos are realized while for lower values a transition to hemispherical dynamos occurs also agrees with the results obtained earlier for $P = 1$ (Grote *et al.*, 2001). When dynamo at even lower values of P_m than shown in Fig. 17 would be obtained a further transition to quadrupolar dynamos can be expected. Problems of numerical resolution have prevented us so far from reaching Rayleigh numbers beyond 10^7 . It is remarkable, however, that a dynamo at a magnetic Prandtl number as low as 0.1 could be attained which is considerably lower than the lowest value attainable in the case $P = 1$. The time averaged magnetic and kinetic energy densities for the dynamos of Fig. 17 have been plotted in Fig. 18. Also shown are the corresponding values of Ohmic and viscous dissipation. While in both cases magnetic and kinetic values are quite comparable, a tendency toward an increasing ratio of Ohmic to viscous dissipation with decreasing P_m can be noticed. It appears that for the parameter range that has been investigated the Ohmic dissipation is the most important ingredient in determining the equilibrium amplitude of the magnetic field.

As is well known from Rayleigh-Bénard convection the level of turbulence strongly increases with decreasing Prandtl number for a given supercritical Rayleigh number owing to the increasing importance of the momentum advection term in the equation of motion. This effect is also observed in rotating systems, but it is moderated by the presence

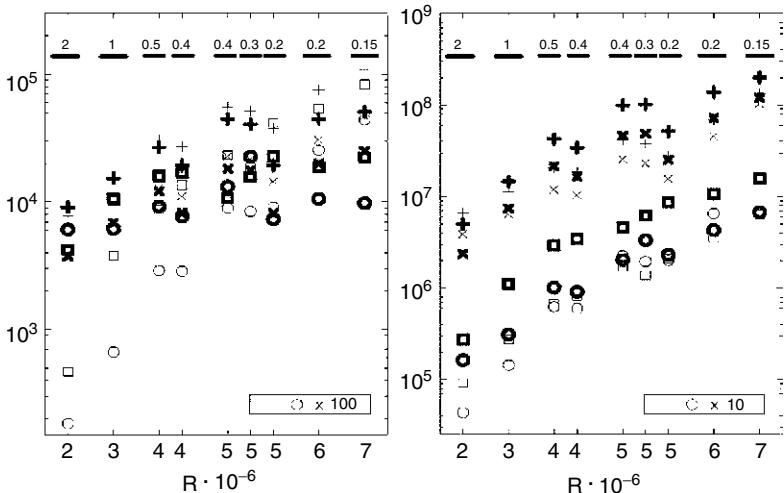


Figure 18. Time averaged magnetic (heavy symbols) and kinetic (light symbols) energy densities (left side) and corresponding dissipations (right side) have been plotted for $P = 0.1$, $\tau = 10^5$. The magnetic Prandtl numbers are indicated in the figure. Energy densities and dissipation densities of axisymmetric field components are indicated by circles (poloidal) and squares (toroidal) while plus-signs and lying crosses denote the corresponding quantities for fluctuating poloidal and toroidal components, respectively. The mean poloidal kinetic energy densities have been multiplied by the factor 100, its dissipation by factor 10.

of the magnetic field. The latter appears to damp the time dependence of convection and the effective Prandtl number appears to be increased by the contribution of Ohmic dissipation. In fact computations of dynamos can be performed more efficiently than corresponding cases of convection without magnetic variables because of lower demands for numerical resolution.

The breaking of the differential rotation by the Lorentz force found in earlier work (Grote and Busse, 2001) continues to be an important effect in low Prandtl number dynamos. The anticorrelation between the strength of the magnetic field and the energy of differential rotation which is given by E_t^m is clearly evident in the time records shown in Fig. 19 (see color insert following page 234). Also apparent is the intermittent nature of the dynamos in this case which varies between a more dipolar and a more hemispherical structure. While the latter type dynamos always exhibit an oscillatory character, changes in the polarity of the dipolar state only occur in an aperiodic fashion in the case of Fig. 19. An example of such a reversal is shown in Fig. 20.

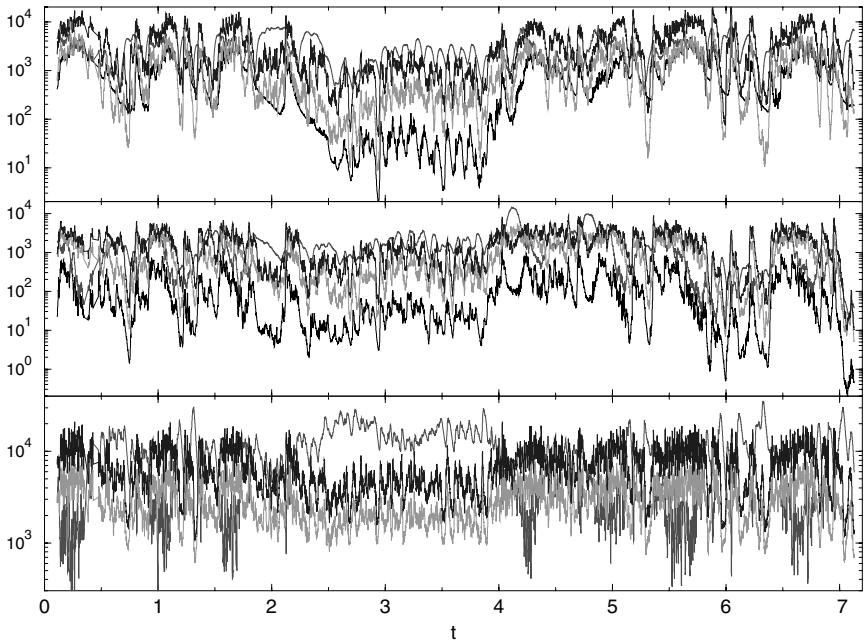


Figure 19. Time series of a convection driven dynamo with $\tau = 3 \times 10^4$, $R = 8.5 \times 10^5$, $P = 0.1$, $P_m = 1$. (See color insert.) The first, second and third plot from the top show energy densities of the dipolar and quadrupolar components of the magnetic field and of the velocity field, respectively. The mean and fluctuating toroidal energy densities are indicated by red and blue lines, while the mean and fluctuating poloidal energy densities are black and green, respectively.

It is remarkable to see that the reversal occurs in a similar way as an oscillation in that magnetic flux with the new polarity is created close to the equatorial plane while the old flux is pushed to higher latitudes and dissipated in the polar regions. We shall return to the phenomenon of dipole oscillations in Section 9.

8. Torsional oscillations

Torsional oscillations have long been postulated for the Earth's core (Braginsky, 1970) and there is some growing evidence that they manifest themselves in the special phenomenon of geomagnetic secular variation known as “jerks”. See the recent paper by Bloxham *et al.* (2002) and references quoted therein. The abrupt changes in the time derivative of the declination of the geomagnetic field have been noticed at

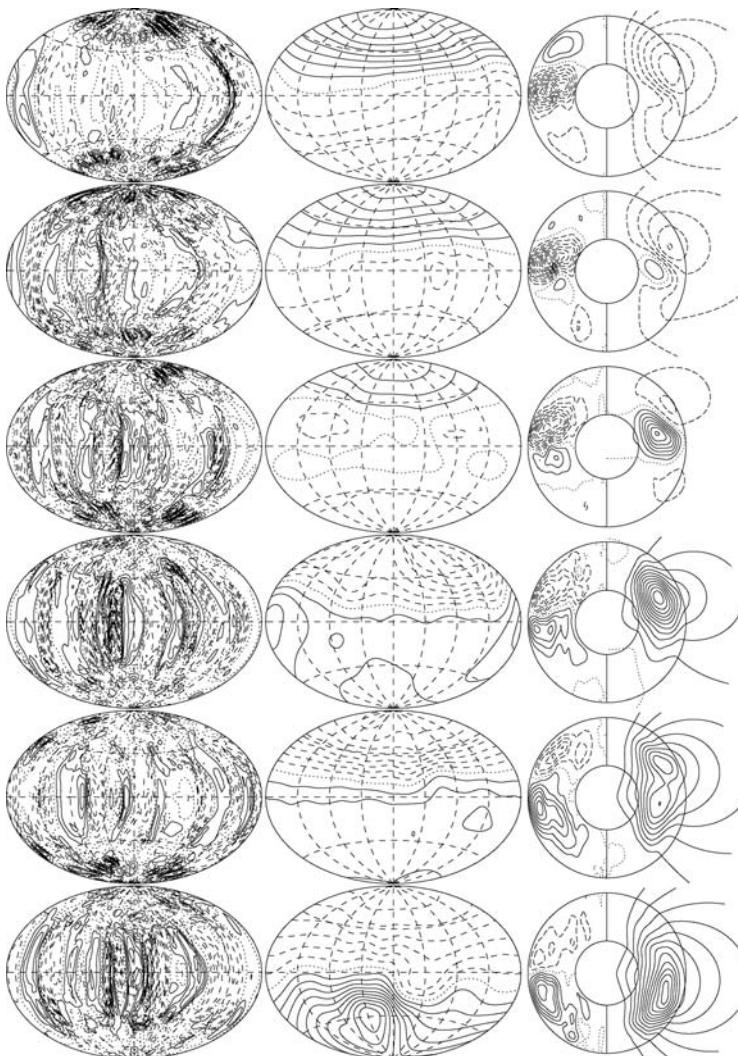


Figure 20. A magnetic field polarity reversal shown by a time sequence of equidistant plots (top to bottom) covering the time span from $t = 5.777$ to $t = 6.017$ of the time series of Fig. 19. The first column shows lines of constant u_r on the surface $r = r_i + 0.5$. The second column exhibits lines of constant B_r on the surface $r = r_o + 0.7$. The circles in the right column indicate lines of constant \bar{B}_ϕ in their left halves and meridional field lines, $r \sin \theta \partial h / \partial \theta = \text{const.}$, in their right halves.

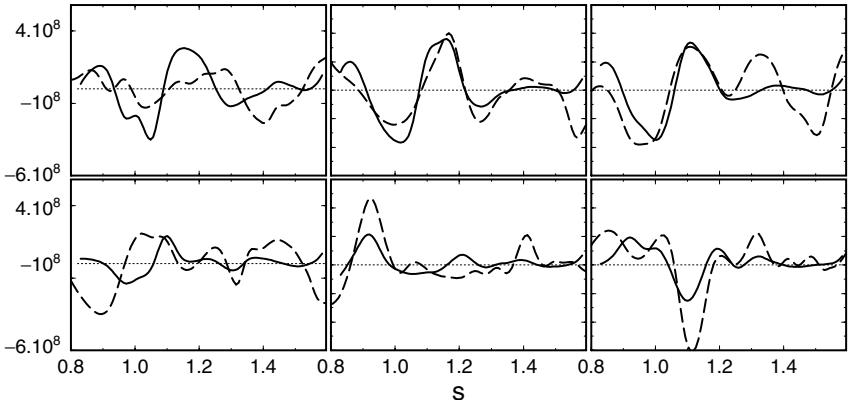


Figure 21. Time sequence of equidistant plots (first row, then second row) with $\Delta t = 0.0003$ of the left-hand side (dashed line) and right-hand side (solid line) of equation (11) in the case $P = 0.1$, $\tau = 10^5$, $R = 3 \cdot 10^6$ and $P_m = 1$.

several observatories in the past, but they were usually interpreted as a more local phenomenon since some observatory records do not exhibit them. The analysis of Bloxham *et al.* (2002) has demonstrated, however, that the morphology of the local geomagnetic field determines whether the “jerks” are observed or not. The fact that we have found spontaneous torsional oscillations in our dynamo computations opens new possibilities for a connection between geomagnetic observations and dynamo theory. An example of a torsional oscillation is shown in Fig. 21. The dashed and solid lines correspond to left and right sides of the dissipationless equation of torsional oscillations, which have been

$$\frac{\partial^2}{\partial t^2} \bar{u}_\phi = \frac{1}{s} \left[\frac{s}{z_+ - z_-} \int_{z_+}^{z_-} \bar{B}_s^2 dz \frac{\partial}{\partial s} \bar{u}_\phi \right], \quad (11)$$

extracted from the results of the dynamo computations. Since the latter include the dissipation effects in the presence of turbulent convection which are neglected in (11), only a rough agreement between left- and right-hand sides of (11) can be expected. The s -coordinate in (11) refers to the distance from the axis and z_\pm denote the values of $\pm\sqrt{r_o^2 - s^2}$, where coaxial cylinders with radius s intersect the outer spherical boundary.

9. Oscillating dipolar dynamos

In one of the first studies (Grote *et al.*, 2000) of the present series the impression was gained that dipolar dynamos are non-oscillatory while hemispherical and quadrupolar always occur in an oscillatory manner. Since then it has become apparent that dipolar dynamos also often oscillate. An example is shown in Fig. 22. As in the case of hemispherical and quadrupolar dynamos magnetic flux of the new polarity first appears close to the equatorial plane and then propagates to higher latitudes where it is finally dissipated. A characteristic feature of dipolar oscillations appears to be that they are not entirely symmetric with respect to the two signs of the magnetic field. This asymmetry can already be noticed in the case of Fig. 22, but it is more strongly exhibited in the case of Fig. 23. Finally, the asymmetry may become so strong that the oscillations can hardly be inferred from watching the poloidal field at some distance from the outer boundary of the shell as is evident from Fig. 24. We are calling this version of the oscillatory dipolar dynamo the “invisible” one. Only close to the equator does the poloidal field exhibit a reversal at distances not too far from the sphere.

The “invisible” oscillating dynamo occurs at Prandtl numbers larger than unity. These dynamos are characterized by strong polar tubes of

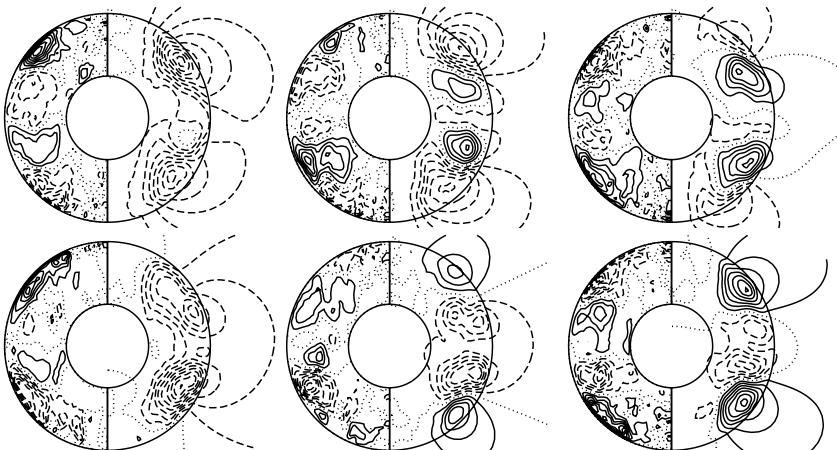


Figure 22. Oscillating dipolar dynamo with $\Delta t = 0.003$ (clockwise) in the case of $P = 1$, $\tau = 5 \times 10^3$, $R = 1.4 \times 10^6$ and $P_m = 1$. The left half of each plot shows lines of constant \bar{B}_ϕ while the right half indicates meridional field lines, $r \sin \theta \partial h / \partial \theta = \text{const}$.

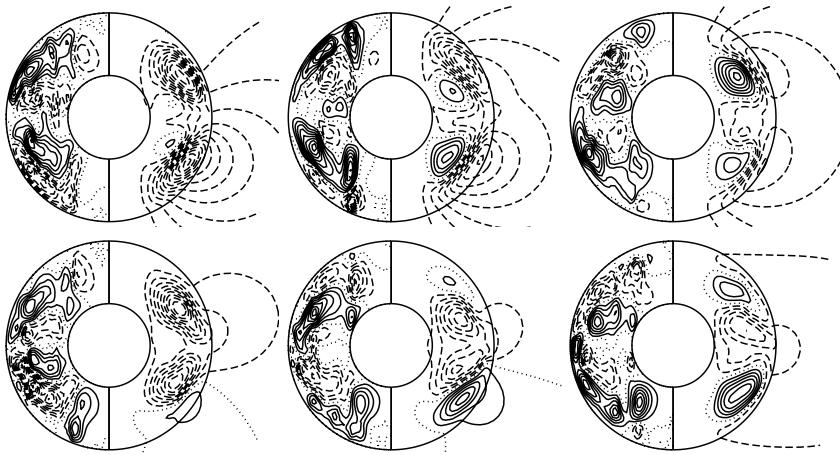


Figure 23. Oscillating dipolar dynamo for $P = 1$, $\tau = 10^4$, $R = 5.4 \times 10^5$ and $P_m = 4$. The time sequence with $\Delta t = 0.04$ starts with upper left plot and continues clockwise to the lower left plot such that about a period is covered. The left half of each plot shows lines of constant \bar{B}_ϕ while the right half indicates meridional field lines, $r \sin \theta \partial h / \partial \theta = \text{const}$.

zonal magnetic flux which do not participate in the oscillation. The latter is confined to the region outside the tangent cylinder. Although the poloidal field is also located mostly outside the tangent cylinder the presence of the polar flux tubes prevents its participation in the oscillation except for the small reversal region close to the equator.

Another type of dipole oscillations which is frequently encountered in low Prandtl number dynamos is shown in Fig. 25. The magnetic field structure is characterized by a strong hemispherical component. But this component switches nearly periodically between the northern and the southern hemisphere. While the mean azimuthal field becomes almost quadrupolar during the switchover (see the plot 5 of Fig. 25) the poloidal field attains its most dipolar structure at this point.

In general oscillating dipolar dynamos are found in the region of the parameter space between the regimes of non-oscillatory dipolar dynamos and of hemispherical dynamos. This distinction becomes blurred, of course, at higher values of the Rayleigh number when the onset of convection in the polar region tends to destroy the symmetry properties of dynamos.

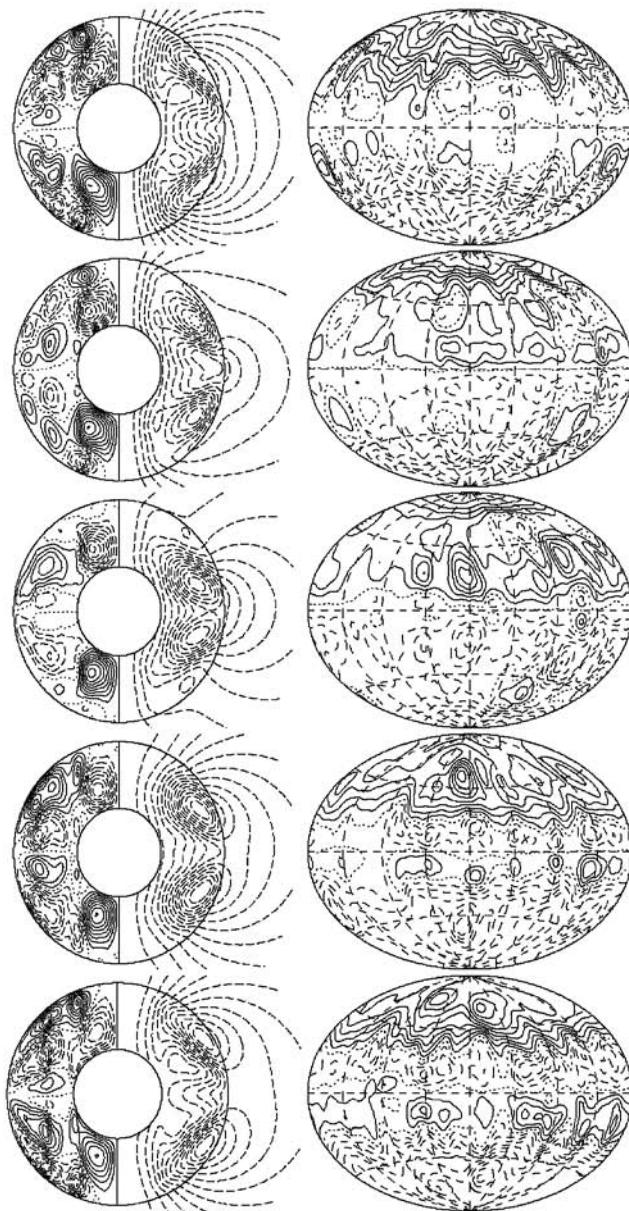


Figure 24. “Invisible” oscillating dynamo with $\Delta t = 0.06$ in the case of $P = P_m = 5$, $\tau = 5 \times 10^3$ and $R = 6 \times 10^5$. The circles in the first column indicate lines of constant \bar{B}_ϕ in their left halves and meridional field lines, $r \sin \theta \partial h / \partial \theta = \text{const.}$, in their right halves. The second column exhibits lines of constant B_r on the surface $r = r_o + 0.4$.

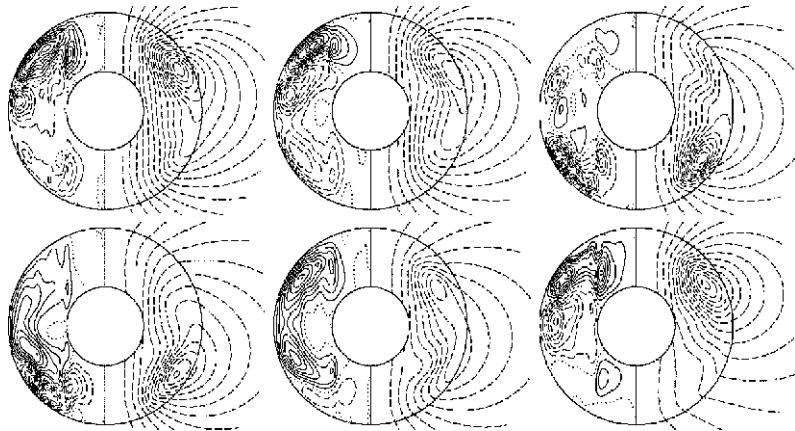


Figure 25. Oscillating dynamo for $P = 0.1$, $\tau = 10^5$, $R = 4 \times 10^6$ and $P_m = 0.4$. The time sequence of equidistant plots (first row, then second row) with $\Delta t = 0.012$ covers about one period. The left half of each plot shows lines of constant \bar{B}_ϕ while the right half shows meridional field lines, $r \sin \theta \partial h / \partial \theta = \text{const}$.

10. Discussion and outlook

Throughout the preceding sections an electrically insulating core has been assumed. A finite conductivity of the same order as that of the fluid shell may be expected to influence the time dependence of the dynamo (Hollerbach and Jones, 1995). When such a finite core conductivity was introduced in the course of the present computations, however, little influence was seen. The magnetic field penetrating into the inner core is rather weak and has little effect on the chaotic dynamo in the outer core. This conclusion was reached previously by Wicht (2002).

A strong effect is exerted by no-slip conditions in place of stress-free conditions at the outer boundary of the spherical shell. In particular, the differential rotation is significantly reduced in the case of $P = 1$ (Kutzner and Christensen, 2002) and the ω -effect of the dynamo is nearly absent. Dynamos with a no-slip condition on the boundary thus resemble those obtained at much larger values of P . It must be expected, however, that the difference between results for stress-free and those for no-slip boundaries will vanish in the limit of large τ . It therefore appears to be preferable to use stress-free conditions at the outer boundary for planetary applications.

The ultimate goal of dynamo simulations are results which can be compared quantitatively with the properties of the geomagnetic field and its variations in time. Reversals and torsional oscillations appear

to be particularly attractive possibilities in this respect. The low Prandtl number regime is well suited for this purpose because viscous dissipation plays a lesser role than in the case of $P \geq 1$. Higher values of the Coriolis number can thus be reached in the numerical computations. While current estimates of the effective Prandtl number in the outer core of the Earth range around $P = 0.1$ it could well be that the diffusivity of light elements should be used instead of the thermal diffusivity. Compositional buoyancy rather than thermal buoyancy is generally believed to be the major energy source of convection (Lister and Buffett, 1995) and the role played by multiple sources of buoyancy will thus require more detailed studies. The possibility of new dynamical phenomena associated with double buoyancy has been pointed out (Busse, 2002b) and is presently investigated numerically.

References

- Ahlers, G. and Xu, X., “Prandtl number dependence of heat transport in turbulent Rayleigh-Bénard convection,” *Phys. Rev. Lett.* **86**, 3320–3323 (2001).
- Ardes, M., Busse, F. H. and Wicht, J., “Thermal convection in rotating spherical shells,” *Phys. Earth Planet Inter.* **99**, 55–67 (1997).
- Bloxham, J., Zatman, S. and Dumberry, M., “The origin of geomagnetic jerks,” *Nature* **240**, 65–68 (2002).
- Braginsky, S. I., “Torsional magnetohydrodynamic vibrations in the Earth’s core and variations in day length,” *Geomagn. Aeron.* **10**, 1–8 (1970).
- Busse, F. H., “Thermal instabilities in rapidly rotating systems,” *J. Fluid Mech.* **44**, 441–460 (1970).
- Busse, F. H., “Recent developments in the dynamo theory of planetary magnetism,” *Annu. Rev. Earth Plan. Sci.* **11**, 241–268 (1983).
- Busse, F. H., “Homogeneous dynamos in planetary cores and in the laboratory,” *Annu. Rev. Fluid Mech.* **32**, 383–408 (2000).
- Busse, F. H., “Convection flows in rapidly rotating spheres and their dynamo action,” *Phys. Fluids* **14**, 1301–1314 (2002a).
- Busse, F. H., “Is low Rayleigh number convection possible in the Earth’s core?” *Geophys. Res. Letts.* **29**, GLO149597 (2002b).
- Busse, F. H., Grote, E. and Simitev, R., “Convection in rotating spherical shells and its dynamo action,” in: *Earth’s Core and Lower Mantle* (Eds. C. A. Jones, A. M. Soward and K. Zhang), pp. 130–152, Taylor and Francis (2003).
- Busse, F. H., Grote, E., and Tilgner, A., “On convection driven dynamos in rotating spherical shells,” *Studia geoph. et geod.* **42**, 211–223 (1998).
- Chandrasekhar, S., *Hydrodynamic and Hydromagnetic Stability*, Clarendon Press, Oxford (1961).

- Dormy, E., Valet, J.-P. and Courtillot, V., "Numerical models of the geodynamo and observational constraints," *Geochem. Geophys. Geosyst.* **1**, Paper No. 2000GC 000062 (2000).
- Eltayeb, I. A., "Hydromagnetic convection in a rapidly rotating fluid layer," *Proc. R. Soc. Lond. A* **326**, 229–254 (1972).
- Eschrich, K.-O. and Rüdiger, G., "A second-order correlation approximation for thermal conductivity and Prandtl number of free turbulence," *Astron. Nachr.* **304**, 171–180 (1983).
- Glatzmaier, G. A. and Roberts, P. H., "A three-dimensional convection dynamo solution with rotating and finitely conducting inner core and mantle," *Phys. Earth Planet. Inter.* **91**, 63–75 (1995).
- Grote, E. and Busse, F. H., "Dynamics of convection and dynamos in rotating spherical fluid shells," *Fluid Dyn. Res.* **28**, 349–368 (2001).
- Grote, E., Busse, F. H. and Simitev, R., "Buoyancy driven convection in rotating spherical shells and its dynamo action," in: *High Performance Computing in Science and Engineering '01* (Eds. E. Krause and W. Jäger), pp. 12–34, Springer (2001).
- Grote, E., Busse, F. H. and Tilgner, A., "Computations of convection driven spherical dynamos," in: *High Performance Computing in Science and Engineering '01* (Eds. E. Krause and W. Jäger), pp. 13–25, Springer (1999).
- Grote, E., Busse, F. H., and Tilgner, A., "Regular and chaotic spherical dynamos," *Phys. Earth Planet. Inter.* **117**, 259–272 (2000).
- Fearn, D. R., "Thermal and magnetic instabilities in a rapidly rotating fluid sphere," *Geophys. Astrophys. Fluid Dynam.*, **14**, 103–126 (1979).
- Hollerbach, R. and Jones, C., "On the magnetically stabilizing role of the Earth's inner core," *Phys. Earth Planet. Inter.* **87**, 171–181 (1995).
- Jones, C. A., Soward, A. M. and Mussa, A. I., "The onset of thermal convection in a rapidly rotating sphere," *J. Fluid Mech.* **405**, 157–179 (2000).
- Kono, M. and Roberts, P. H., "Recent geodynamo simulations and observations of the geomagnetic field," *Rev. Geophys.* **40**, 1013–1061 (2002).
- Kutzner, K. and Christensen, U., "From stable dipolar towards reversing numerical dynamos," *Phys. Earth Planet. Inter.* **131**, 29–45 (2002).
- Lister, J. R. and Buffett, B. A., "The strength and efficiency of thermal and compositional convection in the Earth's core," *Phys. Earth Planet. Inter.* **91**, 17–30 (1995).
- Pino, D., Mercader, I. and Net, M., "Thermal and inertial mode of convection in a rapidly rotating annulus," *Phys. Rev. E* **61**, 1507–1517 (2000).
- Simitev, R. and Busse, F. H., "Patterns of convection in rotating spherical shells," *New J. Phys.*, in press (2003).
- Tilgner, A. and Busse, F. H., "Finite amplitude convection in rotating spherical fluid shells," *J. Fluid Mech.* **332**, 359–376 (1997).
- Wicht, J., "Inner-core conductivity in numerical dynamo simulations," *Phys. Earth Planet. Inter.* **132**, 281–302 (2002).
- Yano, J.-I., "Asymptotic theory of thermal convection in rapidly rotating systems," *J. Fluid Mech.* **243**, 103–131 (1992).

- Zhang, K., “On coupling between the Poincaré equation and the heat equation,” *J. Fluid Mech.* **268**, 211–229 (1994).
- Zhang, K., “On coupling between the Poincaré equation and the heat equation: no-slip boundary condition,” *J. Fluid Mech.* **284**, 239–256 (1995).
- Zhang, K. and Busse, F. H., “On the onset of convection in rotating spherical shells,” *Geophys. Astrophys. Fluid Dynam.* **39**, 119–147 (1987).
- Zhang, K. and Busse, F. H., “Convection driven magnetohydrodynamic dynamos in rotating spherical shells,” *Geophys. Astrophys. Fluid Dynam.* **49**, 97–116 (1989).

13 Laboratory experiments on liquid metal dynamos and liquid metal MHD turbulence

Stephan Fauve

*Laboratoire de Physique Statistique de l'Ecole Normale Supérieure CNRS UMR 8550, 24 Rue Lhomond,
75231 Paris cedex 05, France*

Daniel P. Lathrop

IREAP, IPST, and the Department of Physics, University of Maryland, College Park, MD 20742, USA

1. Introduction

Most of astrophysical objects that involve the flow of a large enough volume of electrically conducting fluid possess a magnetic field. We recall in [Table 1](#) orders of magnitude of the magnetic fields and some associated relevant parameters for planets, stars and our Galaxy (Zeldovich *et al.*, 1983).

It is difficult to compare the data in Table 1 because these astrophysical objects have strongly different physical properties. A magnetohydrodynamic description is probably valid for the Earth core which is made of liquid metal, but classical hydrodynamics certainly breaks down both for rarefied plasmas (where the mean free path is no longer small compared to the characteristic length on which the velocity varies), and for very dense stars where quantum and relativistic effects are important. However, we observe that the magnetic field B increases with density ρ . It does not seem to depend monotonically on the typical size L , but considering the typical magnetic energy of the object $B^2L^3/2\mu_0$ where μ_0 is the magnetic permeability of vacuum, we find the expected ordering from the Galaxy to the Earth. We may also consider the typical

Table 1. Approximate values of the magnetic fields for astrophysical objects

	Our galaxy	Sun	Jupiter	Earth	White dwarfs	Neutron stars
B (T)	10^{-10}	10^{-4}	4×10^{-4}	10^{-4}	$10^2 - 10^4$	$10^6 - 10^9$
ρ (kg/m ³)	2×10^{-21}	1	10^3	10^4	10^{10}	10^{19}
L (m)	10^{19}	2×10^8	5×10^7	3×10^6	10^7	10^4
η (m ² /s)	10^{17}	10^3	10	3		
L^2/η (years)	10^{13}	10^6	10^7	10^4		
$B^2 L^3 / 2\mu_0$	10^{43}	4×10^{22}	10^{22}		2×10^{17}	
$B^2 L \eta / 2\mu_0$	10^{22}	10^9		4×10^7	10^6	

value of Joule dissipation. To wit, we divide the magnetic energy by the characteristic magnetic diffusion time L^2/η , with $\eta = (\mu_0\sigma)^{-1}$ where σ is the electrical conductivity of the medium. We thus get an idea of the minimum amount of power which is necessary to maintain the magnetic field against Joule dissipation. Again, the ordering from the Galaxy to the Earth looks reasonable. We note that these energies and powers have been strongly underestimated. First, they are estimated from the visible part of the surface magnetic field in the case of the Sun or the planets. Thus the poloidal field is underestimated and the toroidal component is ignored. Second, we have assumed that the lengthscale of the gradients of the magnetic field is the size L of the conducting medium. Magnetic energy at smaller scales will lead to a shorter diffusion timescale and thus to a higher dissipated power. The very large astrophysical scales lead to long typical diffusion times L^2/η for the magnetic field. However, for the magnetic fields of the sun or planets, this timescale is smaller than the age of these objects, even when considering the largest lengthscale L and without invoking turbulent diffusivity. Moreover, these magnetic fields have a complex dynamics in time and space. One thus needs to find their generation mechanism.

It has been first proposed by Larmor (Larmor, 1919) that the magnetic fields of “celestial bodies” may be produced by internal motions of conducting matter. Therefore, we have to take into account the typical magnitude of their velocity field, V , and the kinetic viscosity, ν , and thus obtain two independent dimensionless parameters for our problem. We can choose the magnetic Reynolds number, R_m , and the kinetic

Reynolds number, Re :

$$R_m = \frac{V L}{\eta}, \quad Re = \frac{V L}{\nu}. \quad (1)$$

Their ratio is the magnetic Prandtl number, $Pr_m = \mu_0 \sigma \nu = \nu/\eta$. It only depends on the fluid and is usually very small (except for some astrophysical plasmas). Magnetic field self-generation can be obtained only for large R_m for which Joule dissipation can be overcome (for most known fluid dynamos, the dynamo threshold R_{mc} is roughly in the range 10–100). Therefore, the kinetic Reynolds number is very large and the flow is strongly turbulent. This is the case of the Earth or the Sun and also the context of experiments with liquid metals for which $Pr_m < 10^{-5}$. Several dynamo regimes are displayed in Fig. 1. Direct numerical simulations are possible for values of Pr_m orders of magnitude larger than the realistic ones for the Sun, the Earth or laboratory experiments, first because it is not possible to handle a too large difference between the timescale of diffusion of the magnetic field and the one of momentum. Second, although the dynamo threshold should be some curve $R_{mc} = f(Pr_m)$ in Fig. 1, one does not expect a large variation of R_{mc} on Pr_m for a given forcing of the flow. Therefore, a small Pr_m dynamo occurs for

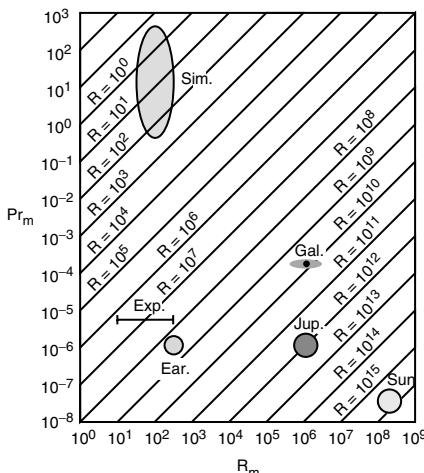


Figure 1. Relevant parameters for the Earth (Ear.), simulations (Sim.), Jupiter (Jup.), the Sun, estimates for galaxy (Gal.), and experiments (Exp.), showing the Reynolds number Re , magnetic Prandtl number Pr_m , and the magnetic Reynolds number R_m . Note that $R_m = Re Pr_m$ (some estimates from Zeldovich *et al.*, 1983).

large Re and necessitates the resolution of the small spatial scales generated by turbulence. Strongly developed turbulence is also unpleasant for the experimentalist. Indeed, the power needed to drive a turbulent flow scales like $P \propto \rho L^2 V^3$ and we have

$$R_m \propto \frac{1}{\eta} \left(\frac{PL}{\rho} \right)^{1/3}. \quad (2)$$

This formula has simple consequences: first, taking liquid sodium (the liquid metal with the highest electric conductivity), $\eta \approx 0.1 \text{ m}^2/\text{s}$, $\rho \approx 10^3 \text{ kg m}^{-3}$, and with a typical lengthscale $L \approx 1\text{m}$, we get $P \approx R_m^3$, thus a mechanical power larger than 100 kW is needed to reach a dynamo threshold of the order of 50. Second, it appears unlikely to ever operate experimental dynamos at R_m large compared with R_{mc} . Indeed, it costs eight times more power to reach $2R_{mc}$ than to reach the dynamo threshold. In conclusion, most experimental dynamos should have the following characteristics:

- they bifurcate from a strongly turbulent flow regime,
- they operate in the vicinity of their bifurcation threshold.

Early experiments have been performed by Lehnert (1957). A swirling flow has been generated by rotating a disk in a cylindrical vessel containing 58 l of liquid sodium. Although a self-generated dynamo regime has not been reached, induction measurements have been performed with an applied axial field. The generation of a toroidal (ω -effect) and poloidal field components have been observed. Large values of R_m (up to roughly 25) have been achieved in the liquid sodium cooling system of breeder nuclear reactors but no clear-cut results have been observed concerning the dynamo effect. The generation of magnetic fields has been reported in some cases, but ascribed to a combined effect of a thermoelectric current due to the temperature difference between liquid sodium and steel boundaries and magnetic field amplification by the flow (Kirk et al., 1984).

The first homogeneous dynamo has been operated by Lowes and Wilkinson (1963, 1968) by rotating two cylindrical ferromagnetic cylinders in electrical contact in cylindrical holes drilled in a block of the same material. Self-generation of a magnetic field was achieved at high enough rotation rates as predicted by rotor-dynamo models (Herzenberg, 1958). Using a flow that consists of solid body rotation in two separate domains in electrical contact allowed to generate a dynamo without

using the largest amount of power to compensate turbulent dissipation. The first homogeneous fluid dynamos have been operated more than three decades later in liquid sodium with a much larger amount of power: in Karlsruhe (Stieglitz and Müller, 2001) using a flow in an array of pipes set up in order to mimic a spatially periodic flow proposed by G.O. Roberts (1972), and in Riga (Gailitis *et al.*, 2001) using a Ponomarenko-type flow (Ponomarenko, 1973). Although there were no doubts about self-generation of magnetic fields by Roberts' or Ponomarenko-type laminar flows, these experiments have displayed several interesting features:

- The observed thresholds are in rather good agreement with theoretical predictions (Busse *et al.*, 1996; Rädler *et al.*, 1998; Gailitis *et al.*, 2002) made by considering only the laminar mean flow and neglecting the small-scale turbulent fluctuations that are present in both experiments.
- The nature of the dynamo bifurcation, stationary for the Karlsruhe experiment or oscillatory (Hopf) in the Riga experiment, is also in agreement with laminar models.
- On the contrary, the saturation level of the magnetic field, due to the back-reaction of the Lorentz force on the flow, cannot be predicted with a laminar flow model and different scaling laws exist in the supercritical dynamo regime depending on the magnitude of the Reynolds number (Pétrélis and Fauve, 2001).
- Although field reversals have not been observed in the Karlsruhe and Riga experiments, small-scale turbulent fluctuations of the magnetic field are well developed.

These observations raise the following questions:

- What is the effect of turbulence, or of the magnitude of the Reynolds number, on the dynamo threshold R_{mc} ? Is it possible to design a laboratory dynamo for which R_{mc} strongly depends on Re for a given driving of the flow?
- What is the mechanism responsible for magnetic field fluctuations in the vicinity of the dynamo threshold: an on-off intermittency effect (Sweet *et al.*, 2001) or chaotic advection of the mean magnetic field by the turbulent flow?
- What is the mechanism for field reversals? Is it possible to observe them in laboratory experiments?

Our interest for the above problems is of course biased by our background: as already mentioned, laboratory dynamos operate in the vicinity of threshold but at very high values of the Reynolds number. This gives rise to a very interesting example of instability that differs in many respects from usual hydrodynamical instabilities. The dynamo bifurcation occurs from a base state which is fully turbulent and this has strong consequences on the saturation level of the magnetic energy in the supercritical regime and also on the nature of the magnetic field fluctuations about their mean. In some cases, we also expect that this may affect the value of the dynamo threshold itself.

There are of course other questions motivated by astrophysical or geophysical dynamo regimes. Some of them, which are also of fundamental interest, concern the effect of rotation.

- Are rotating flows helpful for self-generation, i.e., do they lead to smaller values of the dynamo threshold or to a smaller value of the minimum amount of power needed to reach the dynamo threshold?
- Is the scaling for the magnetic energy above the dynamo threshold changed in the case of rapidly rotating flows, as suggested in the case of convective dynamos (Roberts, 1988)?
- Is rotation the key ingredient to get a subcritical bifurcation to the dynamo regime?

We will try to discuss the above questions in connection with existing or planned laboratory dynamo experiments. The paper is organized as follows: in Section 2, we analyze the results of Karlsruhe and Riga experiments with a particular emphasis on the saturation of the magnetic field above the dynamo threshold. Section 3 is devoted to experiments involving fully three-dimensional flows of liquid sodium, i.e., with less geometrical constraints than the Karlsruhe and Riga experiments. Motivations for rapidly rotating flows together with the description of existing or planned experiments are presented in Section 4. Finally, the main open questions are emphasized in Section 5.

2. Self-generation in Riga and Karlsruhe

The past decade has seen unprecedented progress in experiments aimed at understanding the self-generation of magnetic fields by flowing conductors. This progress has been made possible by a number of factors: improved technology for handling liquid metal flows, a better

understanding of the instabilities involved in the relevant class of partial differential equations, and the key ingredient — the persistence of a few unswerving individuals.

Within a year, both experiments in Riga, Latvia, and in Karlsruhe, Germany, produced saturated magnetic fields from experiments employing flowing liquid sodium. In this article we will not enter the otherwise amusing debate of precedence, but will instead focus on the results, and what can be learned by a careful examination of the field amplitude and power dissipation observations.

In the first two subsections here, we outline the published results of the two experimental groups. In the third, we focus on our own interpretations of these phenomena, and point to lessons and issues for future experiments.

2.1. *From the Ponomarenko dynamo to the Riga experiment*

The experiment carried out by Gailitis, Lielausis, Platacis, Dement'ev, Cifersons, Gerbeth, Gundrum, Stefani, Christen and Will self-generated in a flow substantially motivated by the simple model of a dynamo known as the Ponomarenko dynamo (Ponomarenko, 1973). In that simple analytical model a conducting cylinder moves axially while rotating — a helical motion, while the conductor outside the cylinder is stationary. That motion, in an unbounded domain, gives rise to an axially periodic in space, periodic in time magnetic field instability. Fitting this type of motion into a (necessarily finite in space) laboratory experiment, along with having sufficiently large magnetic Reynolds numbers, allowed the Riga group, after long persistence, to self-generate a magnetic field.

Their flow is driven by a single propellor, generating helical flow down a central cylindrical cavity (see Fig. 2b). The return flow is in an annulus surrounding this central flow. Just outside of these nested features is an annular region of nearly stationary flow. These chambers are divided by thin stainless steel walls, which the experimentalists took great care to ensure were wetted to allow currents to flow through them.

Figure 3 displays the growth and saturation of a time periodic magnetic field at high enough rotation rate. The experimental results displayed dynamo growth rates in good agreement with kinematic theory (Gailitis *et al.*, 2002). In addition, they give information that a kinematic dynamo model cannot predict. In particular the Riga group had made detailed observations of the magnetic field saturation value and the power dissipation needed to drive the flow. These measurements give indications of the effect of Lorentz forces in the flow in order to

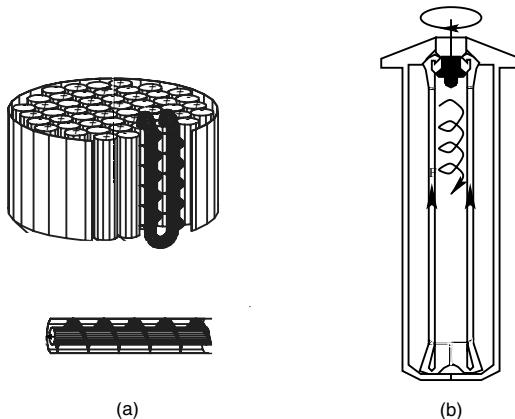


Figure 2. Schematics for the experiments from Karlsruhe (a) and Riga (b) which show how helical flow is forced by guiding the sodium through steel channels (from Stieglitz *et al.*, 2002 and Gailitis *et al.*, 2002).

bring about the saturated state. Dynamo generation does correspond to an increase in the required mechanical power ([Fig. 4a](#)). However, a puzzling result is displayed in Fig. 4b: the amplitude of the magnetic field for supercritical rotation rates does not seem to show the universal $\sqrt{R_m - R_{mc}}$ law. In addition, the form of the law seems to depend on the location of the measurement point (Fig. 4b).

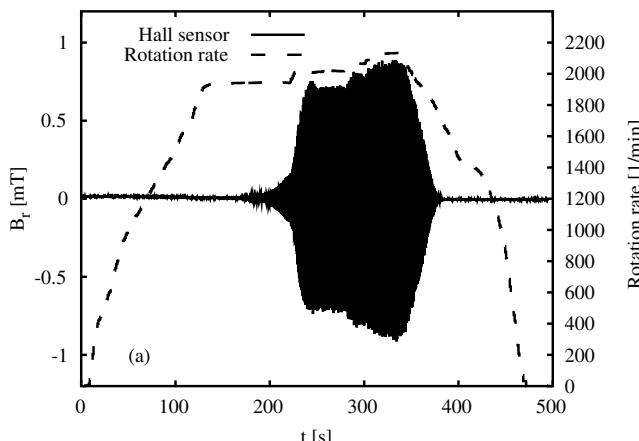


Figure 3. Time traces of the magnetic field from the Riga experiment. Note how the magnetic field oscillations grow after the rotation rate is increased above a threshold (from Gailitis *et al.*, 2002).

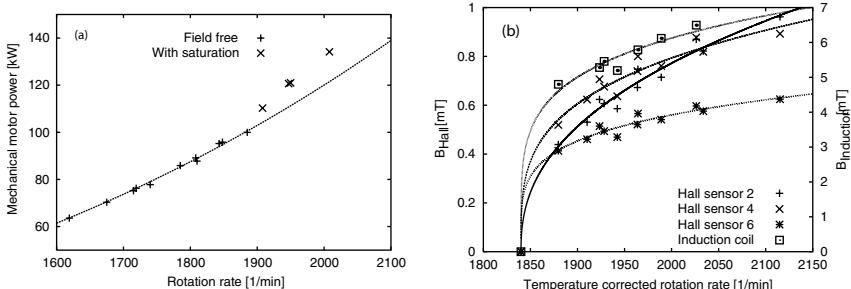


Figure 4. (a) Results for the power consumption vs. rotation rate from the Riga experiment (see Gailitis *et al.*, 2002, for details of the experiment). (b) Magnetic field amplitude as the rotation rate is raised above the critical rotation rate (data courtesy of Frank Stefani).

2.2. From the Roberts flow to the Karlsruhe experiment

The experiment in Karlsruhe, Germany, was motivated by a kinematic dynamo model developed by G.O. Roberts (Roberts, 1972). The idealized flow extends to all space, while by necessity, the experiment is a spatially truncated version. The flow has a two-dimensional square array of vortices, with axial flow in each vortex chosen such that every vortex has the same sign helicity. Such a flow is quite efficient at self-generation (in the sense of generating at a low magnetic Reynolds number based on the lengthscale of the vortices). To establish an analogous flow in the laboratory, Steiglitz and Müller used three external pumped loops of sodium to drive a flow in an intricate array of baffled tubes (see Fig. 2a). The flow rate of both the central axial component of the flow, and the helical outer portion of the flow could be independently adjusted. When the flow rates exceeded a critical value, the system generated a quasi-steady magnetic field as shown in Fig. 5. This generation comes at a cost in the power necessary to drive the flow. Both the pressure drop above threshold and the magnetic field amplitude (Fig. 6) were carefully measured in those experiments. Relatively small fluctuations were observed in the magnetic field amplitude, which shows the effects of the underlying small-scale turbulence.

2.3. Expected and unexpected results: the nature of the dynamo bifurcation

It is well known that the nature of an instability and the form of the equation governing the amplitude of the neutral modes in the vicinity

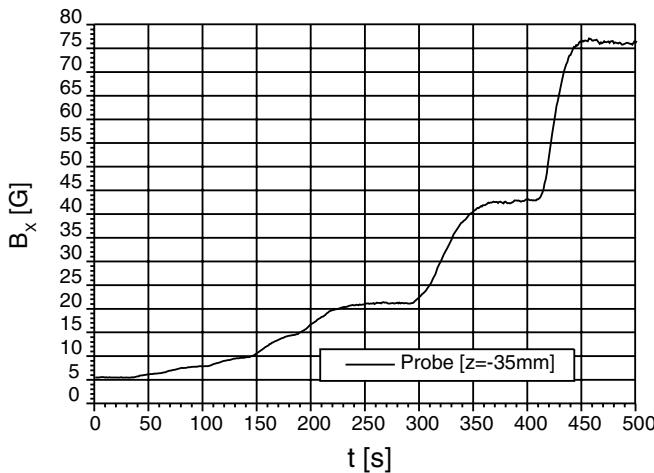


Figure 5. Magnetic field time recordings from the Karlsruhe experiment show small turbulent fluctuations atop a large steady self-generated field (figure from Stieglitz and Müller, 2002).

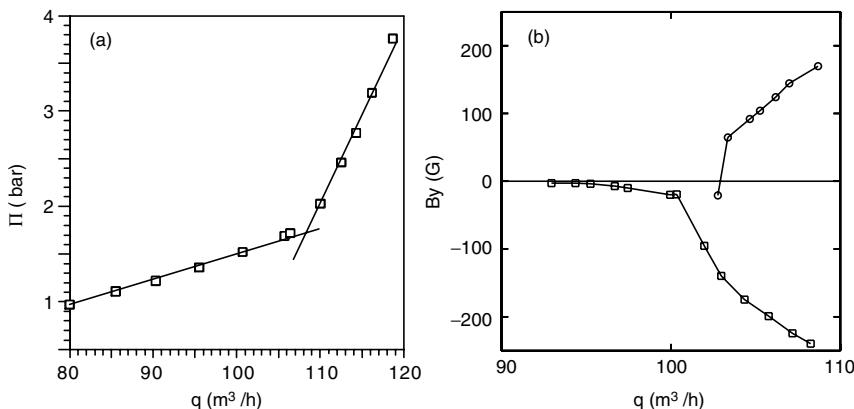


Figure 6. (a) Observed increase in pressure after the onset of magnetic field self-generation in Karlsruhe. Below the onset flow rate, the pressure conforms to the expected hydraulic dependence ($\Pi \sim q^2$, where Π is the pressure drop and q the volume flow rate). The dependence appears linear due to the limited range in flow rates. Above the onset, an extra pressure drop is observed. (b) The magnetic field amplitude increases above the critical flow rate in the Karlsruhe experiment. Another branch of self-generation is only available by imposing an initial field. This other branch is disconnected from the main branch. The perturbation of the Earth's field on the experiment has been evoked to explain the lack of up-down symmetry in this bifurcation diagram (Tilgner and Busse, 2002) (figure from Stieglitz and Müller, 2002).

of the bifurcation threshold is affected by symmetries (e.g., Chossat and Lauterbach, 2000). This type of argument can be used to understand why the dynamo generated in the Riga experiment is oscillatory whereas the one in Karlsruhe is stationary (Fauve and Pétrélis, 2003). Ponomarenko's flow is invariant under translations along the z axis, $z \rightarrow z + z_0$. However, it is not invariant under $z \rightarrow -z$ (the helicity is changed to its opposite). In other words, the $+z$ and $-z$ directions are not equivalent. When the neutral mode breaks the translation invariance along z at bifurcation threshold, it naturally begins to drift in a direction depending on the sign of the helicity. This gives an oscillatory behavior generated by a Hopf bifurcation. G. O. Roberts' flow also has a non-zero mean helicity. Moreover, it looks like a periodic array of helical flows with axis parallel to z . Why does the Riga experiment display a Hopf bifurcation while the Karlsruhe one does not? This is due to an additional symmetry: a rotation of angle π along the axis $x = y$ leaves the flow unchanged but would change the direction of propagation for an oscillatory instability. We observe that both the turbulent fluctuations as well as the additional boundary conditions involved in the Karlsruhe and Riga experiments, do not affect the nature of the bifurcations predicted by the laminar flow models of Roberts and Ponomarenko.

2.4. Expected and unexpected results: sensitivity of the dynamo threshold to flow perturbations

Karlsruhe and Riga experiments have shown that the dynamo threshold of Roberts or Ponomarenko type flows are almost unchanged because of the presence of turbulent fluctuations. Writing $\mathbf{V}(\mathbf{r}, t) = \langle \mathbf{V}(\mathbf{r}) \rangle + \mathbf{v}(\mathbf{r}, t)$, where $\langle \cdot \rangle$ stands for temporal average, a correct prediction of the dynamo threshold can be obtained when neglecting the fluctuating velocity field $\mathbf{v}(\mathbf{r}, t)$ (Rädler *et al.*, 1998; Busse *et al.*, 1996; Gailitis *et al.*, 2002).

In addition, it is clear from the same studies that the precise profile of the mean flow $\langle \mathbf{V}(\mathbf{r}) \rangle$ or the boundary conditions (solid boundaries with an electrical conductivity different from the one of the fluid), do not strongly affect the threshold either. A smooth modification of R_{mc} is observed but we emphasize the contrast with the fully three-dimensional flows of Section 3 for which a minor modification of the mean flow strongly change R_{mc} (see Section 3). This relative insensitivity of R_{mc} to small perturbations of the Roberts or Ponomarenko type flows may be due to the two-dimensional nature of these flows (they do

not depend on one spatial Cartesian coordinate). This is also roughly achieved in the Karlsruhe and Riga experiments because of the elongated aspect ratio of their helical motions.

2.5. Expected and unexpected results: scaling law for the magnetic energy generation above onset

Since simple laminar models correctly describe the kinematic dynamo regimes of the Karlsruhe and Riga experiments, we may try to use them in order to describe the nonlinear regimes above the dynamo threshold. To wit, we have to take into account the back-reaction of the magnetic field generated above the dynamo threshold on the velocity field. We thus determine the amplitude of the magnetic field as a function of the distance to the dynamo threshold. This has been performed for the Roberts' flow (Gilbert and Sulem, 1990; Tilgner and Busse, 2001) and for a Ponomarenko-type flow (Núñez *et al.*, 2001). These calculations share a common feature: they predict a saturated magnetic energy density above threshold given by an equation of the form

$$\langle B^2 \rangle \propto \frac{\rho v \mu_0 \eta}{L^2} (R_m - R_{mc}). \quad (3)$$

We call (3) the “laminar scaling”, characterized by the fact that $B \rightarrow 0$ if $v \rightarrow 0$ with all the other parameters fixed. This shows that the dynamo bifurcation is supercritical, in agreement with the observations of Karlsruhe and Riga experiments. However, the predicted mean magnetic field amplitude is roughly 1000 times smaller than the measured ones, of order 10 mT roughly 10% above threshold. Both experiments used liquid sodium ($\eta \approx 0.1 \text{ m}^2/\text{s}$, $\rho \approx 10^3 \text{ kg m}^{-3}$). The inner diameter of the Riga experiment is $L = 0.25 \text{ m}$. The spatial periodicity of the flow used in the Karlsruhe experiment is of the same order of magnitude, within a cylinder of radius 0.85 m and height 0.7 m. The laminar scaling (3) thus predicts a saturated mean field of order $10 \mu\text{T}$ (and even an order of magnitude smaller in the Karlsruhe experiment if the full size of the experiment is taken for L).

It is possible to understand why the laminar scaling has been obtained in most analytical calculations and which is the appropriate one when the dynamo bifurcation occurs with a large Reynolds number flow (Pétrélis and Fauve, 2001). We recall the induction and Navier-Stokes

equations that we restrict to incompressible flows ($\nabla \cdot \mathbf{v} = 0$),

$$\frac{\partial \mathbf{B}}{\partial t} = \nabla \times (\mathbf{v} \times \mathbf{B}) + \eta \nabla^2 \mathbf{B}, \quad (4)$$

$$\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla) \mathbf{v} = -\nabla \left(\frac{p}{\rho} + \frac{B^2}{2\mu_0} \right) + \nu \nabla^2 \mathbf{v} + \frac{1}{\mu_0 \rho} (\mathbf{B} \cdot \nabla) \mathbf{B}. \quad (5)$$

The flow is created, either by moving solid boundaries or by a body force added to the Navier-Stokes equation. We have to develop (4, 5) close to the dynamo threshold in order to derive an amplitude equation for the growing magnetic field. In general, this calculation is tractable only in the unrealistic case $Pr_m \gg 1$ such that the dynamo bifurcates from a laminar flow. For $Pr_m \ll 1$, a lot of hydrodynamic bifurcations occur first and the flow becomes turbulent before the dynamo threshold. The structure of the weakly nonlinear analysis above threshold is as follows: the dynamo bifurcates from a flow field \mathbf{v}_c at $R_m = R_{mc}$. We write (4) in the form

$$L(\mathbf{B}_0) = \mathbf{0}, \quad (6)$$

where \mathbf{B}_0 is the neutral mode at threshold and L is a linear operator that depends on the bifurcation structure (stationary or Hopf bifurcation). Slightly above threshold, we have

$$\mathbf{v} = \mathbf{v}_f + \epsilon \mathbf{v}_1 + \dots, \quad (7)$$

where $\mathbf{v}_f = \mathbf{v}_c + \epsilon \mathbf{v}_p + \dots$, with $\epsilon = (R_m - R_{mc})/R_{mc} \ll 1$; $\epsilon \mathbf{v}_p$ is the order ϵ velocity correction due to the driving of the fluid slightly above threshold; $\epsilon \mathbf{v}_1$ is the leading order flow distortion by the Lorentz force. We have for \mathbf{B} that

$$\mathbf{B} = \sqrt{\epsilon} (\mathbf{B}_0 + \epsilon \mathbf{B}_1 + \dots). \quad (8)$$

We first compute \mathbf{v}_1 from (5) at order ϵ :

$$\frac{\partial \mathbf{v}_1}{\partial t} + (\mathbf{v}_c \cdot \nabla) \mathbf{v}_1 + (\mathbf{v}_1 \cdot \nabla) \mathbf{v}_c = -\frac{1}{\rho} \nabla \left(p_1 + \frac{B_0^2}{2\mu_0} \right) + \nu \nabla^2 \mathbf{v}_1 + \frac{1}{\mu_0 \rho} (\mathbf{B}_0 \cdot \nabla) \mathbf{B}_0. \quad (9)$$

If $Pr_m \gg 1$, the flow is laminar at the dynamo threshold, and the Lorentz force is mostly balanced by the modification of the viscous force, thus

$$v_1 \propto \frac{B_0^2 L}{\mu_0 \rho \nu} \quad (v_1 = |\mathbf{v}_1|). \quad (10)$$

We next have to relate v_1 to the distance from threshold. For a supercritical bifurcation, we expect saturation when the flow distortion, \mathbf{v}_1 , due to the Lorentz force becomes comparable to the one that results from supercritical forcing, \mathbf{v}_p . This is precisely the relation we get from the solvability condition for (4) at order ϵ (Pétrélis and Fauve, 2001). Using $\mathbf{v}_p \propto \mathbf{v}_1$ and $|L(\mathbf{v}_f - \mathbf{v}_c)/\eta| \propto R_m - R_{mc}$ in the vicinity of threshold, we obtain the laminar scaling (3).

Although analytical calculations are not possible at high Re , we can guess the correct scaling in that case by following the same steps as above.

The only difference is that if $Re \gg 1$, we have to balance the Lorentz force with the inertial term instead of the viscous one in (9). This gives

$$v_1 \propto \frac{B_0^2}{\mu_0 \rho v_c} \quad (v_c = |\mathbf{v}_c|). \quad (11)$$

Thus, we obtain for the high Re scaling,

$$\langle B^2 \rangle \propto \frac{\rho \mu_0 \eta^2}{L^2} R_{mc} (R_m - R_{mc}). \quad (12)$$

It differs from the laminar scaling by the factor $R_{mc} Pr_m^{-1/2}$ and is in much better agreement with the experimental measurements (we recall that for liquid metals, $Pr_m < 10^{-5}$).

The reaction of the magnetic field on the flow is characterized by the ratio of the Lorentz force to the characteristic pressure forces driving the flow. This is measured by the interaction parameter, $N = B^2 L / \rho \mu_0 \eta V$. We observe that the high Re scaling corresponds to $N \propto R_m - R_{mc}$, whereas for the low Re scaling, $Q \propto R_m - R_{mc}$, where Q is the Chandrasekhar number ($Q = N Re$).

We note that if we replace ν by the turbulent viscosity, $\nu_t \propto V L$, in the laminar scaling (3) and then use $V \approx \eta R_{mc} / L$ in the vicinity of threshold, we recover the high Re scaling. However, its direct derivation is clearer. It would be interesting to check this scaling in Karlsruhe and Riga experiments by varying the working temperature and therefore η .

In Fig. 7 we directly test (12) using the Riga and Karlsruhe magnetic field data. The envelope of the magnetic field oscillations for Riga, together with the temperature at the time were supplied by Frank Stefani. The temperature changes during the experimental data taking in Riga, necessitating careful treatment of the data. An estimated value of the magnetic diffusivity at each temperature was used in the figure scaling. This is important in order to correctly nondimensionalize the magnetic field values, in addition to the same treatment for the velocity nondimensionalization in the magnetic Reynolds number.

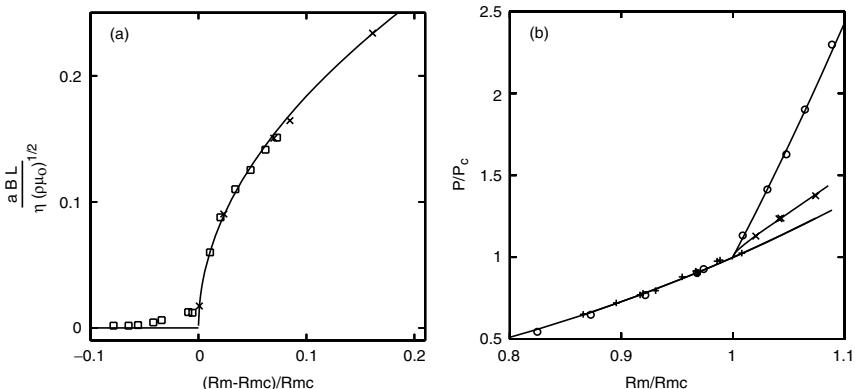


Figure 7. (a) The magnetic field amplitude from both the Karlsruhe (boxes) and Riga (x) experiments can be collapsed onto a single diagram by making the parameters dimensionless, and by fixing a single parameter “a” which characterizes the amplitude of the square root rise in field amplitude B (the envelope of oscillations for Riga, magnitude for Karlsruhe). Here R_{mc} is the onset magnetic Reynolds number for dynamo action, L is the system diameter, η the magnetic diffusivity, ρ the density, and μ the magnetic permeability. See text for more discussion of the data (data from Stieglitz and Müller, 2002, and Frank Stefani (Gailitis *et al.*, 2002)). (b) Power dissipation P for both the Riga (x) and Karlsruhe (○) experiments, scaled by the power at onset of dynamo action P_c , show Kolmogorov dependence below the transition to self-generation, with different deviations above the threshold—perhaps due to the different nature of the saturation mechanisms.

Both the Riga and Karlsruhe magnetic field amplitudes show a square root rise, as (12) predicts. A single parameter, a , allows both data sets to be represented in a single figure. A similar attempt to collapse the power dissipation data for the two experiments succeeds only below the onset of dynamo action. The Karlsruhe power dissipation, in these dimensionless coordinates, is substantially larger than the corresponding measurements for Riga. This may be a sign of the system having difficulty imposing Lorentz forces suitable for saturation in the scale-separated Karlsruhe geometry.

3. Generating a fully three-dimensional turbulent flow in a compact volume

The nearly two-dimensional nature of the large-scale flows of the Karlsruhe and Riga experiments and their relatively low level of turbulent fluctuations probably make these dynamos rather peculiar. This

motivates the study of dynamos generated by flows with less geometrical constraints. Several liquid sodium flows in domains of aspect ratio nearly one have been studied. As mentioned above, early experiments have been conducted by Lehnert (1957). A swirling flow has been generated by rotating a disk in a cylindrical vessel containing 58 l of liquid sodium. As mentioned by Roberts and Jensen (1993), this flow shares some geometrical properties of the Ponomarenko flow: it involves both poloidal and toroidal components with a non-zero mean helicity. It is, however, likely that the three-dimensional nature of the flow when the cylinder has an aspect ratio of order one strongly affects the nature of the dynamo bifurcation compared to the case of a very elongated cylinder. Indeed, kinematic dynamos generated by flows in a sphere have been studied by Dudley and James (1989). They have been classified according to the topologies of their poloidal and toroidal components: $s1 + t1$ stands for one poloidal eddy ($s1$) and one toroidal eddy ($t1$). This is the shape of the flow generated by a propellor in a sphere and has a much larger dynamo threshold than a Ponomarenko-type flow in a finite but elongated cylinder. In addition, the dynamo threshold of these compact flows strongly depends on geometry: minor modifications of the velocity field (without altering its topology) or modifications of the boundary conditions, can suppress the dynamo capability of these flows. Two experiments have been performed using two co-axial propellers (Peffley *et al.*, 2000) or impellers (Bourgoin *et al.*, 2002), counter-rotating in sphere or cylinder of aspect ratio one filled with liquid sodium, and another one is in preparation (Forest *et al.*, 2002). They correspond to $s2 + t2$ flows in the above classification, i.e., two poloidal eddies with the flow directed inward in the mid-plane between the propellers together with two counter-rotating toroidal eddies. Hydrodynamic properties of these flows have been widely studied in the context of von Kármán swirling flows (e.g., Zandbergen and Dijkstra, 1987), i.e., flows generated by rotating two co-axial disks (possibly fitted with blades to increase the driving efficiency). In the counter-rotating case, an $s2 + t2$ type mean flow is generated but the strong shear in the mid-plane due to counter-rotating toroidal eddies generates very large turbulent fluctuations (nearly up to 50% in appropriate geometries). This may be considered as a disadvantage if one is interested by a dynamo as close as possible to known analytical examples, or is an advantage if the motivation is to answer open questions about turbulent dynamos. The rms fluctuations can be strongly decreased by driving the disks in co-rotation (Fauve *et al.*, 1993), thus generating a $s2 + t1$ type flow. It should be noted, however, that the most intense part of this flow is a strong axial vortex

with a diameter small compared to its length (Labbé *et al.*, 1996). The sign of the axial velocity changing on the mid-plane between the disks, the mean flow helicity vanishes. The type of mechanical flow driving described above thus allows to explore various mean flow configurations with different symmetries and different levels of turbulent fluctuations.

3.1. Maryland 30 cm experiment

A sequence of experiments have been conducted at the University of Maryland in liquid sodium, which are of increasing size and drive power. A number of observations have been made in the 30 cm diameter apparatus shown in Fig. 8. Flows are driven in the spherical container by propellers which drive a co- or counter-rotating flow in the two hemispheres. Pumping also occurs largely due to radial outflow from the propellers which return via the poles and equator.

The flow driven by the propellers is highly turbulent ($Re < 7.8 \times 10^6$), but of modest magnetic Reynolds number ($R_m < 65$) based on tip speed of the propellers and the sphere radius. Significant changes

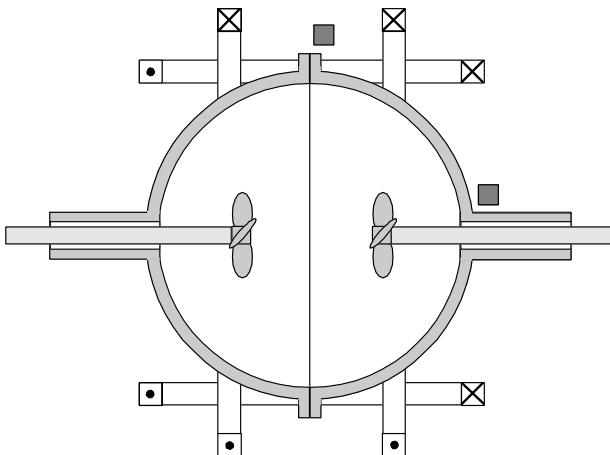


Figure 8. Sketch of the Maryland 30 cm experiment. Counter-rotating propellers drive a flow in sodium within a stationary outer sphere. The propellers can be driven in different axial positions. A number of different propellers have been tested (see Shew *et al.*, 2001). Induced fields are measured at several locations near the shaft and equator using Hall probes. An external magnetic field is applied either transversally or along the axis of the sphere.

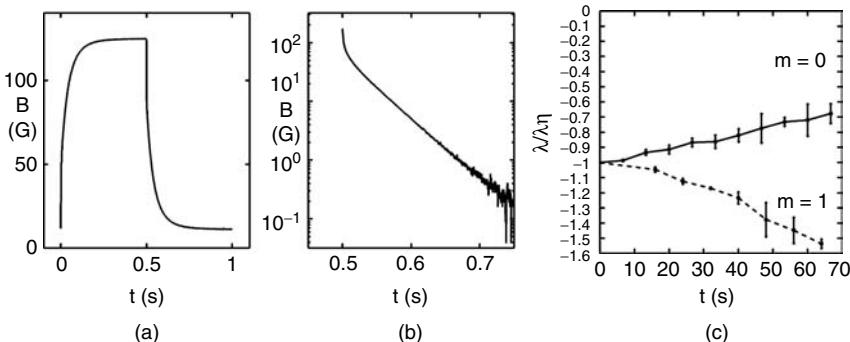


Figure 9. (a) A small external magnetic field is applied by external pulse coils. (b) Rapidly cutting the current to the coils allows a detailed examination of the decaying induced field from the sodium flow. (c) The decay rate (λ) is sensitive to the magnetic Reynolds number and the direction of the applied field ($m = 0$ indicates aligned with the rotation axis, $m = 1$ is for perpendicular to the axis). The data have been normalized to the resistive decay such that for no flow $\lambda/\lambda_\eta = -1$. The change in the decay rate may indicate an approach to self-generation (from Peffley *et al.*, 2000).

in the decay rate of magnetic field have been observed (Fig. 9), but these rates fluctuate in time and depend on changes in geometry such as: baffles, propellor shape and propellor position (Peffley *et al.*, 2000; Shew *et al.*, 2001). Interpreting these changes in decay rate has proven controversial (Tilgner, 2002); an optimistic interpretation suggests a trend toward self-generation.

The dynamics of the induced magnetic field changes significantly when the external field is increased to the point when Lorentz forces become significant in the flow. We have observed a number of different regimes (Sisan *et al.*, 2003) as the interaction parameter is increased (Fig. 10) for the case of the two impellers co-rotating. For small imposed field, the induced magnetic field acts as a passive scalar, being advected and intermittently stretched by the flow. As the external field is increased, the turbulence amplitude is observed to decrease, followed by a rise of coherent oscillations whose frequency changes with applied field strength. Finally, the oscillations, at the highest applied field, lock their main frequency to a sub-harmonic of the impeller rotation rate, and develop modulations. In this final state, the experiment shows system-wide coherence.

These experiments are in the context of several other devices: a smaller 20 cm diameter rotating convection experiment, a 60 cm rotating convection experiment, and construction of a 3 m diameter experiment geometrically similar to the 30 cm device.

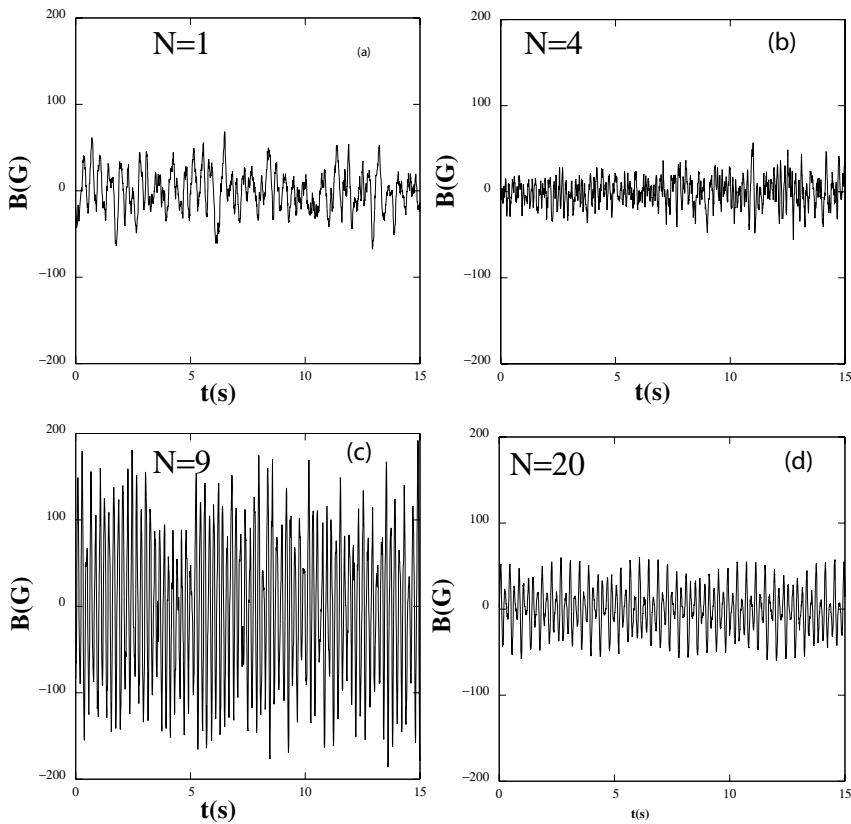


Figure 10. The application of a larger external field can significantly affect the dynamics of the induced field (and presumably the underlying sodium flow). As one increases the relative importance of Lorentz forces to inertia (measured by the interaction parameter $N = B_0^2 L / \rho \mu_0 \eta V$), one sees the induced field act (a) as a passive vector field, (b) with a suppression of turbulence, (c) with large oscillations whose frequency changes with applied field strength and (d) with oscillations locked to a subharmonic of the propellor rotation rate (from Sisan *et al.*, 2003).

3.2. VKS experiment

von Kármán swirling flows

A von Kármán swirling flow of 54 l of liquid sodium (VKS) is driven by rotating two disks in a cylindrical vessel, 410 mm in inner diameter and 400 mm in length (see Fig. 11). The experimental set-up has been described elsewhere (Bourgoin *et al.*, 2002). The radius of the

disks as well as the number and shape of straight or curved blades, have been varied in order to increase the driving efficiency. In the magnetic induction experiments presented here, we use disks of radius $R = 150$ mm, fitted with 8 straight blades of height $h = 10$ mm driven at a rotation frequency up to $f = 30$ Hz. Four baffles, 20 mm in height, have been mounted on the cylindrical vessel inner wall, parallel to its axis, in order to increase the poloidal flow component compared to the toroidal one. A turbulent swirling flow with an integral Reynolds number, $Re = 2\pi R^2 f / \nu$, up to 4×10^6 has been thus driven.

Two coils generate a nearly homogeneous magnetic field \mathbf{B}_0 , parallel or perpendicular to the cylinder axis (see Fig. 11). The three components of the field induced by the flow are measured with a three-dimensional Hall probe, located in the mid-plane between the disks, at an adjustable distance z from the axis of rotation.

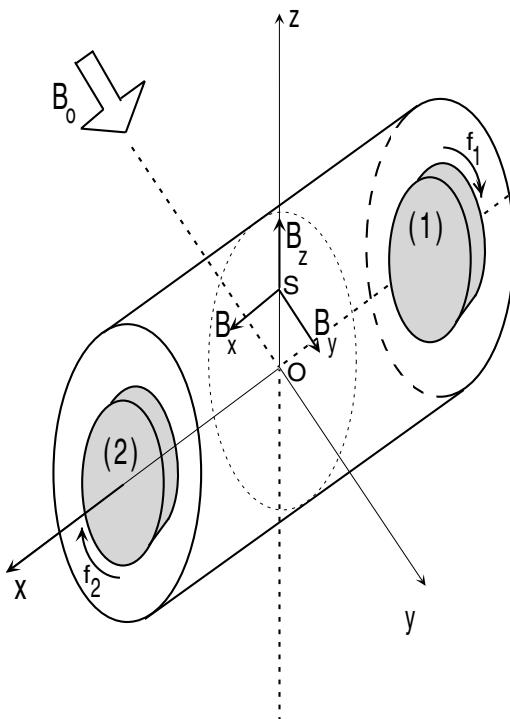


Figure 11. Sketch of the VKS experiment. The external magnetic field is applied either transversally (as shown in the figure) or along the cylinder axis.

Magnetic induction

The equations governing the magnetic field, $\mathbf{B}(\mathbf{r}, t)$, induced by the flow submitted to the field \mathbf{B}_0 , are in the MHD approximation

$$\nabla \cdot \mathbf{B} = 0, \quad (13)$$

$$\frac{\partial \mathbf{B}}{\partial t} = \nabla \times [\mathbf{V} \times (\mathbf{B} + \mathbf{B}_0)] + \eta \nabla^2 \mathbf{B}, \quad (14)$$

where $\mathbf{V}(\mathbf{r}, t)$ is the velocity field.

The maximum field amplitude being $B_0 = 12$ G, the interaction parameter, $N = B_0^2 R / \rho \mu_0 \eta V_M$, which characterizes the ratio of the Lorentz force to inertial forces, is in the range $10^{-4} < N < 10^{-2}$. Thus the effect of the magnetic field on the flow is negligible. This has been checked directly by measuring \mathbf{B} as a function of \mathbf{B}_0 at a constant driving of the flow. The mean induced field $\langle \mathbf{B} \rangle$ where $\langle \cdot \rangle$ stands for average in time, as well as its rms fluctuations in time, \mathbf{B}_{rms} , vary linearly with B_0 , thus showing that the modification of the velocity field \mathbf{V} in (14) can be neglected. Thus, the only relevant dimensionless parameters of our experiments are the magnetic Prandtl number, $Pr_m = \nu / \eta \approx 10^{-5}$, and the magnetic Reynolds number, $R_m = 2\pi R^2 f / \eta$, which is proportional to the rotation frequency f and has been varied up to 40 for radius of the disks $R = 150$ mm.

The three components of the mean magnetic field $\mathbf{B}_0 + \langle \mathbf{B}(\mathbf{r}) \rangle$, at $z = 100$ mm above the rotation axis, are displayed as a function of the rotation frequency in Fig. 12 (\mathbf{B}_0 axial) and in Fig. 12 (\mathbf{B}_0 transverse).

We first observe that there is no induced field B_z . Second, when the rotation of the disks is reversed, $f \rightarrow -f$, the induced field parallel to \mathbf{B}_0 is unchanged whereas the other horizontal component changes sign. Third, the induced field is proportional to f , thus to R_m , on a larger range for \mathbf{B}_0 axial than for \mathbf{B}_0 transverse.

Assuming that the swirling flow has not broken the symmetries of the driving configuration, the first two observations can be understood using the following symmetry transformations:

- (i) the rotation of angle π about the z -axis leaves the flow unchanged but leads to the following transformation of the field components: $(\langle B_x \rangle, \langle B_y \rangle, \langle B_z \rangle) \rightarrow (-\langle B_x \rangle, -\langle B_y \rangle, \langle B_z \rangle)$. \mathbf{B}_0 changes sign in both cases (axial or transverse). The transformation $\mathbf{B}_0 \rightarrow -\mathbf{B}_0$ thus leads to the initial situation with $(\langle B_x \rangle, \langle B_y \rangle, \langle B_z \rangle) \rightarrow (\langle B_x \rangle, \langle B_y \rangle, -\langle B_z \rangle)$. Thus, $B_z = 0$.

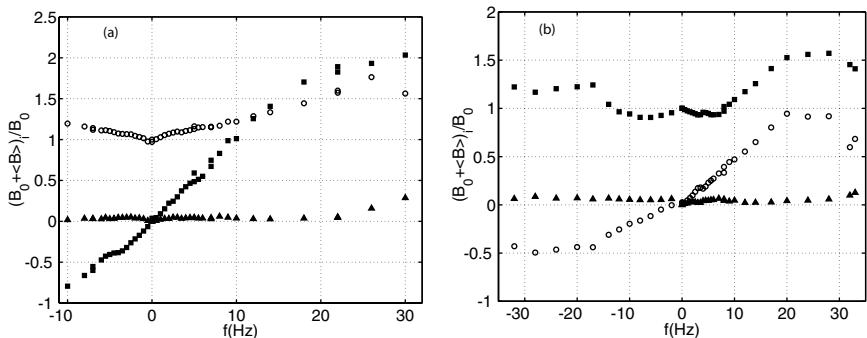


Figure 12. (a) Components of the total mean magnetic field as a function of the rotation frequency with counter-rotating disks and a field \mathbf{B}_0 applied along the cylinder axis. The disk radius is $R = 150$ mm with straight blades. Four baffles are mounted on the inner wall of the cylindrical vessel. The magnetic field is measured at $z = 100$ mm. [(o) = $B_o + \langle B_x \rangle / B_o$, (squares) = $\langle B_y \rangle / B_o$, (Δ) = $\langle B_z \rangle / B_o$]. (b) Components of the total mean magnetic field as a function of the rotation frequency with counter-rotating disks and a field \mathbf{B}_0 applied transversally. The disk radius is $R = 150$ mm with straight blades. Four baffles are mounted on the inner wall of the cylindrical vessel. The magnetic field is measured at $z = 100$ mm. [(o) = $\langle B_x \rangle / B_o$, (squares) = $B_o + \langle B_y \rangle / B_o$, (Δ) = $\langle B_z \rangle / B_o$].

- (ii) the symmetry with respect to the vertical plane, $x0z$, changes the sign of the rotation vector (it is a pseudovector), thus $f \rightarrow -f$. If \mathbf{B}_0 is axial, its sign is also changed and one should perform the additional transformations $\mathbf{B}_0 \rightarrow -\mathbf{B}_0$ and $\mathbf{B} \rightarrow -\mathbf{B}$ in order to recover the initial situation with opposite rotation rates. We thus obtain that $f \rightarrow -f$ implies $(\langle B_x \rangle, \langle B_y \rangle, \langle B_z \rangle) \rightarrow (\langle B_x \rangle, -\langle B_y \rangle, -\langle B_z \rangle)$. For \mathbf{B}_0 transverse, the symmetry with respect to $x0z$ gives, $f \rightarrow -f$ and $(\langle B_x \rangle, \langle B_y \rangle, \langle B_z \rangle) \rightarrow (-\langle B_x \rangle, \langle B_y \rangle, -\langle B_z \rangle)$.

When \mathbf{B}_0 is axial, it is more appropriate to consider cylindrical coordinates. Figure 12 together with rotational invariance about the cylinder axis, show that there is no induced field in the radial direction in the mid-plane between the disks. The largest induced field component is the azimuthal one and corresponds to the ω -effect, first observed in a one-disk configuration by Lehnert (1957).

When \mathbf{B}_0 is transverse, a large axial field component is induced. A rotation of angle π about the cylinder axis, followed by the transformation $\mathbf{B}_0 \rightarrow -\mathbf{B}_0$ shows that $B_x(0, 0, -z) = -B_x(0, 0, z)$. Thus B_x is generated by a current density parallel to \mathbf{B}_0 . In addition, as said above, the sign of B_x depends on the flow helicity. Although we have no scale separation here in order to consider the generation of B_x as an α -effect,

a simple way to understand it relies on the scheme proposed by Parker (1955) of magnetic field induced by cyclonic eddies. It has been also shown by Bourgoin (2003) that electrical boundary conditions may add a contribution to the generation of B_x . The behavior of B_y as a function of the rotation frequency is more complex and strongly nonlinear at rather low R_m compared to the one of B_x . $B_0 + B_y$ first decreases at low R_m . This can be understood as transverse magnetic field expulsion from rotating eddies. This effect is stronger in the one-disk configuration without baffles (Pétrélis *et al.*, 2003) and has been studied in details in a von Kármán flow in gallium (Odier *et al.*, 2000).

Magnetic fluctuations

Direct time recordings of the fluctuations of the components of the magnetic field are displayed in Fig. 13 for an axially applied magnetic field $B_0 = 12\text{G}$. *rms* fluctuations to mean induced field may be as large as 50%. At low R_m , the fluctuating magnetic field is mainly generated from the interaction of the fluctuating velocity field with \mathbf{B}_0 , thus the fluctuation to mean ratio of the induced magnetic field is of the same order of magnitude as the fluctuation to mean ratio of the velocity field. Magnetic field spectra display an inertial range scaling above the disk rotation frequency with a power law fall-off close to $f^{-11/3}$. This corresponds to a Kolmogorov type scaling predicted by Golitsyn (1960) and Moffatt (1961) provided that Taylor hypothesis could be used. A $1/f$ type spectrum is observed at low frequency (Bourgoin *et al.*, 2002). Although a k^{-1} spectrum has been predicted for the galactic magnetic fields (Ruzmaikin and Shukurov, 1982), it is unlikely to be able to use Taylor hypothesis at low frequencies and another explanation should be found (Marié *et al.*, 2003b).

Effect of turbulence and sensitivity of dynamo threshold

Writing $\mathbf{B}(\mathbf{r}, t) = \langle \mathbf{B}(\mathbf{r}) \rangle + \mathbf{b}(\mathbf{r}, t)$, where \mathbf{b} is the fluctuating part of the magnetic field, and similarly for \mathbf{V} , we deduce from (14) the mean induced field equation

$$-\frac{1}{\mu_0\sigma}\nabla^2\langle\mathbf{B}\rangle = \nabla \times (\langle\mathbf{V}\rangle \times \mathbf{B}_0 + \langle\mathbf{V}\rangle \times \langle\mathbf{B}\rangle + \langle\mathbf{v} \times \mathbf{b}\rangle). \quad (15)$$

When the magnetic Reynolds number is small, the first source term on the right-hand side of (15) is the dominant one and we usually get the i^{th} component of the mean induced field $\langle B_i \rangle \propto R_m B_0$. However, we emphasize that when \mathbf{B}_0 is transverse, $\langle B_y \rangle$ behaves nonlinearly at low

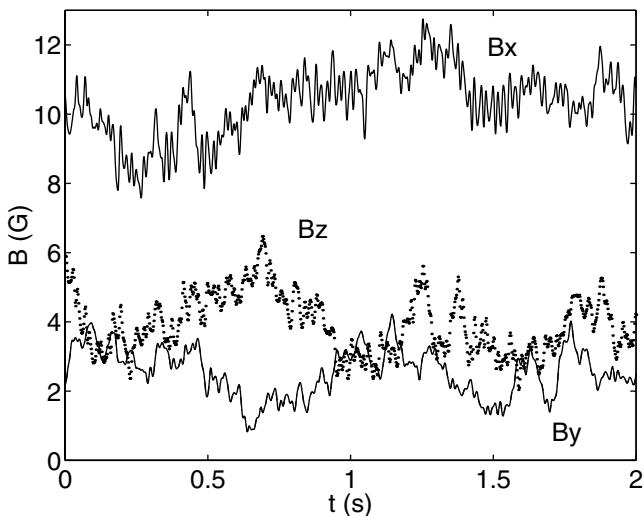


Figure 13. Time recording of the fluctuations of the magnetic field with counter-rotating disks at frequency 8 Hz and a field $B_0 = 12$ G applied axially. The magnetic field is measured at $z = 100$ mm.

R_m (at least at the location of the measurements of Fig. 12). We have thus probed the nonlinear source terms of (15). The relative contribution of the two nonlinear source terms $\nabla \times (\langle \mathbf{V} \rangle \times \langle \mathbf{B} \rangle)$ and $\nabla \times (\langle \mathbf{v} \times \mathbf{b} \rangle)$ depends on the relative level of turbulent fluctuations \mathbf{v} and \mathbf{b} , and also on their correlation. Both the velocity field and the magnetic field display large fluctuations (roughly 50% with counter-rotating disks). An estimate of the effect of $\langle \mathbf{v} \times \mathbf{b} \rangle$ can be obtained from the difference between the measured magnetic field and a numerical computation of the magnetic field induced by the mean flow alone measured in a similar water experiment. This comparison has been performed in a less turbulent one-disk configuration (Pétrélis *et al.*, 2003) and should be done in the counter-rotating geometry.

To wit, a 1/2-scale water experiment has been built and the mean velocity field $\langle \mathbf{V}(\mathbf{r}) \rangle$ has been measured using laser anemometry. A kinematic dynamo code has then been used, either to compute the magnetic field induced by the mean flow, or to find the dynamo threshold corresponding to the mean flow alone (Marié *et al.*, 2003a). These computations have confirmed the mean induction measurements of the VKS experiment. They have also shown that despite local field amplification, the maximum R_m presently reached in the VKS experiment is below the dynamo threshold computed as if the mean flow were acting alone. It is

expected to drive the flow above this threshold in the next experimental run.

In addition, these numerical simulations have displayed a rather surprising behavior. Slightly different shapes of the blades of the propellers, that lead only to minor modifications of the measured mean flow, may have huge effects on the dynamo threshold. When R_m is increased, the dynamo growth rate continuously increases and becomes positive with one type of propellers whereas it goes up to a maximum and then decreases before reaching zero with slightly different propellers. Thus, the dynamo capability of this flow is very sensitive to minor modifications. It is not clear whether this singular behavior subsists in the presence of turbulence or if it results from the fact that fluctuations have been removed from the calculation. In any case, it would be of great theoretical interest to understand the reason of this extreme sensitivity of fully three-dimensional laminar dynamos. Another practically more important question is of course to understand the effect of turbulent fluctuations on these laminar dynamo thresholds.

4. Motivations for rapidly rotating flows

Although the first experiments which showed self-generating fields did not have system-wide rotation, there is significant interest in understanding the effect of (planetary) rotation on field generation. Indeed, there is a widespread belief that global rotation may assist generation by lowering the threshold Reynolds number.

Global rotation, especially when the Rossby number is small ($Ro = U/2\Omega L \ll 1$, where U is the typical magnitude of the velocity field in the rotating reference frame), causes the flow to become quasi-two-dimensional. This is effected through the action of Coriolis accelerations. It does not take too much imagination to see the elongation present in both the Riga and Karlsruhe experiments, and to wonder if their baffles have recreated in the flow patterns comparable to what rotation might cause in planetary bodies. Rotation, by breaking the parity symmetry in the governing equations can cause the flow to generate global helicity — helpful for self-generation.

There is however no clear-cut demonstration or experimental evidence that rotation increases the dynamo capability of a flow. The exact meaning of this sentence itself is not clear. Does it mean that one expects a lower critical magnetic Reynolds number for a rotating flow? If yes, how should R_m be defined with and without rotation since it is known that rotation generates new lengthscales? A better defined criterion is to evaluate the power required to reach the dynamo threshold

with and without rotation. Other questions about the effect of rotation concerns the dynamic dynamo problem: how does it affect the saturated magnetic energy above the dynamo threshold? Does it modify the nature of the dynamo bifurcation? Before discussing these problems, we will first consider several MHD experiments in rapidly rotating flows motivated by astrophysical or geophysical situations.

4.1. Astrophysical and geophysical dynamos

Due to these motivations, and the general motivation to better understand the effects of rotation on flows in nature, a number of groups have undertaken experiments which incorporate global rotation (Fig. 14; see [color insert](#) following page 234).

Researchers in Grenoble, France, have undertaken a number of rotating convection experiments, and are in construction for a differentially rotating sodium experiment (Cardin *et al.*, 2002). Their overall goal is to understand the dynamics of rotating geophysical flows, and states where Coriolis and Lorentz forces are dominant.

Another project is underway in Socorro, New Mexico, USA, where axial jets are driven in a rotating flow, to study the effect of helicity generation in accretion disks. Preliminary experiments have been conducted in water (Colgate *et al.*, 2002), while a scaled-up sodium experiment has been fabricated.

A rotating convection sodium experiment has been fabricated and is undergoing initial experiments in Maryland, USA. This experiment is a concentric spherical device, with a heated outer 60 cm sphere and a cooled 20 cm inner sphere. The goal of the project is to understand the effects of Coriolis and Lorentz forces on rotating convection, and the effects of convection on induced magnetic fields.

Precession of a rotating volume of fluid generates tidal forces that have been considered as a possible source of power driving planetary dynamos (Malkus, 1968). Gans (1970) performed experiments with a cylinder filled with liquid sodium. The rotation rate was increased up to 3600 rpm for a precession rate of 50 rpm. An external magnetic field amplification has been reported but no self-generation. Another precession driven dynamo experiment is planned (Léorat *et al.*, 2001).

Finally, we mention the idea of a transient dynamo experiment using a Ponomarenko type flow generated in a torus when its rotation is abruptly stopped (Frick *et al.*, 2001). A fixed divertor inside the torus forces a helical motion of the inertially continuing flow. A Ponomarenko type flow is thus generated in a torus and one may expect to observe a transient dynamo.

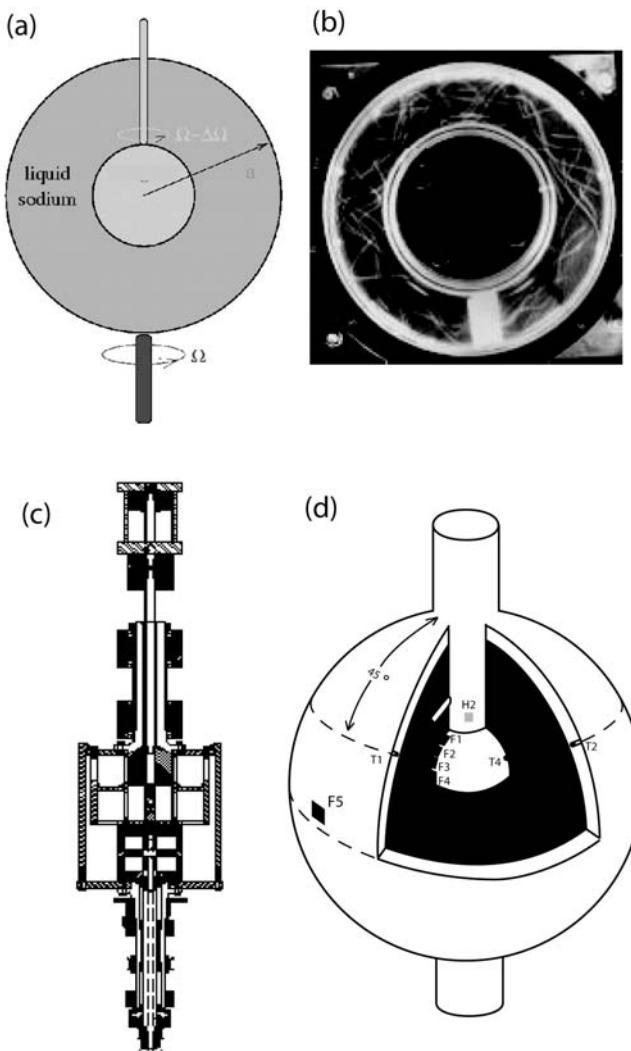


Figure 14. (a) Cross section of the Grenoble sodium experiment, which has differentially rotating inner and outer spheres. The inner sphere is a permanent magnet (see Cardin, *et al.*, 2002). (b) A sodium experiment is planned in Perm, where a rotating torus of liquid is rapidly decelerated, leading to a helical motion of the liquid within the torus (from Frick *et al.*, 2002). (c) In New Mexico, a rotating Couette flow is planned with axial jets of liquid to also stimulate helical flows in sodium (from Colgate *et al.*, 2002). (d) A Maryland experiment with a 60 cm rotating convection experiment, where centrifugal forces drive convection between a heated outer sphere and a cooled inner sphere (see Shew *et al.*, 2002). (**See color insert.**)

4.2. Dynamo bifurcations in rapidly rotating flows

Rotation imposes a strong constraint on the flow that tends to become nearly two-dimensional. The lengthscale l of the flow in a direction perpendicular to Ω is thus much smaller than the one along Ω , $l \ll L$. When convection is generated in a rotating sphere, the flow concentrates in columns of diameter $l \propto LE^{1/3}$ where $E = v/\Omega L^2$ is the Ekman number (Roberts, 1968; Busse, 1970). This type of flow can generate a large scale magnetic field on lengthscale L via an α -effect (Busse, 1975). Plane layer models (Childress and Soward, 1972; Soward, 1974) display most of the important features of spherical geometries (Roberts, 1988). Increasing the rotation rate too much delays the linear instability onset of self-generation because more and more power is necessary to overcome dissipation at small scale l . However, for finite amplitude magnetic fields, the Lorentz force suppresses the rotational constraint and allows large scale motions leading to much smaller viscous and ohmic dissipation. A subcritical “strong field” branch thus exists below the linear stability onset (St. Pierre, 1993). The strong field scaling corresponds to a balance between the Coriolis and the Lorentz forces, thus $B_{\text{strong}}^2 \propto \rho\Omega/\sigma$. A weak field regime may be stable above the linear threshold but it becomes unstable for an order one Chandrasekhar number. It corresponds to the low Reynolds number scaling (3), thus $B_{\text{weak}}^2 \propto \rho v \mu_0 \eta / L^2$. A very interesting feature of dynamos generated by rapidly rotating flows is thus the subcritical nature of the bifurcation. Consequently, the questions related to the effect of rotation on the linear dynamo threshold are of secondary importance. The mean magnetic energy of finite amplitude dynamo solutions deserves more attention and is strongly affected by rotation.

4.3. Scalings of magnetic energy

The weak field branch has been computed analytically with different models (Childress and Soward, 1972; Soward, 1974; Busse, 1975). These computations assume the flow laminar with a simple geometry whereas astrophysical or geophysical flows are turbulent. We have already emphasized in Section 2 that this strongly affects the mean magnetic energy. However, both the low (3) and high Reynolds number scaling (12) give orders of magnitude too small values for the magnetic field of the Earth or of the Sun. For the Earth, taking $\rho \approx 10^4 \text{ kg m}^{-3}$, $\sigma \approx 3 \cdot 10^5 \text{ S m}^{-1}$ and $L \approx 3 \cdot 10^6 \text{ m}$, gives $B_{\text{weak}} \approx 0.5 \mu\text{G}$. Moreover, even if we replace L by $l \propto LE^{1/3}$, thus changing the lengthscale and correspondingly the field by a factor 10^5 , does not lead to a large enough correction!

One way to solve the problem is to find a scaling law which does not involve any length-scale at all. We may use the same type of arguments as the ones of Section 2. The only difference with the low and high Reynolds number scalings found there, is that there is the additional Coriolis term, $2\Omega \times \mathbf{v}_1$, in (9). If it is the dominant one, we get

$$v_1 \propto \frac{B_0^2}{\mu_0 \rho \Omega L}. \quad (16)$$

This gives

$$\langle B^2 \rangle \propto \rho \Omega \mu_0 \eta (R_m - R_{mc}). \quad (17)$$

Note that we have obtained the strong field scaling, but without assuming that there is a balance between the Lorentz force and the total Coriolis force. Only the Coriolis force related to the velocity perturbation balances the Lorentz force; this gives the additional term $R_m - R_{mc}$ in the expression of the mean magnetic energy. Although this scaling law looks fine ($\sqrt{\rho \Omega \mu_0 \eta} \approx 10$ G), it is not obvious that a perturbative analysis can be worked out that way and we should be cautious in the absence of an explicit analytical example leading to the strong field scaling.

5. Concluding remarks

Decades after the discovery of the first analytical examples of simple laminar fluid dynamos, self-generation has been observed by the Karlsruhe and Riga groups. These experiments have shown that, in geometrically well-controlled flows, small scale turbulent fluctuations have little effect on the nature of the dynamo bifurcation and its threshold value. However, the large Reynolds number of these flows should be taken into account in order to predict the correct order of magnitude of the mean magnetic energy generated by the dynamo process.

Both the Karlsruhe and Riga experiments were designed after a careful optimization of the dynamo capability of the flow, maintained in a nearly two-dimensional configuration. This geometrical constraint provides a rather robust value of the threshold, thus easily predictable using an approximation of the real flow. On the contrary, in a less constrained fully three-dimensional flow, slight modifications may strongly affect the threshold. This is unpleasant for the experimentalist but gives new degrees of freedom and makes the dynamo bifurcation problem richer: indeed, above linear threshold, the Lorentz force may drive the flow toward a geometrical configuration with a larger dynamo

capability. This would generate a subcritical branch of finite amplitude solutions for the magnetic field. It is strongly believed that such a mechanism operates for the geodynamo because the Lorentz force destroys the geostrophic nature of rapidly rotating flows. It would be of great interest to observe a subcritical dynamo bifurcation experimentally and in particular to find whether it is possible to reach a regime in which most dissipation is of ohmic nature. This is believed to be the regime of the geodynamo whereas the Karlsruhe and Riga experiments have displayed dominant viscous dissipation.

Finally, it would be of great interest to observe the dynamical behavior of the magnetic field generated by a strongly turbulent flow. The fluid velocity being involved multiplicatively in the induction equation, one may expect an intermittent behavior of the magnetic field in the vicinity of the dynamo threshold. However, bursts of magnetic fields within long time intervals with nearly zero field are not observed in natural dynamos. Stars or planets display a large coherent magnetic field, for instance, quasi-stationary for the Earth or oscillatory for the Sun. It should be understood if this just results from filtering or is due to another mechanism.

Acknowledgments

SF acknowledges the members of the VKS team, M. Bourgoin, J. Burgete, A. Chiffaudel, F. Daviaud, L. Marié, P. Odier, F. Pétrélis, and J.-F. Pinton, with whom experiments of Section 3.3 have been conducted, and A. Nunez and F. Pétrélis, with whom saturation of the magnetic field above the dynamo threshold has been studied. D.P.L would like to gratefully acknowledge the support of the National Science Foundation and the Research Corporation, and advice from T. Antonsen, and E. Ott. The Maryland experiments could not have been conducted without the efforts and advice of D. Martin, N. Peffley, A.B. Cawthorne, D. Sisan, and W. Shew.

References

- Bourgoin, M., *Thèse de doctorat*, ENS-Lyon (2003).
Bourgoin, M., Marié, L., Pétrélis, F., Gasquet, C., Guigon, A., Luciani, J. B., Moulin, M., Namer, F., Burgete, J., Chiffaudel, A., Daviaud, F., Fauve, S., Odier, P. and Pinton, J. F., “Magnetohydrodynamics measurements in the von Kármán sodium experiment,” *Phys. Fluids* **14**, 3046–3058 (2002).

- Busse, F. H., "Thermal instabilities in rapidly rotating systems," *J. Fluid Mech.* **44**, 441–460 (1970).
- Busse, F. H., "A model of the geodynamo," *Geophys. J. R. Astron. Soc.* **42**, 437–459 (1975).
- Busse, F. H., Müller, U., Stieglitz, R. and Tilgner, A., "A two-scale homogeneous dynamo, and extended analytical model and an experimental demonstration under development," *Magnetohydrodynamics* **32**, 235–248 (1996).
- Cardin, P., Brito, D., Jault, D., Nataf, H.-C. and Masson, J.-P., "Towards a rapidly rotating liquid sodium dynamo experiment," *Magnetohydrodynamics* **38**, 177–210 (2002).
- Childress, S. and Soward, A. M., "Convection driven hydromagnetic dynamo," *Phys. Rev. Lett.* **29**, 837–839 (1972).
- Chossat P. and Lauterbach, R., "Methods in equivariant bifurcations and dynamical systems," in: *Advances Series in Nonlinear Dynamics*, vol. **15** World Scientific (2000).
- Colgate, S. A., Pariev, V. I., Beckley, H. F., Ferrel, R., Romero, V. D. and Weatherall, J. C., "The New Mexico $\alpha\Omega$ dynamo experiment: Modelling astrophysical dynamos," *Magnetohydrodynamics* **38**, 129–142 (2002).
- Dudley, M. L. and James, R. W., "Time-dependent kinematic dynamos with stationary flows," *Proc. R. Soc. Lond. A* **425**, 407–429 (1989).
- Fauve, S. and Pétrélis, F., "The dynamo effect," in: *Peyresq Lectures on Non-linear Phenomena*, Vol. II (Ed. J-A Sepulchre), pp. 1–64, World Scientific, Singapore (2003).
- Fauve, S., Laroche, C. and Castaing, B., "Pressure fluctuations in swirling turbulent flows," *J. Physique II* **3**, 271–278 (1993).
- Forest, C. B., Bayliss, R. A., Kendrick, R. D., Nornberg, M. D., O'Connell, R. and Spence, E. J., "Hydrodynamic and numerical modeling of a spherical homogeneous dynamo experiment," *Magnetohydrodynamics* **38**, 107–120 (2002).
- Frick, P., Denisov, S., Khripchenko, S., Noskov, V., Sokoloff, D. and Stepanov, R., "A nonstationary dynamo experiment in a braked torus," in: *Dynamo and Dynamics, a Mathematical Challenge* (Eds. P. Chossat, D. Armbruster and I. Oprea), pp. 1–8, Kluwer Academic Publishers (2001).
- Frick, P., Noskov, V., Denisov, S., Khripchenko, S., Sokoloff D., Stepanov, R. and Sukhanovsky, A., "Non-stationary screw flow in a toroidal channel: way to a laboratory dynamo experiment," *Magnetohydrodynamics* **38**, 107–120 (2002).
- Gailitis, A., Lielausis, O., Platacis, E., Dement'ev, S., Cifersons, A., Gerbeth, G., Gundrum, T., Stefani, F., Christen M. and Will, G., "Magnetic field saturation in the Riga dynamo experiment," *Phys. Rev. Lett.* **86**, 3024–3027 (2001).
- Gailitis, A., Lielausis, O., Platacis, E., Dement'ev, E., Cifersons, A., Gerbeth, G., Gundrum, T., Stefani, F., Christen, M. and Will, G. "Dynamo experiments at the Riga sodium facility," *Magnetohydrodynamics* **38**, 5–14 (2002).
- Gans, R. F., "On hydromagnetic precession in a cylinder," *J. Fluid Mech.* **45**, 111–130 (1970).

- Gilbert, A. D. and Sulem, P. L., "On inverse cascades in alpha effect dynamos," *Geophys. Astrophys. Fluid Dynam.* **51**, 243–261 (1990).
- Golitsyn, G. S., "Fluctuations of the magnetic field and current density in a turbulent flow of a weakly conducting fluid," *Sov. Phys. Dokl.* **5**, 536–539 (1960).
- Herzenberg, A., "Geomagnetic dynamos," *Phil. Trans. R. Soc. Lond. A* **250**, 543–583 (1958).
- Kirko, G. E., Telichko, M. T. and Sheinkman, A. G., "Self-excitation of a magnetic field (MHD dynamo) in the central part of a nuclear reactor with liquid-metal cooling," *Sov. Phys. Dokl.* **28**, 502–503 (1984).
- Labbé, R., Pinton, J.-F. and Fauve, S., "Study of the von Kármán flow between coaxial corotating disks," *Phys. Fluids* **8**, 914–922 (1996).
- Larmor, J., "How could a rotating body such as the sun become a magnet?" Rep. 87th Meeting Brit. Assoc. Adv. Sci., Bournemouth, Sept. 9–13, 1919, pp. 159–160, John Murray London (1919).
- Lehnert, B., "An experiment on axisymmetric flow of liquid sodium in a magnetic field," *Arkiv för Fysik* **13**, 109–116 (1957).
- Léorat, J., Lallemand, P., Guermond, J. L. and Plunian, F., "Dynamo action, between numerical experiments and liquid sodium devices," in: *Dynamo and Dynamics, a Mathematical Challenge* (Eds. P. Chossat, D. Armbruster and I. Oprea), 25–33, Kluwer Academic Publishers (2001).
- Lowes, F. J. and Wilkinson, I., "Geomagnetic dynamo: a laboratory model," *Nature* **198**, 1158–1160 (1963)
- Lowes, F. J. and Wilkinson, I., "Geomagnetic dynamo: an improved laboratory model," *Nature* **219**, 717–718 (1968).
- Malkus, W. V. R., "Precession of the Earth as the cause of geomagnetism," *Science* **160**, 259–264 (1968).
- Marié, L., Burguete, J., Daviaud, F. and Léorat, J., "Numerical study of homogeneous dynamo based on experimental von Kármán type flows," *Eur. Phys. J. B*, to appear (2003a).
- Marié, L., Pétrélis, Bourgoin, M., Chiffaudel, A., Daviaud, F., Fauve, S., Odier, P. and Pinton, J. F., "1/f noise in magnetohydrodynamic turbulence," to be published (2003b).
- Moffatt, H. K., "The amplification of a weak applied magnetic field by turbulence in fluids of moderate conductivity," *J. Fluid Mech.* **11**, 625–635 (1961).
- Nuñez, A., Pétrélis, F. and Fauve, S., "Saturation of a Ponomarenko type fluid dynamo," in: *Dynamo and Dynamics, a Mathematical Challenge* (Eds. P. Chossat, D. Armbruster and I. Oprea), pp. 67–74, Kluwer Academic Publishers (2001).
- Odier, P., Pinton, J. F. and Fauve, S., "Advection of a magnetic field by a turbulent swirling flow," *Phys. Rev. E* **58**, 7397–7401 (1998).
- Odier, P., Pinton J.-F. and S. Fauve, "Magnetic induction by coherent vortex motion," *Eur. Phys. J. B* **16**, 373–378 (2000).
- Parker, E. N., "Cosmical magnetic fields," *Astrophys. J.* **122**, 293–314 (1955).
- Peffley, N. L., Cawthorne, A. B. and Lathrop, D. P., "Toward a self-generating magnetic dynamo: The role of turbulence," *Phys. Rev. E* **61**, 5287–5294 (2000).

- Pétrélis, F. and Fauve, S., "Saturation of the magnetic field above the dynamo threshold," *Eur. Phys. J. B* **22**, 273–276 (2001).
- Pétrélis, F., Bourgoin, M., Marié, L., Burgete, J., Chiffaudel, A., Daviaud, F., Fauve, S., Odier, P. and Pinton, J. F., "Nonlinear magnetic induction by helical motion in a liquid sodium turbulent flow," *Phys. Rev. Lett.* **90**, 174501-1–4 (2003).
- Ponomarenko, Yu. B., "Theory of the hydromagnetic generator," *J. Appl. Mech. Tech. Phys.* **14**, 775–778 (1973).
- Rädler, K.-H., Apstein, E., Rheinhardt, M. and Schüler, M., "The Karlsruhe dynamo experiment. A mean field approach," *Studia geophys. geod.* **42**, 224–231 (1998).
- Roberts, G. O., "Dynamo action of fluid motions with two-dimensional periodicity," *Phil. Trans. R. Soc. Lond. A* **271**, 411–454 (1972).
- Roberts, P. H., "On the thermal instability of a rotating fluid sphere containing heat sources," *Phil. Trans. R. Soc. Lond. A* **263**, 93–117 (1968).
- Roberts, P. H., "Future of geodynamo theory," *Geophys. Astrophys. Fluid Dynam.* **44**, 3–31 and references therein (1988).
- Roberts, P. H. and Jensen, T. H., "Homogeneous dynamos: theory and practice," *Phys. Fluids* **B5**, 2657–2662 (1993).
- Ruzmaikin, A. A. and Shukurov, A. M., "Spectrum of the galactic magnetic fields," *Astrophys. Space Sci.* **82**, 397–407 (1982).
- Shew, W. L., Sisan, D. R. and Lathrop, D. P., "Hunting for dynamos: eight different liquid sodium flows," in: *Dynamo and Dynamics, a Mathematical Challenge* (Eds. P. Chossat, D. Armbruster and I. Oprea), pp. 83–92, Kluwer Academic Publishers (2001).
- Shew, W. L., Sisan, D. R. and Lathrop, D. P., "Mechanically forced and thermally driven flows in liquid sodium," *Magnetohydrodynamics* **38**, 107–120 (2002).
- Sisan, D. R., Shew, W. L. and Lathrop, D. P., "Lorentz force effects in magneto-turbulence," *Phys. Earth Planet. Int.* **135**, 137–159 (2003).
- St. Pierre, M. G., "The strong field branch of the Childress-Soward dynamo," in: *Theory of Solar and Planetary Dynamos* (Eds. M. R. E. Proctor, P. C. Matthews and A. M. Ruchlidge), pp. 295–302, Cambridge University Press (1993).
- Soward, A. M., "A convection-driven dynamo: I. The weak field case," *Phil. Trans. R. Soc. Lond. A* **275**, 611–646 (1974).
- Stieglitz, R. and Müller, U., "Experimental demonstration of a homogeneous two-scale dynamo," *Phys. Fluids* **13**, 561–564 (2001).
- Stieglitz, R. and Müller, U., "Experimental demonstration of a homogeneous two-scale dynamo," *Magnetohydrodynamics* **38**, 27–34 (2002).
- Sweet, D., Ott, E., Finn, J., Antonsen, T. M. and Lathrop, D., "Blowout bifurcations and the onset of magnetic activity in turbulent dynamos," *Phys. Rev. E* **63**, 066211, 1–4 (2001).
- Tilgner, A., "Onset of dynamo action in an axisymmetric flow," *Phys. Rev. E* **66**, 017304, 1–4 (2002).
- Tilgner, A. and Busse, F. H., "Saturation mechanism of a model of the Karlsruhe dynamo," in *Dynamo and Dynamics, a Mathematical Challenge* (Eds.

- P. Chossat, D. Armbruster and I. Oprea), pp. 109–116, Kluwer Academic Publishers (2001).
- Tilgner, A. and Busse, F. H., “Simulation of the bifurcation diagram of the Karlsruhe dynamo,” *Magnetohydrodynamics* **38**, 35–40 (2002).
- Zandbergen, P. J. and Dijkstra, D., “von Kármán swirling flows,” *Annu. Rev. Fluid Mech.* **19**, 465– (1987).
- Zeldovich, Ya. B., Ruzmaikin, A. A. and Sokoloff, D. D., *Magnetic Fields in Astrophysics*, Gordon and Breach, New York (1983).

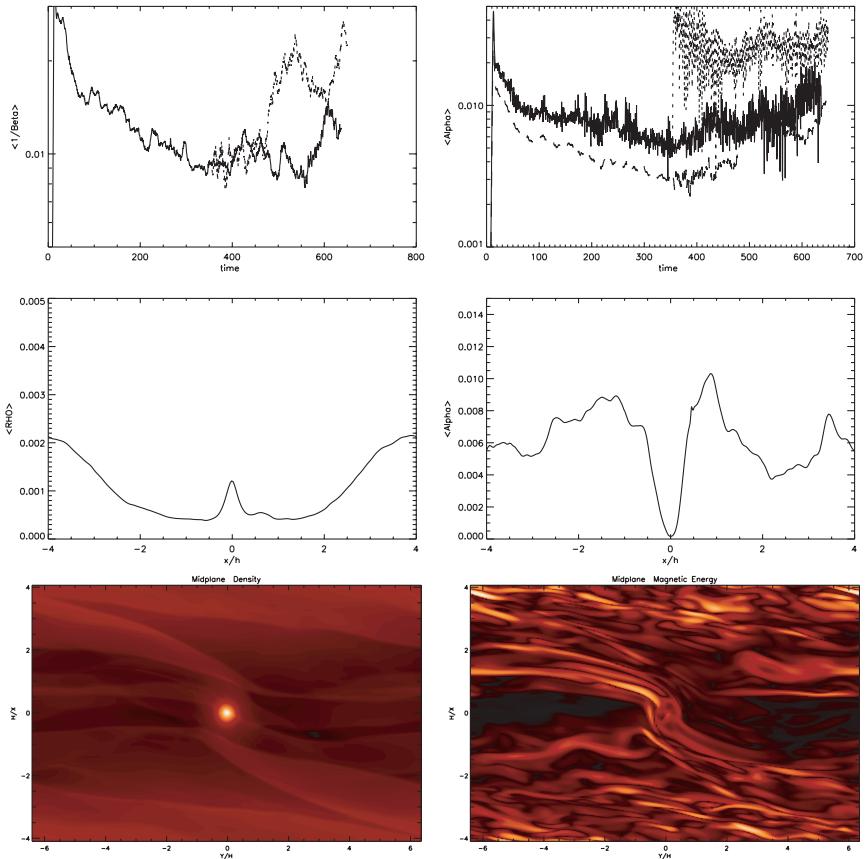


Figure 2.22. This figure illustrates the results of local simulations carried out in a shearing box. The first simulation had no perturbing planet while in the second a perturbing planet with $m_p/M_\odot = 2H^3/R^3$ was introduced at time 353 and continued from then. The upper left panel shows $1/\langle \beta \rangle$ against time for the first simulation (full line) and the second beyond $t = 353$ (dashed line). The upper right panel shows $\langle \alpha \rangle$ against time for the first simulation (full line). The uppermost dotted curve corresponds to the continuation with the planet present and is dominated by the Reynolds' stress contribution. The contribution of the magnetic stress is given by the lowest curve with times beyond 353 corresponding to the second simulation. The left middle panel shows the vertically and y averaged density near the end of the second simulation indicating the gap. The right middle panel shows the vertically and azimuthally averaged total stress as a function of x near the end of the second simulation. The lower left panel is a contour plot of the midplane density near the end of the second simulation and the lower right panel is a contour plot of the midplane magnetic energy near the end of the second simulation.

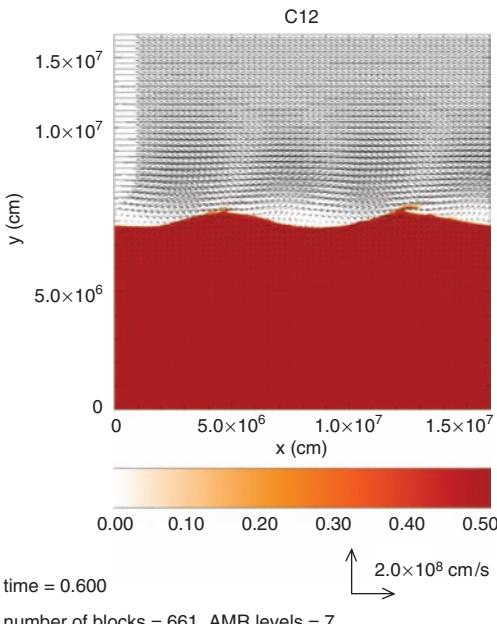


Figure 3.4. A snapshot of a two-dimensional simulation of a breaking gravity wave on the surface of a white dwarf, using the Flash code (from Alexakis *et al.*, 2004a). The lines and arrows are meant to indicate the nature of the flow/wind overlying the perturbed interface.

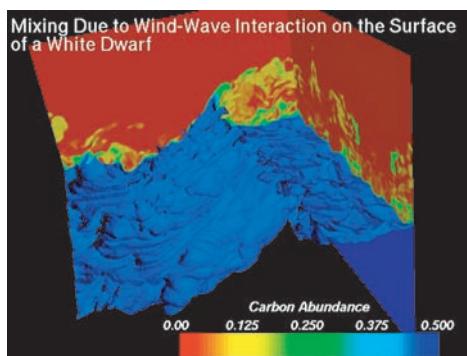


Figure 3.5. A snapshot of a three-dimensional simulation of a breaking gravity wave on the surface of a white dwarf. The color scale indicates carbon abundance.

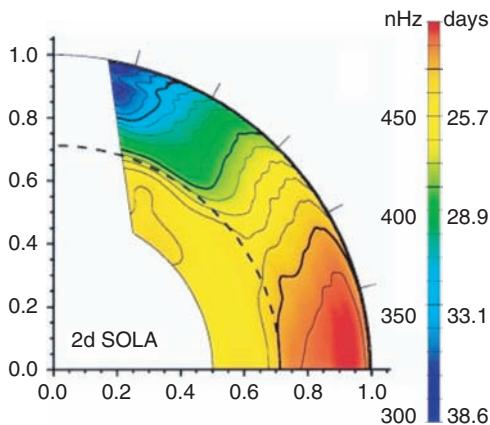


Figure 7.1. Inversion for the interior solar rotation rate $\Omega/2\pi$ with radius and latitude using the two-dimensional SOLA inversion technique (from Schou et al., 1998). The dashed line represents the base of the convection zone.

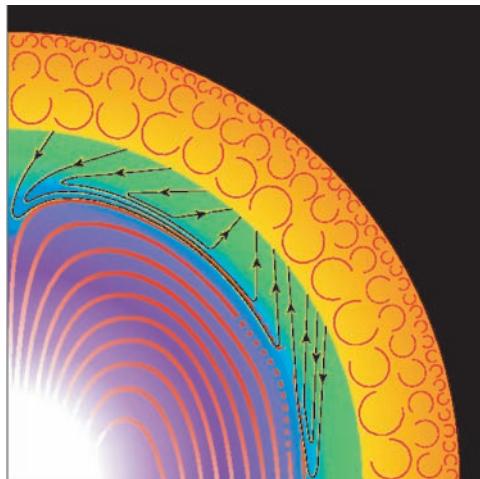


Figure 7.3. Schematic representation of a meridional quadrant of the Sun (after Gough and McIntrye, 1998). The arrows represent the tachocline circulation, which follows surfaces S of constant specific angular momentum in the (green) body of the tachocline (whose thickness has been exaggerated by a factor 5), and is deflected by the magnetic field in the (blue) diffusive boundary layer (whose thickness has been exaggerated by a factor 50).

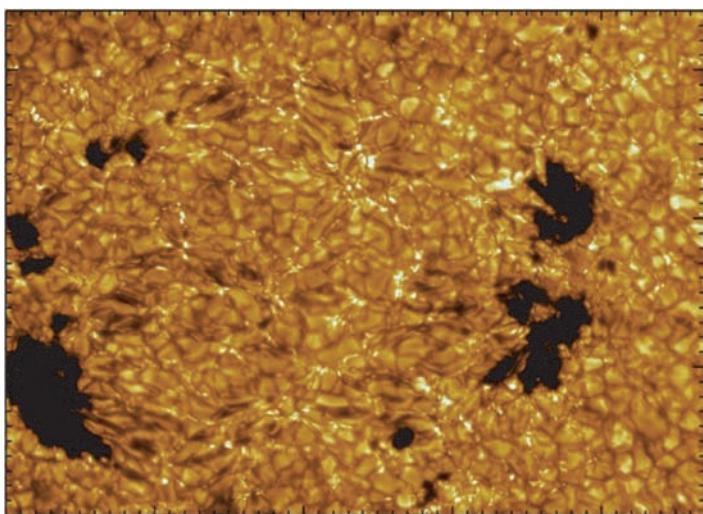


Figure 8.1. Filtergram of an active region. Sunspots, pores and magnetic bright points can be clearly seen. (Courtesy T. E. Berger.)

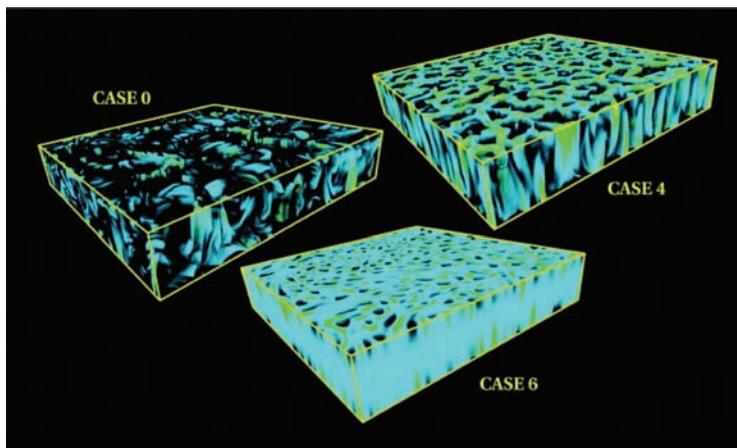


Figure 8.6. Volume rendering for magnetic fields in Boussinesq magnetoconvection (from Cattaneo *et al.*, 2002). Case 0 is a dynamo, with no imposed field, while cases 4 and 6 have moderate and large imposed fields, respectively.

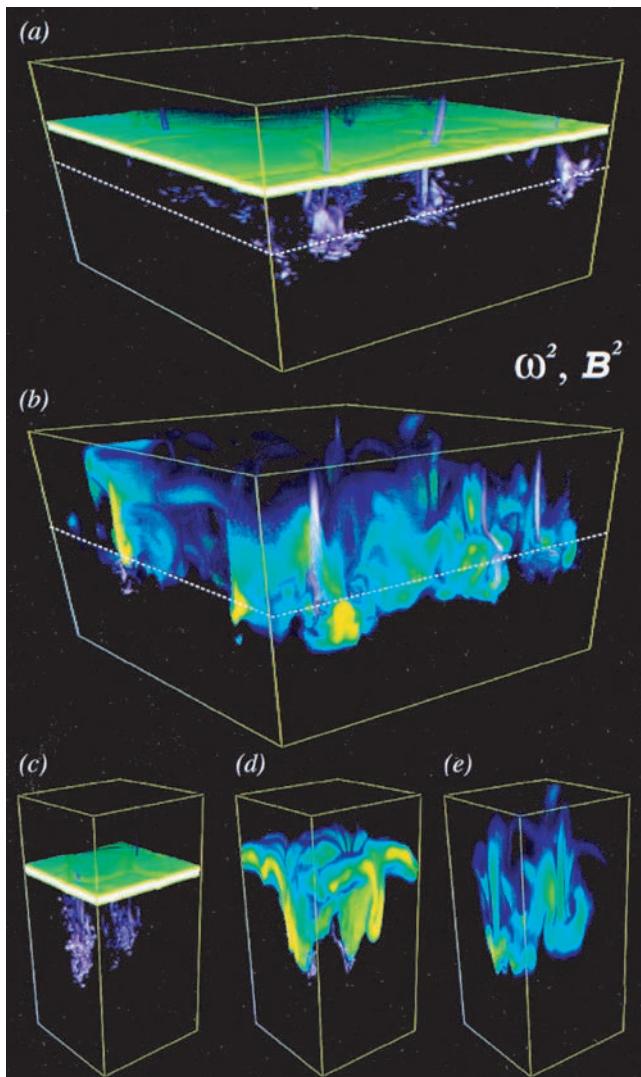


Figure 8.13. Pumping of horizontal fields by three-dimensional convection above a stably stratified layer (from Tobias *et al.*, 2001). (a) Initial configuration with a layer of magnetic field inserted in the unstable convection zone. (b) Later time, showing concentrations of the magnetic field in the stable region. (c)–(e) Volume renderings for a subvolume of the full domain centered around a couple of coherent downflows. The strong plumes pump magnetic flux downward and amplify the magnetic energy locally by inductive processes.

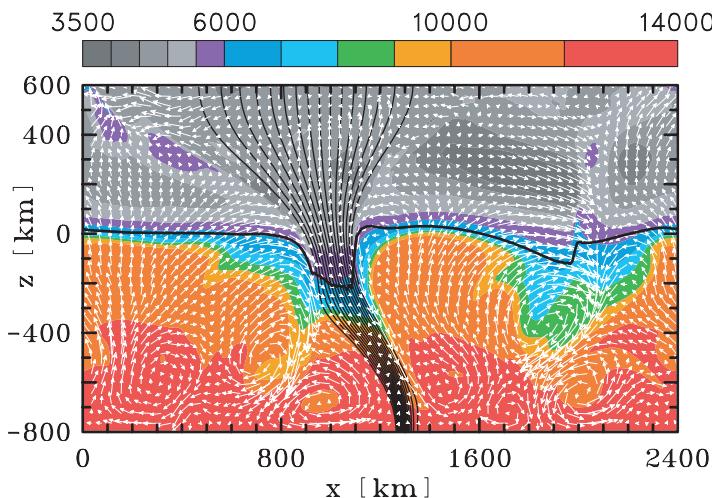


Figure 8.15. Two-dimensional magnetoconvection with radiative transfer (courtesy of O. Steiner). An intense flux tube can be seen, together with the depression of the surface due to evacuation and cooling.

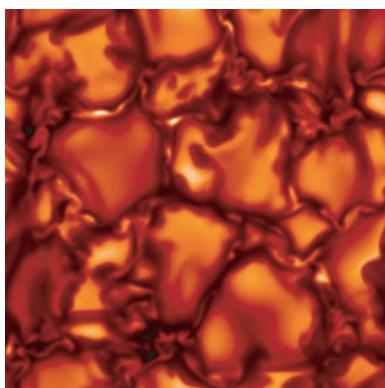


Figure 8.16a. Three-dimensional magnetoconvection with radiative transfer (from Vögler and Schüssler, 2003). Left image shows field intensity, right image vertical field strength. Averaged imposed field is 200G. Note the prominent mesogranular scale of the magnetic field.

Figure 11.1. Snapshot of the 3D magnetic field structure for a Glatzmaier-Roberts geodynamo simulation, illustrated with a set of magnetic lines of force, which are blue where directed inward and gold where directed outward. The axis of rotation is vertical and centered in the image. The lines are drawn out to two Earth radii. The field is a smooth, dipole-dominated, potential field outside the core. (Reprinted with permission from the Annual Review of Earth and Planetary Sciences, Volume 30 ©2002 by Annual Reviews, www.annualreviews.org.)

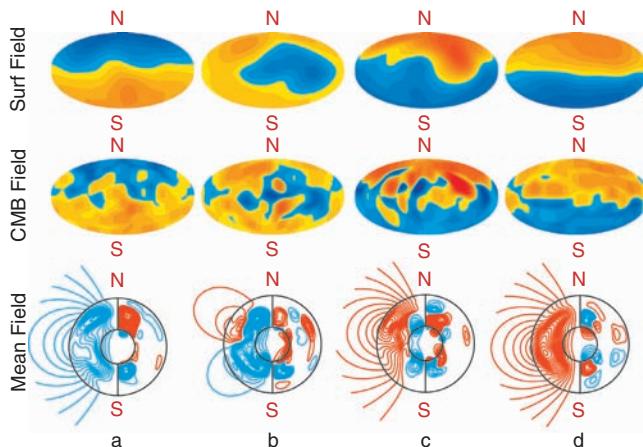
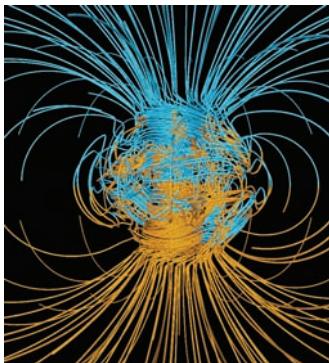


Figure 11.4. A sequence of snapshots of the longitudinally averaged magnetic field through the interior of the core and of the radial component of the field at the core-mantle boundary and at what would be the surface of the Earth, displayed at roughly 3000-year intervals spanning a dipole reversal from a geodynamo simulation. In the plots of the average field, the small circle represents the inner core boundary and the large circle is the core-mantle boundary. The poloidal field is shown as magnetic field lines on the left-hand sides of these plots (blue is clockwise and red is counterclockwise). The toroidal field direction and intensity are represented as contours (not magnetic field lines) on the right-hand sides (red is eastward and blue is westward). Aitoff-Hammer projections of the entire core-mantle boundary and surface are used to display the radial field (with the two different surfaces displayed as the same size). Reds represent outward directed field and blues inward field; the surface field, which is typically an order of magnitude weaker, was multiplied by 10 to enhance the color contrast. (Reproduced with permission of *Nature*, from figure 2, Glatzmaier, G. A., Coe, R. S., Hongre, L. and Roberts, P. H., “The role of the Earth’s mantle in controlling the frequency of geomagnetic reversals,” *Nature* **401**, 885–890 (1999), copyright by Nature, www.nature.com.)

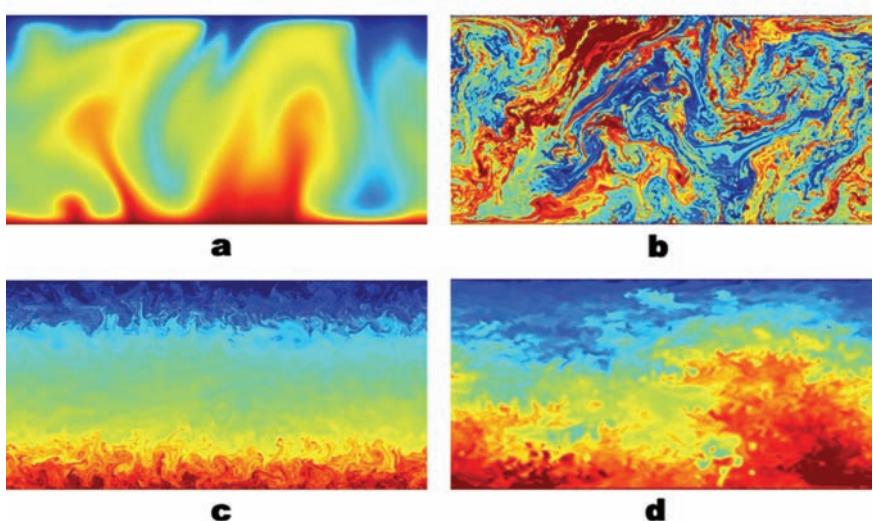


Figure 11.5. Snapshots of the entropy for four different 2D anelastic magnetoconvection calculations. Reds are high entropy; blues are low. Gravity is downward, the bottom boundary is hot and the top is cold. (a) Rotating and magnetic, $Ra = 3 \times 10^6$, $Ek = 10^{-4}$, $Pr = 1$, $Q = 10^4$, $q = 1$, $N_\rho = 0.2$ and $n = 1$, (b) nonrotating and nonmagnetic, $Ra = 10^{12}$, $Pr = 1$, $N_\rho = 0.2$ and $n = 1$, (c) rotating but nonmagnetic, $Ra = 3 \times 10^{12}$, $Ek = 10^{-9}$, $Pr = 1$, $N_\rho = 0.2$ and $n = 1$, (d) rotating and magnetic, $Ra = 10^{11}$, $Ek = 10^{-9}$, $Pr = 1$, $Q = 10^8$, $q = 1$, $N_\rho = 2.5$ and $n = 1.5$.

Entropy

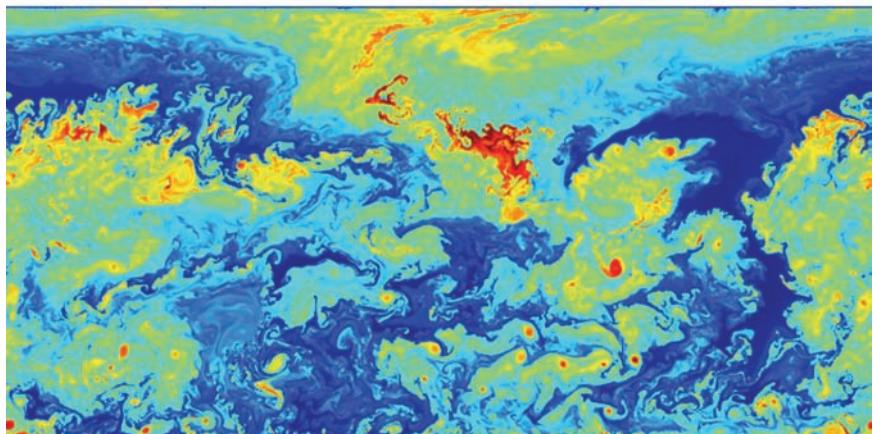


Figure 11.7. A snapshot of the entropy for a non-magnetic case with $Ra = 2 \times 10^{12}$, $Ek = 10^{-9}$, $Pr = 0.1$, $N_\rho = 5$ and $n = 1.5$.

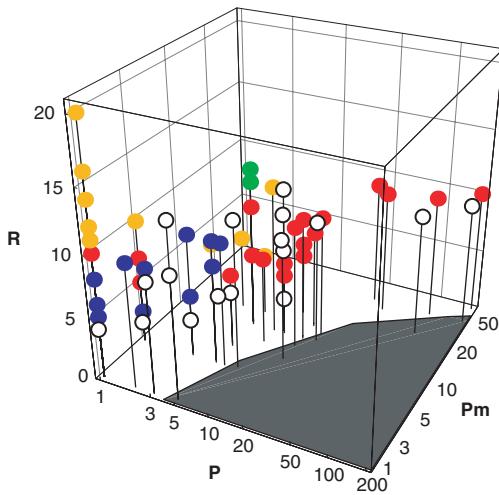


Figure 12.12. Dynamo solutions indicated by red (dipolar), blue (quadrupolar), green (hemispherical) and yellow (mixed symmetry) balls in the $R - P - P_m$ parameter space. No dynamo solution could be obtained for values of P, P_m in the shaded region. The values of the ordinate should be multiplied by the factor 10^5 .

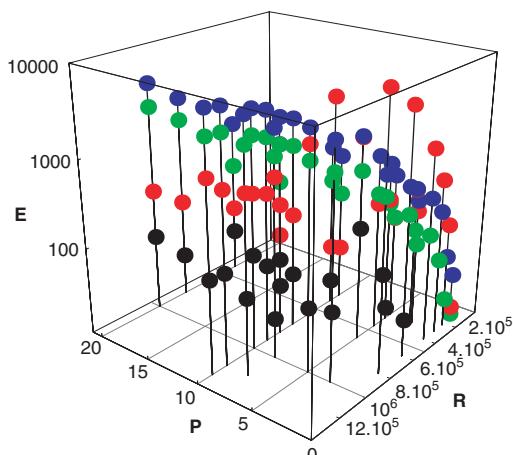


Figure 12.14. Dependences of energy densities of the axisymmetric toroidal (red), axisymmetric poloidal (black, multiplied by 10^2), non-axisymmetric toroidal (blue) and non-axisymmetric poloidal (green) components of motion on R and P in the case of $\tau = 5 \times 10^3$. The energy densities have been multiplied by P^2 and thus are measured in terms of the thermal scaling.

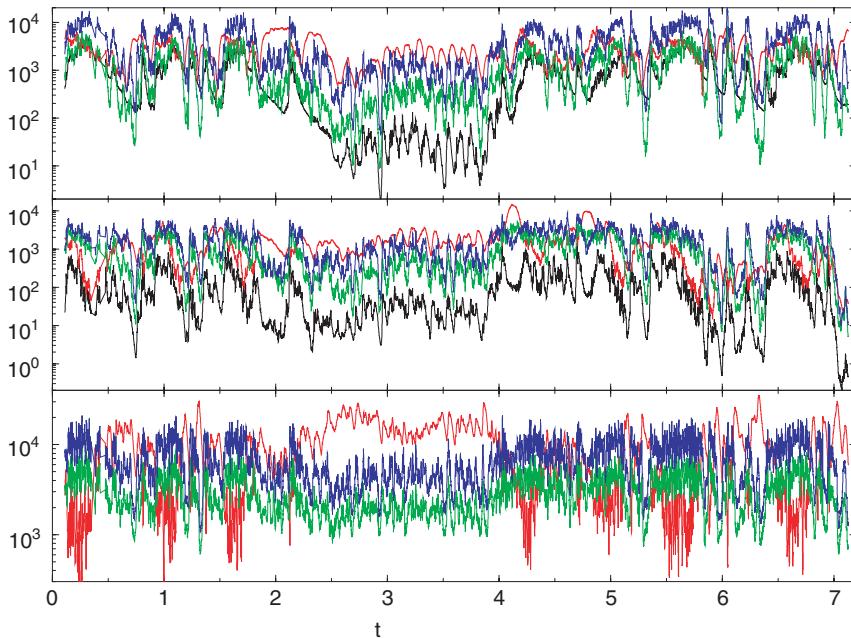
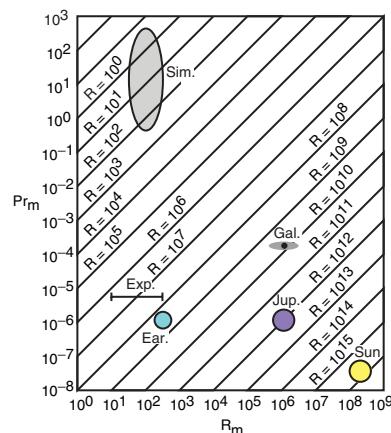


Figure 12.19. Time series of a convection driven dynamo with $\tau = 3 \times 10^4$, $R = 8.5 \times 10^5$, $P = 0.1$, $P_m = 1$. The first, second and third plot from the top show energy densities of the dipolar and quadrupolar components of the magnetic field and of the velocity field, respectively. The mean and fluctuating toroidal energy densities are indicated by red and blue lines, while the mean and fluctuating poloidal energy densities are black and green, respectively.

Figure 13.1. Relevant parameters for the Earth (Ear.), simulations (Sim.), Jupiter (Jup.), the Sun, estimates for galaxy (Gal.), and experiments (Exp.), showing the Reynolds number Re , magnetic Prandtl number P_m , and the magnetic Reynolds number R_m . Note that $R_m = Re P_m$ (some estimates from Zeldovich *et al.*, 1983).



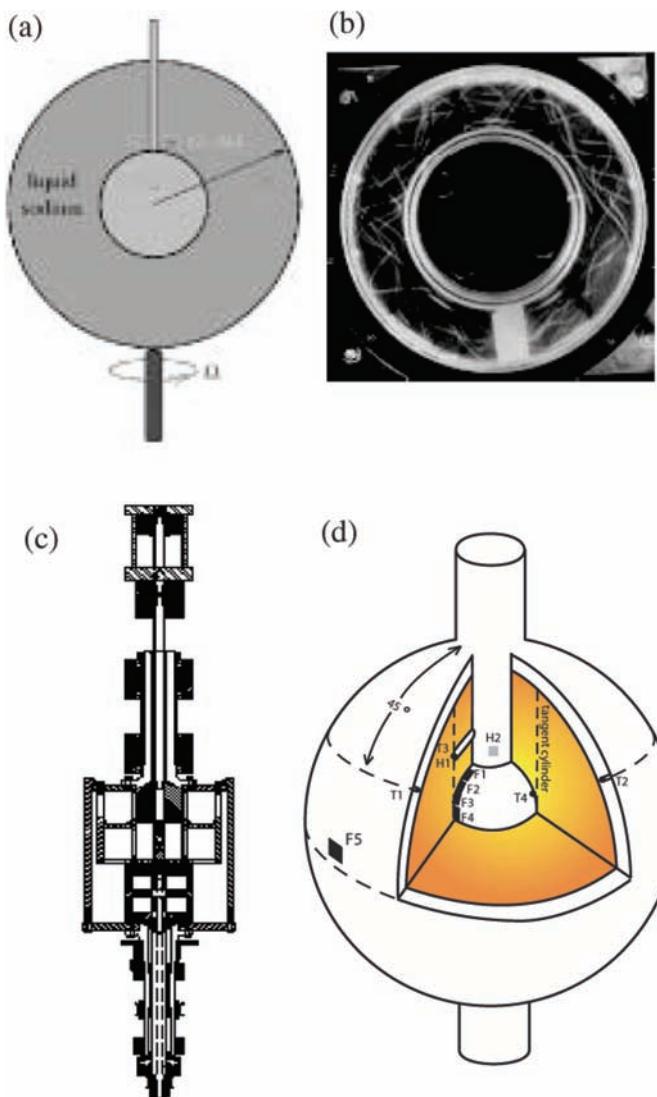


Figure 13.14. (a) Cross section of the Grenoble sodium experiment, which has differentially rotating inner and outer spheres. The inner sphere is a permanent magnet (see Cardin, *et al.*, 2002). (b) A sodium experiment is planned in Perm, where a rotating torus of liquid is rapidly decelerated, leading to a helical motion of the liquid within the torus (from Frick *et al.*, 2002). (c) In New Mexico, a rotating Couette flow is planned with axial jets of liquid to also stimulate helical flows in sodium (from Colgate *et al.*, 2002). (d) A Maryland experiment with a 60 cm rotating convection experiment, where centrifugal forces drive convection between a heated outer sphere and a cooled inner sphere (see Shew *et al.*, 2002).