MAGNETOHYDRODYNAMICS AND THE EARTH'S CORE

Selected works of Paul Roberts

Edited by

Andrew Soward

University of Exeter, UK

Taylor & Francis
Taylor & Francis Group
LONDON AND NEW YORK

MAGNETOHYDRODYNAMICS AND THE EARTH'S CORE

The fluid mechanics of astrophysics

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 Zeldovich, 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
S I Vaintshtein, 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 Antonio Ferriz-Mas and Manuel Nuñez

Volume 10
Magnetohydrodynamics and the Earth's Core
Selected works of Paul Roberts
Edited by A M Soward

MAGNETOHYDRODYNAMICS AND THE EARTH'S CORE

Selected works of Paul Roberts

Edited by

Andrew Soward

University of Exeter, UK

Taylor & Francis
Taylor & Francis Group
LONDON AND NEW YORK

First published 2003 by Taylor & Francis
11 New Fetter Lane, London EC4P 4EE
Simultaneously published in the USA and Canada by Taylor & Francis Inc,
29 West 35th Street, New York, NY 10001

Taylor & Francis is an imprint of the Taylor & Francis Group

This edition published in the Taylor & Francis e-Library, 2005.

Publisher's Note
This book has been prepared from camera-ready-copy.

Printed and bound in Great Britain by T J International Ltd, Padstow, Cornwall

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Cataloging in Publication Data
A catalog record for this title has been requested.

ISBN 0-203-27222-X Master e-book ISBN

To Maureen

CONTENTS

Foreword xi
Commentary on publications xii
The 1999 John Adam Fleming Medal xxi
Acknowledgements xxv

PART 1
Geomagnetism and dynamos **1**

Propagation of induced fields through the core 3
P.H. ROBERTS

On analysis of the secular variation – A hydromagnetic constraint: theory 15
P.H. ROBERTS AND S. SCOTT

Some comments on the theory of homogeneous dynamos 30
R.D. GIBSON AND P.H. ROBERTS

Nearly symmetric hydromagnetic dynamos 43
J.G. TOUGH AND P.H. ROBERTS

Magnetic core–mantle coupling by inertial waves 52
P.H. ROBERTS

A three-dimensional kinematic dynamo 54
S. KUMAR AND P.H. ROBERTS

Are planetary dynamos driven by gravitational settling? 78
DAVID E. LOPER AND PAUL H. ROBERTS

A study of conditions at the inner core boundary of the Earth 80
DAVID E. LOPER AND PAUL H. ROBERTS

Structure of the Earth's inner core 86
DAVID R. FEARN, DAVID E. LOPER AND PAUL H. ROBERTS

Time-dependent electromagnetic core–mantle coupling 88
M. STIX AND P.H. ROBERTS

On topographic core–mantle coupling 100
PAUL H. ROBERTS

A three-dimensional convective dynamo solution with rotating and finitely conducting inner core and mantle 107
GARY A. GLATZMAIER AND PAUL H. ROBERTS

A three-dimensional self-consistent computer simulation of a geomagnetic field reversal 120
GARY A. GLATZMAIER AND PAUL H. ROBERTS

An anelastic evolutionary geodynamo simulation driven by compositional and thermal convection 127
GARY A. GLATZMAIER AND PAUL H. ROBERTS

Rotation and magnetism of Earth's inner core 141
GARY A. GLATZMAIER AND PAUL H. ROBERTS

PART 2
Fluid mechanics and MHD **147**

Characteristic value problems posed by differential equations arising in hydrodynamics and hydromagnetics 149
P.H. ROBERTS

On the motion of a liquid in a spheroidal cavity of a precessing rigid body 169
K. STEWARTSON AND P.H. ROBERTS

The ellipticity of a slowly rotating configuration 189
S. CHANDRASEKHAR AND P.H. ROBERTS

Singularities of Hartmann layers 197
P.H. ROBERTS

The critical layer in stratified shear flow 211
P. BALDWIN AND P.H. ROBERTS

On the hydromagnetics of rotating fluids 229
IBRAHIM A. ELTAYEB AND PAUL H. ROBERTS

A transformation of the stellar wind equations 232
P.H. ROBERTS

A hamiltonian theory for weakly interacting vortices 234
P.H. ROBERTS

A unified approach to mean field electrodynamics 245
P.H. ROBERTS AND A.M. SOWARD

MAC waves 261
P.H. ROBERTS

On the motion of a fluid that is incompressible in a generalized sense and its relationship to the Boussinesq approximation 264
R.N. HILLS AND P.H. ROBERTS

Nonlinear theory of localized standing waves 272
BRUCE DENARDO, ANDRÉS LARRAZA, SETH PUTTERMAN AND PAUL ROBERTS

Ideal instabilities in rapidly rotating MHD systems that have critical layers 276
SIKUN LAN, WEIJIA KUANG AND PAUL H. ROBERTS

Structure and stability of a spherical implosion 304
P.H. ROBERTS AND C.C. WU

Thermal inertial waves in a rotating fluid layer: Exact and asymptotic solutions 310
KEKE ZHANG AND PAUL H. ROBERTS

The decay of bubble oscillations 318
P.H. ROBERTS AND C.C. WU

PART 3
Superfluidity **321**

Roton–phonon relaxation times in helium II 323
R.J. DONNELLY AND P.H. ROBERTS

Nucleation of quantized vortex rings by ions in helium II 325
RUSSELL J. DONNELLY AND PAUL H. ROBERTS

Theory of the onset of superflow 330
PAUL H. ROBERTS AND RUSSELL J. DONNELLY

Motion of ions at finite velocities in helium II 335
DONALD M. STRAYER, RUSSEL J. DONNELLY AND PAUL H. ROBERTS

Dynamics of rotons 340
P.H. ROBERTS AND R.J. DONNELLY

A simple theory of temperature-dependent energy levels: application to He II 342
P.H. ROBERTS AND R.J. DONNELLY

The electron bubble in liquid helium 345
P.H. ROBERTS

Calculation of the static healing length in helium II 363
P.H. ROBERTS, R.N. HILLS AND R.J. DONNELLY

Superflow in restricted geometries 367
R.J. DONNELLY, R.N. HILLS AND P.H. ROBERTS

Non-linear hydrodynamics and a one fluid theory of superfluid He^4 371
S.J. PUTTERMAN AND P.H. ROBERTS

What happens to vortex rings that shrink? 375
C.A. JONES AND P.H. ROBERTS

PART 4
Mixed phase regions **377**

On the life cycle of a mush 379
R.N. HILLS, P.H. ROBERTS AND A.M. SOWARD

Relaxation effects and the evolution of a mush 405
R.N. HILLS AND P.H. ROBERTS

Publication list 422

FOREWORD

Between Thursday 15th and Saturday 17th July 1999 the Paul Roberts' Seventieth Anniversary Meeting was held at the School of Mathematical Sciences at the University of Exeter. During that meeting a number of Paul's close colleagues gave substantial talks on topics which focused on their past research interactions, examples of which will be found amongst the selected works that appear here. As part of those celebrations, we encouraged Paul to collect together those papers, that he saw fit, into this single volume. His research contributions are remarkable both in their multitude, diversity, depth of content and international importance. A particular merit of this volume is Paul's interesting choice of papers, which includes some very fine articles that appeared in diverse and possibly remote journals. It is therefore extremely valuable to have them all readily accessible in this one volume, even though it can only catch the flavour of his remarkable achievements.

Paul has, however, kindly provided a Commentary of the papers included to help guide us through; he has put his work into historical context focusing for us on the issues of the time. The result is a wonderful compendium giving a clear overview of many of Paul's seminal contributions, which have served as the foundation for much subsequent research.

It was extremely timely that in 1999 the American Geophysical Union honoured Paul with the award of the John Fleming Medal. The complete Citation, made by Masaru Kono, is included here as well as Paul's Response. Any further Editorial words seem superfluous; the Citation, the Response and Paul's Commentary say it all.

Andrew Soward, 2000

COMMENTARY ON PUBLICATIONS

Anyone who troubles to look at my publication list, is sure to wonder how someone can possibly have moved so randomly amongst such a diverse set of topics, and this is meant to explain how it happened. It also talks about some interesting scientists I have met.

My career in research started in 1951 when Herman Bondi, whose lectures to us as undergraduates had been so inspirational, agreed to advise me for the next three years. To my disappointment, he said he had run out of astrophysical projects suitable for Ph.D. students, but suggested that I should see whether fluid dynamos could exist, or whether Cowling's theorem was the foretaste of a stronger result. Having made essentially no progress in a year, apart from learning something about the young subject of magnetohydrodynamics (MHD), it was time to switch topic and advisor. Keith Runcorn took me on, to work on geomagnetic subjects, such as the screening of the magnetic signals from the core by the mantle. I made one interesting discovery that overturned a popular belief of the time, namely that the sources of the geomagnetic secular variation all lay near the surface of the core (e.g., Lowes and Runcorn, 1951). This error was the result of a mis-application of the skin-depth argument familiar from the electrodynamics of solid conductors. My new found knowledge of MHD had paid off: I could point out that in a fluid, sources deep in the core could transmit magnetic signals to its surface with essentially no diminution. Runcorn practically fell out of his chair, and then told me I must have made a mistake. That kind of reaction to one's research tells one that one has either made a monumental error or has really succeeded in doing something worthwhile. In fact people I mixed with in 1952 had little knowledge of MHD. I didn't get round to publishing my idea until 1955 [3,8,12]* and by that time Runcorn had (characteristically) already published the result himself with a very vague acknowledgment to myself (Runcorn, 1954).

In 1954, I moved to the Yerkes Observatory of the University of Chicago to become a research associate of Chandrasekhar, who was able to convince me that many new and exciting things happened outside Cambridge. Chandra guided largely by example, working longer hours than anyone else; he even had some of us working on Christmas Day. He was held in great awe, there being widespread speculation that his first name was not Subrahmanyan but Superman. I learned how a successful research scientist is supposed to operate. Chandra also introduced me to turbulence theory [4], stability theory (on which he was working at the time), and noise and stochastic processes (on which he had been working previously). I made use of the latter in my interaction [11] with a remarkable Jesuit priest,

* Numbers in square brackets refer to papers in the reference list at the end of the book.

Flor Bertiau, who was also at Yerkes at that time. (Later Bertiau paid me a unique compliment by dedicating a small book on computing algorithms to me; see Bertiau, 1977. He also commissioned three short reviews [6,7,9].)

Chandra taught me the importance of publishing what one has done ("It's the record that counts," he would say). Inspired by this, I acquired a distinction that few can match: my first *two* publications [1,2] were seriously flawed, a fact I soon realized though I was too discouraged and embarrassed to do anything about it (but see Skiles, 1972: Wentzel, 1960). Chandra was interested in the nascent CTR field and suggested that I analyze the stability of twisted magnetic fields. Computers were primitive in those days, and I was pleased to get [5] a new, exactly soluble model, later used by Trehan. Chandra convinced me that it is vital to present one's work well. He himself claimed that everything he wrote had been through at least four drafts. Busy man though he was, he found time to give me lessons on how to organize mathematics in print. He himself would count the number of symbols in a long displayed equation, so that he could instruct the compositors where it should be broken.

I found Chandra a fair-minded man, as the following anecdote illustrates. He wrote a paper about waves on an electrically conducting ocean permeated by a vertical magnetic field. I had the temerity to tell him that his solution was fundamentally wrong. He took this well, but he was such a resourceful debater that within an hour he had convinced me that it was I who was in error, not him. After a few weeks I was no longer convinced, and returned to the fray, but with the same outcome. After I returned to England, I decided, with some trepidation, to put my point of view to the test by submitting it for publication. I was very relieved when the editor accepted it without comment (see Appendix to [22]). The editor was Chandra himself. Who was right? Of course, I still think I was, and was encouraged in this when later I came across something Evry Schatzman had written but which I can no longer trace.

I was brought down to Earth after my all too short a time in Chicago by having to return to England to do my military service. Having discharged this in a year, I moved as an ICI Fellow to the Physics Department at King's College of Durham University (but situated in Newcastle); in 1963 this became the independent University of Newcastle upon Tyne. Here I made friends with Tomomasa Tatsumi from Kyoto, then a British Council Scholar. I took him to my parents' home in Aberystwyth for Christmas in 1955 and at Easter in 1956, and we wrote a paper [16] on MHD turbulence. Unlike Chandra's papers on this subject, we did not assume that the field-flow correlation $\overline{\nabla B}$ was zero, and in this sense we anticipated the later developments of Yoshizawa (1990), but we also missed an opportunity, since we could easily have built on Parker's (1955) epoch-breaking paper, and have formalized the α-effect as Steenbeck, Krause and Rädler (1966) did later; we might even have discovered helicity (Steenbeck and Krause, 1966) well ahead of the field. With results [10] certainly (better) forgotten, I applied the methods of [16] to turbulent diffusion, and Bob Kraichnan, then at the Courant Institute, invited me there for six months to work on the same topic [17,18,90]. There he taught me the direct interaction approximation that he had recently developed.

At that time, the head of my department, Keith Runcorn, did not believe in continental drift, and favored polar wandering to explain mysterious paleomagnetic results. He had not located a solution to the random walk problem on a sphere, and suggested I should tackle it, something I was loth to do, because I was sure the solution must be "out there somewhere." Having been unsuccessful in finding it, I delighted in another opportunity of applying what Chandra had taught me about stochastic processes. To my chagrin, the Professor of Theoretical Physics, Stanley Rushbrooke, told me soon after that Harold Ursell had obtained

my solution a decade earlier, but had not published it because he thought someone else must have done so already! This brought me into contact with that remarkably talented and scholarly mathematician at Leeds University, Harold Ursell, a man who seemed equally at home in many branches of pure and applied mathematics. As an undergraduate, he developed (Ursell, 1927) a method of dealing with imperfect gases in statistical mechanics, later to become the "Ursel–Meyer expansion." He was, I believe, one of the first people to recognize the power of diagram expansions in theoretical physics.

Ursell and I agreed to collaborate on the random walk problem. He had obtained new results on convergence and asymptotic properties; making use of the tensor calculus I had learned from Bondi in Part 3 of the Math Tripos in Cambridge, I had generalized to random walks on Riemannian manifolds, a new direction for Ursell but one in which he quickly took the lead. This paper also brought me into closer contact with another remarkable man, who "out of the blue" wrote to say that our solution must be related to elliptic functions (and so it was, though somewhat obliquely). This was Ronald Fisher who, perhaps because of his very poor eyesight, had developed an extraordinary geometrical way of thinking. His wide interests embraced both geophysics and, of course, probability theory. My pleasure in random walk problems was renewed many years later when Pat Moran wrote suggesting that a generalization of [13] to random rotations would be of interest. This resulted in another enjoyable bout with randomness [147]. Papers [13] and [147] have found applications in earthquake research; e.g., see Kagan and Knopoff (1985). The final irony was that the random walk problem on a sphere had indeed been solved (in the diffusion limit) by Perrin early in the 20th century.

Chandra was at this time involved with solving eigenvalue problems of hydrodynamic stability by variational methods. Although these were strictly non-self-adjoint systems, he found ways of formulating them so that he could bound the eigenvalues from one side. I became interested in seeing how accurate the direct non-self-adjoint variational formulation would be. The results were surprisingly good [14] and impressed Chandra, despite the fact that the bounding property was lost. He took up the topic himself; see Chandrasekhar (1961a, Appendix IV; 1961b). Although I had by now largely forgotten my first Ph.D. topic, I remembered it now and set up the adjoint dynamo problem, which has since found a place in fast dynamo theory: e.g., Childress and Gilbert (1995). See also [47,188], and Proctor (1977). At about this time, I collaborated with Frank Lowes on Earth currents [20] and with Raymond ("Spike") Hide on an article [15] and two reviews [19,25]. The latter contains in Section XI a foretaste of topographic core–mantle coupling, one of the novel ideas for which Hide is so well known.

After a brief spell as a Lecturer in Physics at Newcastle, I returned to Yerkes Observatory in 1961 as an Associate Professor of Astronomy. I found Chandra rapidly developing the virial method for studying the stability of rotating, self-gravitating fluid "stars" of uniform density; see also [24]. As rotation increases, the sequence of Maclaurin spheroids bifurcates, and triaxial Jacobi ellipsoids become possible. Beyond this bifurcation point, both equilibria are possible and, in the absence of dissipation, both are stable. Since however the Jacobi ellipsoid has the lower energy, it had always been supposed that the Maclaurin spheroid is unstable when viscosity is allowed for. This had never been demonstrated by direct solution of the fluid equations, and I realized that this was a good problem to solve by asymptotic methods. At that time, Keith Stewartson, then the Head of Applied Mathematics at Durham University, was visiting the Mathematics Research Center in Madison a few miles up the road from Yerkes, and he guided me through a knotty difficulty raised by a boundary layer singularity [27]. During this period, I interested Keith in the question of what flows will be

set up in the fluid contained in a spheroidal cavity (such as the Earth's core), rotating and precessing about an axis fixed in space. I pointed out that, to a modern fluid dynamicist, Poincaré's argument was unsatisfactory, even though it might well have led to the right answer. Keith was perhaps then at the height of his powers, and it was impressive to see the speed at which he progressed [29]. One day we were arguing about (and correcting) the Bondi and Lyttleton (1953) formulation; the next day, it seemed, everything was done. We later wrote on spherical precession [36], where the flow is driven by viscous stresses rather than pressure differences. This work put me in contact with another remarkable man, Harold Jeffreys, who however felt we had advanced not one iota beyond Poincaré. As a result of one of Harold's suggestions, I wrote, for the one and only time, a paper [42] in a single day! To distance myself from the unreality of uniform density in the Maclaurin and Jacobi sequences, I commenced a series of studies on rotating polytropes [28,30,34,38,40], but these were soon eclipsed by James (1964). I collaborated with Chandra for the one and only time [31] in a paper that put bounds on the ellipticity of a slowly rotating configuration of varying density.

My work with Stewartson introduced me to the boundary layer singularity, and I was lucky enough to be able to determine [50] how Hartmann layers connect on either side of the singularity that arises where the imposed magnetic field is tangential to the wall of an MHD duct; see also Waechter (1969) and Grasman (1971).

While in Yerkes, I also continued collaborating with some of the faculty at the Rutherford College of Technology, now the University of Northumbria. With one of these, I wrote two papers. One [32] is best forgotten; the other [41], published after an enormously long gestation period (for Stan Scott is a very thorough man), became so well known that the source is now rarely referenced. The motivation underlying both papers was to extrapolate the geomagnetic field down to the core surface, to learn what it is like "down there," and to find out something about the core motions that maintain the geodynamo. Paper [41] was popular, probably because it gave hope, though to what extent that hope is justified is somewhat uncertain: see Love (1999).

I returned to Britain in 1963 as Professor of Applied Mathematics at the University of Newcastle upon Tyne, and quickly acquired a talented Ph.D. student, Dennis Gibson now Vice-Chancellor of the University of Technology of Brisbane. Our first paper [47] included a simplified approach to the Herzenberg (1958) two-rotor dynamo, one that was used by Moffatt in his MHD book (1978) and also by Brandenburg *et al*. (1998). But Dennis' best use of this technique was to find a three-rotor dynamo paradoxically even simpler than Herzenberg's. Our second paper [56] made it seem unlikely that the Bullard-Gilman (1954) dynamo could maintain a magnetic field; others have since come to the same conclusion, including Pekeris *et al*. (1973). Later Subodh Kumar and I [98] created a successful dynamo that has since been very thoroughly studied by Gubbins *et al*. (2000). Recently, Dennis and I collaborated again in a paper on integral equations and MHD duct flow [218] that has since been included in a book [252].

Back in Britain, I maintained contact with Russell Donnelly, then a Professor in the Physics Department at Chicago University. He had already interested me in superfluid mechanics, or super fluid mechanics as I had begun to regard it. I was lucky enough to make a useful contribution early [58], when I isolated perhaps the main mechanism through which ions can be trapped by vortices, an idea that Donnelly (1965) found very exciting and developed. Russ also asked a very good question: Since the speed U of a quasi-particle is obtained from its Hamiltonian H by differentiation with respect to its momentum (impulse) p, i.e., $U = \partial H/\partial p$, why is this not true of a circular vortex ring moving in a classical inviscid fluid?

For example, for the "standard" sequence of uniform core models with circulation κ in a fluid of density ρ, one has $H \sim \frac{1}{2}\rho\kappa R[\ln(R/a) - \frac{7}{4}]$, $p \sim \pi\rho R^2$ and $U = (\kappa/4\pi R)[\ln(R/a) - \frac{1}{4}]$, asymptotically for large R/a. Because of its second term, the expansion apparently fails to obey $U = \partial H/\partial p$. I was very perplexed, since the inviscid fluid is a conservative system with its own Hamiltonian density, and it was even shown in Art. 167 of Lamb's *Hydrodynamics* (1945) that Hamilton's equations should be obeyed. But that failure to satisfy $U = \partial H/\partial p$ is only superficial, the result of holding a constant in the differentiation, rather than the net vorticity. Once this is recognized, all is well [62]. Russ also referred me to the winning Adam's prize essay of Thomson (1883), and I found that his theory of the binary collision of circular vortex rings did not obey Hamiltonian dynamics either! Closer scrutiny revealed an arithmetic error in his analysis ($26 + 8 = 32$) which, when corrected, satisfactorily brought his analysis up to Hamiltonian expectations. In this paper [85], I also gave a more general proof of the relation $U = \partial H/\partial p$ for an isolated ring, for any R/a all the way, in the case of the standard vortex sequence, to Hill's spherical vortex. An interesting aftermath of this work was the development of the idea of velicity by Butke (1993); see also Kuzmin (1983). This is called "impulse" by Russo and Smereka (1999).

Russ Donnelly also introduced me to the nonlinear Schrödinger equation (NLSE) as a model of superfluid helium at 0°K. This interesting approach sparked a sequence of contributions [68,92,141,153,251,255,261,262,265,266,267,268,269], one describing how boundary layer separation and drag arise in a superfluid, through the shedding of quantized vortex lines; see also [115,157]. But one wants to get away from 0°K. Landau published his celebrated two-fluid theory in 1941, but that was before quantized vorticity was known to exist, and therefore, naturally enough, it fails when superfluid vorticity is present. One of the ideas that set off a lifetime collaboration and friendship with Ron Hills was the thought that we should modify the Landau theory in a way that incorporated superfluid vorticity [110,117,125,130]; see also [81,111,123]. I believe our generalization is potentially useful in being applicable where Landau's theory is not. Later Seth Putterman and I developed a quasi-classical, one-fluid model of superfluidity [142,143] which achieved much the same objectives.

Russ also gave me the opportunity [60,67] of applying what I'd learned about stochastic processes to the problem of nucleation of vorticity in helium, and introduced me to other interesting topics [59,61,69,83,91,103,119,120,122,126]. He was also keen to develop a useful thermodynamics when energy levels are temperature dependent. In this context, he arranged that we should informally present our work [106,109] to another outstanding scientist, Richard Feynman, who was incisive but friendly in discussions, and totally unpretentious.

I maintained an interest in stability questions. In the early 1960s, it was often said that a flow would be more stable in an electrically conducting fluid if a magnetic field were present than it would be without it. I doubted this but (alas, another missed opportunity!), the system I chose to demonstrate this was unconvincing [33; see also 37]; meanwhile Kakutani (1964) was already doing better. Ibrahim Eltayeb and I analyzed a more successful example [65] from rotating magnetoconvection, showing how a Coriolis force could, by opposing the Lorentz force, reduces the critical Rayleigh number required for convection. Stewartson and I generalized to weakly nonlinear convection [95,99]. An amusing later twist showed that a magnetic field could permit convection to occur at negative Rayleigh numbers, i.e., in a bottom-heavy layer [121,127]. This created some surprise in a meeting at the Battelle Institute in Seattle, and one distinguished member of the audience told me I had been very naughty misleading the audience! Subsequently Soward (1979) created his own neat model

and Acheson (1980) provided the physical explanation. I continued to study thermal instability [23,39,42,45,52,53]. Ilya Prigogine was a frequent visitor to Chicago University. I met him during the time he and Glansdorff were developing their local potential method (see e.g., Glansdorff and Prigogine, 1971); he advocated applying this to the nonlinear Bénard problem, and I attempted [45] to so, even though I became a little unsure of the approach [44]. The happiest result of the entire episode was perhaps the attractive asymptotics problem solved by Keith Stewartson in the appendix to [45]. Keith transmitted to me his liking (but, alas, not his competence) in asymptotics; I in turn transmitted my liking to others, particularly to my students Paul Baldwin [63,80] and Weijia Kuang [183,192,200]. This was one area in which Chandra taught me nothing; he seemed to distrust asymptotics, and this led to some minor obscurities (see [237]).

I argued that the onset of convection in a rapidly rotating sphere containing heat sources would be in a cartridge belt of cells round the rotation axis. The credit for this observation usually goes to Busse (1970), largely on the basis, I feel, of the sketch he drew of the cartridge belt, which is often copied or redrawn. Never underestimate the power of graphics! (He also re-derived my equations for the cartridge belt by a better method, and selected the correct eigenfunction, i.e., the one with the smallest marginal Rayleigh number; I had accidently chosen the second smallest.) This problem was attacked several times subsequently before Jones *et al*. (2000) administered the *coup de grâce*.

Soon after I returned to Britain in 1963, I was lucky enough to meet Edward Bullard with the news that "some plasma physicist or other in Moscow" had come up with a new theory of the geomagnetic field, and that it "might be worth looking into." I read Braginsky's paper (1964a) with scepticism, turning to awe as I gradually realized that this strikingly original but massively technical *tour de force* was completely accurate and potentially useful. I set out with another student, Graham Tough, to generalize Braginsky's paper (Tough, 1967) and to add some dynamics [54].

While on sabbatical leave at the National Center for Atmospheric Research (NCAR) in Boulder, I met Michael Stix, who was avidly following the research of the Steenbeck–Krause–Rädler group in Potsdam. We decided that it would serve the dynamo community well if we presented translations of the bulk of their work [71]. We also devised new solar dynamo models [77], making use of programs that I had written for a different purpose, that of benchmarking Braginsky's (1964b) models and developing them [78]. While at NCAR, Bernard Durney introduced me to the solar wind [72,75], and this led to my first collaboration with Andrew Soward [76].

I determined to form links across the iron curtain with both Moscow and Potsdam. This led to collaborations with Fritz Krause [108,136] and to a controversy [88,89] that lurched on for some time, with interesting input from Andrew Soward [97,100]. Collaboration with Stanislav Braginsky was at first difficult [104] because of Soviet restrictions, but later led to integrations of his model—*Z* dynamo [166,179,216] and later still, after he had joined me in UCLA, to our monumental work on core MHD [221,270]. His (1963) paper got me interested in the dynamics of mixed phase regions, initially in application to the Earth's solid inner core, but afterwards in more general metallurgical contexts. Both Dave Loper and Ron Hills became deeply involved, and it was a wonderful experience. Papers [134,138] established, to me rather convincingly, that there is a region of mixed solid and liquid phases, generally called a "mush" or a "mushy layer," at the top of the inner core. We created theories for the motion and evolution of mixed phase regions [112,133,146,158,189,164,167, 172,175,181, 196,201,207,242]. I was particularly pleased with [185], which laid the ghost of superheated ice, and with [140], which provided a basis for the theoretical understanding

of the chimneys that form in a dendrite layer. This has been imitated by others (the sincerest form of flattery!) but is as nothing compared with the sequel [264]. During a short stay at Nordita in 1980, Chris Pethick got me slightly interested in neutron stars [135] and in type II superconductivity, about which I made a small but interesting discovery [137].

My extremely successful collaboration with Gary Glatzmaier started in 1993 with 2D models [208], and led to the first numerical simulations of the geodynamo that could be said to be geophysically faithful, being fully nonlinear and fully 3D [225,226,232,234,235,237, 238,241,246,253,256,258,263]. In particular, this was the first time that field reversals had been realistically simulated, as distinct from the earlier crude but suggestive 1D models; see e.g., [64,136]. Computer capabilities and availability had just reached the level where such large-scale simulations were possible, and our work was contemporaneous with the wonderful model of Kageyama *et al*. (1995) that was fully compressible (though too compressible to be a model of the Earth). Gary and I were able to achieve greater realism by making use of the anelastic theory developed by Stanislav and myself [221,270] for a realistic model of the core. Since that time there have been many Boussinesq models of the geodynamo, too many to refer to here.

After I moved to UCLA in 1986, Seth Putterman entered a new field, sonoluminescence (SL), in a big way, and his enthusiasm was infectious. I started a fruitful collaboration with Cheng-Chin Wu, who attempted to educate me in the theory and techniques of compressible flow and shocks. We produced the first complete shock wave simulation of SL [206,211,213] and proposed that the observed light was bremsstrahlung emission, an idea that is now currently favored. We derived new results concerning the stability of the imploding shock wave [231,233], and added something to the understanding of rectified diffusion, the process that determines the size of the oscillating bubble [243]. We also examined the linear stability of the bubble surface [248]; see also [244,249], What gave me most pleasure was the discovery [247] that something new and useful could be said in a field on which several eminent scientists had previously trodden: the initial-value problem for surface stability cannot be solved using the infinite, though incomplete, set of discrete eigenvectors.

I have aimed in this commentary to tell the reader something about the environment in which I have worked, the interesting scientists with whom I have interacted, the mistakes I have made and the opportunities I have missed, in short, a tale typical of the average scientist's life. My selection of papers favors those that (a) feature my collaborators, (b) are short, and (c) are likely to be lost without trace unless I make this effort to save them [115,124,185,190,201]. Concerning (b), it should be clear that I value highly some papers that were too lengthy to include.

Paul H. Roberts, Los Angeles

References

Acheson, D.J., 1980, "Stable density statification as a catalyst for instability," *J. Fluid Mech.*, **96**, 723–733.

Bertiau, F.C., 1977, *Programs for Pocket Calculators HP-67 and HP-97 in the Field of Theoretical and Observational Astronomy*, in association with E. Fierens, Leuven: University Press.

Bondi, H. and Lyttleton, R.A., 1953, "On the dynamical theory of the rotation of the Earth, II. The effect of precession on the motion of the liquid core," *Proc. Camb. Phil. Soc.*, **49**, 498–515.

Braginsky, S.I., 1963, "Structure of the F layer and reasons for convection in the Earth's core," *Soviet Phys. Dokl.*, **149**, 8–10 [*Dokl. Akad. Nauk. CCCP*, **149**, 1311–1314].

Braginsky, S.I., 1964a, "Self-excitation of a magnetic field during the motion of a highly conducting fluid," *Soviet Phys. JETP*, **20**, 726–735 (1965).

Braginsky, S.I., 1964b, "Kinematic models of the Earth's hydromagnetic dynamo," *Geomag. and Aeron.*, **4**, 572–583.
Brandenburg, A., Moss, D. and Soward, A.M., 1998, "New results for the Herzenberg dynamo: steady and oscillatory solutions" *Proc. R. Soc. Lond.*, **454**, 1283–1300.
Bullard, E.C. and Gilman, H., 1954, "Homogeneous dynamos and terrestrial magnetism," *Philos. Trans. R. Soc. Lond. A* **247**, 213–278.
Busse, F.H., 1970, "Thermal instabilities in rapidly rotating systems," *J. Fluid Mech.*, **44**, 441–460.
Butke, T.F., 1993, "Velicity methods: Lagrangian numerical methods which preserve the Hamiltonian structure for incompressible fluid flow," in *Vortex Flows and Related Numerical Methods*, eds. J.T. Beale, G.H. Cottet and S. Halberson. Dordrecht: Kluwer.
Chandrasekhar, S. 1943, "Stochastic problems in physics and astronomy," *Rev. Mod. Phys.*, **15**, 1–89. Reprinted in Wax (1954).
Chandrasekhar, S. 1961a, *Hydrodynamic and Hydromagnetic Stability*, Oxford: University Press.
Chandrasekhar, S. 1961b, "Adjoint differential systems in the theory of hydrodynamic stability," *J. Math. & Mech.*, **10**, 683–690.
Childress, S. and Gilbert, A.D., 1995, *Stretch, Twist, Fold: The fast Dynamo*, New York, Springer.
Donnelly, R.J., 1965, "Theory of the interaction of ions and quantized vortices in rotating helium II," *Phys. Rev. Letts.*, **14**, 39–41.
Gibson, R.D., 1968, "The Herzenberg dynamo, II," *Q. J. Mech. Appl. Math.*, **21**, 257–287.
Glansdorff, P. and Prigogine, I., 1971, *Thermodynamic Theory of Structure, Stability and Fluctuations*, London: Wiley-Interscience.
Grasman, J., 1971, *On the Birth of Boundary Layers*, Amsterdam: Mathematical Center.
Gubbins D., Barber, C.N., Gibbons, S. and Love, J.J., 2000, "Kinematic dynamo action in a sphere II. Symmetry selection," *Proc. R. Soc. Lond.*, *A***456**, 1669–1683.
Herzenberg, A., 1958, "Geomagnetic dynamos," *Phil. Trans. R. Soc. Lond.*, *A***250**, 543–585.
James, R.A., 1964. "The structure and stability of rotating gas masses," *Astrophys. J.*, **140**, 552–582.
Jones, C.A., Soward, A.M. and Mussa A.I., 2000, "The onset of thermal convection in a rapidly rotating sphere," *J. Fluid Mech.*, **405**, 157–179.
Kagan, Y.Y. and Knopoff, L., 1985, "The first-order statistical moment of the seismic moment tensor," *Geophys. J. Int.*, **81**, 429–444.
Kageyama, A., Sato, T., Watanabe, K., Horiuchi, R., Hayashi, T., Todo, Y., Watanabe, T.H. and Takamaru, H., 1995, "Computer simulation of a magnetohydrodynamic dynamo. II," *Phys. Plasmas*, **2**, 1421–1431.
Kakutani, T., 1964, "The: hydromagnetic stability of the modified plane Couette flow in the presence of a transverse magnetic field," *J. Phys. Soc. Japan*, **19**, 1041–1057.
Kuzmin. G.A., 1983, "Ideal incompressible hydrodynamics which preserve the Hamiltonian structure of incompressible fluid flow," *Phys. Lett.*, **96***A*, 88–90.
Lamb, H., 1945, *Hydrodynamics*, 6th edition. Cambridge: University Press. Reprinted by Dover Publications.
Landau, L.D., 1941, "The theory of superfluidity of helium II", *J. Phys. USSR*, **5**, 71–90.
Love, J.J., 1999, "A critique of frozen-flux inverse modelling of a nearly steady geodynamo," *Geophys. J. Intern.*, **138**, 353–365.
Lowes, F.J. and Runcorn, S.K., 1951, "The analysis of the geomagnetic secular variation," *Philos. Trans. R. Soc. Lond.*, *A***243**, 525–546.
Moffatt, H.K., 1978, *Magnetic Field Generation in Electrically Conducting Fluids*, Cambridge: University Press.
Parker, E.N., 1955, "Hydromagnetic dynamo models," *Astrophys. J.*, **122**, 293–314.
Pekeris, C.L., Accad, Y. and Shkoller, B., 1973, "Kinematic dynamos and the Earth's magnetic field," *Philos. Trans. R. Soc. Lond.*, *A***275**, 425–461.
Proctor, M.R.E., 1977, "The role of mean circulation in parity selection by planetary magnetic fields," *Geophys. Astrophys. Fluid Dynam.*, **8**, 311–324.
Runcorn, S.K., 1954, "The Earth's core," *Trans. Am. Geophys. Un.*, **35**, 49–63.

Russo, G. and Smereka, P., 1999, "Impulse formulation of the Euler equations; general properties and numerical methods," *J. Fluid Mech.*, **391**, 189–209.

Skiles, D.D., 1972, "The laws of reflection and refraction of incompressible magnetohydrodynamic waves at a fluid solid interface," *Phys. Earth Planet. Inter.*, **5**, 90–98.

Soward, A.M., 1979, "Thermal and magnetically driven convection in a rapidly rotating fluid layer," *J. Fluid Mech.*, **90**. 669–684.

Steenbeck, M. and Krause, F., 1966, "Erklärung stellarer und planetarer Magnetfelder durch einen turbulenzbedingten Dynamomechanismus," *Z. Naturforsch.*, **21a**, 1285–1296. Translation in [71].

Steenbeck, M., Krause, F. and Rädler, K.-H., 1966, "Berechnung der mittleren Lorentz-Feldstärke $\overline{\mathbf{v}\times\mathbf{B}}$ für ein elektrisch leitendes Medium in turbulenter, durch Coriolis-Kräfte beeinflußter Bewegung," *Z. Naturforsch.*, **21a**, 369–376. Translation in [71].

Thomson, J.J., 1883, *A Treatise on the Motion of Vortex Rings*, London: Macmillan.

Tough, G., 1967. "Nearly symmetric dynamos," *Geophys. J. R. Astr. Soc.*, **13**, 393–396; corrigendum *ibid*, **15**, 343 (1969).

Ursell, H.D., 1927, "The evaluation of Gibbs' phase-integral for imperfect gases," *Proc. Camb. Phil. Soc.*, **23**, 685–697.

Waechter, R.T., 1969, "Steady magnetohydrodynamic flow in an insilating circular pipe," *Mathematika*, **16**, 249–262.

Wax, N., 1954. *Noise and Stochastic Processes*, ed. N. Wax. New York: Dover Publications.

Wentzel, D.G., 1960, "Hydromagnetic equilibria," *Astrophys. J., Supp. Ser.*, **47**, 187–232.

Yoshizawa, A., 1990, "Self-consistent turbulent dynamo modeling of reversed field pinches and planetary magnetic fields," *Phys. Fluids*, **B2**, 1589–1600.

THE 1999 JOHN ADAM FLEMING MEDAL

Paul H. Roberts was awarded the 1999 John Adam Fleming Medal at the AGU Fall Meeting Honors Ceremony, which was held on December 15,1999, in San Francisco, California. The medal recognizes original research and technical leadership in geomagnetism, atmospheric electricity, aeronomy, and related sciences.

Citation

The spectacularly realistic simulations of magnetic field generation in the Earth's core since 1995 have received so much publicity that most geophysicists are well aware of the remarkable progress that has been made in the last few years. The dynamo theory, however, is not new but, rather, a long-standing problem in geophysics. The idea that the magnetic fields of the Earth and Sun arise from dynamos operating in their interiors was first put forward by Joseph Larmor in 1919, but Thomas Cowling ruled out two-dimensional dynamos by his famous theorem in 1933. The period when Paul Roberts was working for his Ph.D. in Cambridge coincides with the time when Walter Elsasser and especially Edward Bullard were making significant efforts on the homogeneous dynamo problem. Paul's advisor at Cambridge suggested that he might try to solve the dynamo problem.

Paul received his Ph.D. from Cambridge University in 1954; his dissertation was on theoretical geomagnetism, but there was nothing on the dynamo. He spent post-doctorate years at the University of Chicago and then returned to England and started to work at Newcastle upon Tyne on the problems of magnetohydrodynamics (MHD) of the Earth's core. His early results include a paper written with Stan Scot about the possibility of estimating core surface motion using secular variation data and the frozen flux approximation. In 1968, Paul wrote about the convection in a rapidly rotating fluid sphere. This and a later paper by Fritz Busse in 1970 showed that the motions would be predominantly two-dimensional and confined to a set of convection rolls aligned parallel to the rotational axis.

In 1975, with Subodh Kumar, Paul presented one of the first successful spherical kinematic dynamos. David Gubbins and Chaim Pekeris and his colleagues came up with other successful models at about the same time. Two other papers from 1972 (one with Michael Stix) became classics of the so called α-effect dynamo theory. Paul returned to MHD dynamos in the 1980s and worked with Gary Glatzmaier and Stanislav Braginsky. The papers with Stanislav on so-called model-Z dynamos demonstrated the subtle balance between Coriolis, Lorentz, and buoyancy forces in the core.

The recent developments of the fully three-dimensional MHD models have been really remarkable. Gary and Paul presented their first geodynamo models in 1995, at about the same time as Akira Kageyama, Tetsuya Sato, and their colleagues reported on their successful stellar model. Gary and Paul simplified the problem by using the Boussinesq approximation and obtained a spectacular demonstration of a geomagnetic polarity reversal. They also claimed that the inner core is rotating faster than the mantle by a few degrees per year. The seismologists rose to the challenge, and it is now an exciting if controversial branch of their subject. Their work was followed by several more Boussinesq models by other groups: Jeremy Bloxham and Weijia Kuang; Peter Olson, Uli Christensen, and Gary Glatzmaier; Shigeo Kida and Hideaki Kitauchi; and Ataru Sakuraba and myself.

In 1996, Gary and Paul decided to move toward more realistic models that include both the compressibility of the Earth and the fact that the fluid core is cooled from the top but freezes from the bottom. The necessary theoretical apparatus was at hand. In 1995, Stanislav Braginsky and Paul published the most thorough investigation ever undertaken of core MHD. Gary and Paul made this the basis of the more realistic model that is now central to their work. They have recently, with Robert Coe and Lionel Hongre, shown how the statistics and reversal characteristics of the geomagnetic field depend on the assumed pattern of heat flow from core to mantle.

There are about 250 papers listed in Paul's publication list, and they reveal his broad range of interests, from sonoluminescence to superfluidity. These are remote from geomagnetism, and it would be inappropriate for me to dwell on them here. Instead, I would like to draw

ACKNOWLEDGEMENTS

We are grateful to the following editors and publishers for their permission to reproduce, without charge, previously published material which was co-authored by Paul Roberts.

Citation and response reproduced with permission of AGU; EOS, Vol 81, No 7, p66, 1999, copyright by the American Geophysical Union.

Reproduced with permission from Springer-Verlag, *Extrait des Annales de Geophysique*, Vol 15, No 1, pp. 75–86. Copyright 1959.

Reprinted with the permission of The Microbiology Research Foundation. Copyright 1965.

Reprinted from *Magnetism and the Cosmos*, 1996, pp. l08–120 with kind permission from Pearson Education.

Reprinted from *Physics of the Earth and Planetar Interiors*, Vol 1, pp. 288–296. Copyright 1968 with permission from Elsevier Science.

Reprinted from *Physics of the Earth and Planetary Interiors*, Vol 8, pp. 389–390. Copyright 1974 with permission from Elsevier Science.

Printed with permission of The Royal Society. Copyright 1975.

Reprinted from *Physics of the Earth and Planetary Interiors*, Vol 20, pp. 192–193. Copyright 1979 with permission from Elsevier Science.

Reprinted from *Physics of the Earth and Planetary Interiors*, Vol 24, pp. 302–307. Copyright 1981 with permission from Elsevier Science.

Reprinted by permission from *Nature*, Vol. 292, pp. 232–233. Copyright 1981, Macmillan Magazines Ltd.

Reprinted from *Physics of the Earth and Planetary Interiors*, Vol 36, pp. 49–60. Copyright 1984 with permission from Elsevier Science.

Copyright 1988. Gordon & Breach.

Reprinted from *Physics of the Earth and Planetary Interiors*, Vol 91, pp. 63–75. Copyright 1995 with permission from Elsevier Science.

Reprinted by permission from *Nature*, Vol. 377, No. 6546, pp. 232–233. Copyright 1981, Macmillan Magazines Ltd.

Reprinted from *Physica D*, Vol 97, Glatzmaier and Roberts' An anelastic evolutionary geodynamo simulation driven by compositional and thermal convection, pp. 81–94. Copyright 1997 with permission from Elsevier Science.

Reprinted with permission from Science. Copyright 1996, American Association for the Advancement of Science.

Reproduced with permission of the *Journal of Mathematical Analysis and Applications*. Copyright 1960.

Reprinted with permission of Cambridge University Press.

your attention to two salient features of the list. The first is that Paul is asked to write a review of dynamo theory just about every year! In my opinion, this reflects how people rely on his judgment in evaluating the progress made in the subject. Second, the list of Paul's coauthors is truly impressive, showing both that he has always been in the mainstream of scientific developments and that his work has had a strong influence on other people.

"In summary, Paul has made an extraordinary contribution to geophysics, especially to the dynamo theory. He is thus a most fitting recipient of the John Adam Fleming Medal of the American Geophysical Union. It gives me great pleasure to introduce Paul Roberts to you."—*Masaru Kono, Institute for Study of the Earth's Interior, Okayama University, Tottori-Ken, Japan*

Response

"I am grateful to Masaru Kono for his kind words and for initiating the proposal that led to the honor that the American Geophysical Union has conferred upon me. I would also like to thank others who supported his proposal. I can fairly claim to be one of the doyens of geodynamo theory, one of the old fogies of the subject. It all started for me at the outset of my research career when I asked my advisor in Cambridge to suggest suitable thesis topics and he proposed that I should either prove that fluid dynamos could not exist or find a working model. This was a daunting prospect, indeed, for a starting post-graduate student! After a year, I switched topics and advisors; it would be another six years before two fluid dynamos were independently devised, albeit ones that were geophysically unrealistic. During my year of failure I learned a lot about magnetohydrodynamics—then a young subject—and was able to demonstrate that the sources of the geomagnetic secular variation did not have to be as was commonly supposed at the time within 100 km of the core surface a result of erroneously applying the electrodynamics of solid conductors to fluids. With hindsight it seems so obvious, but at the time it did not. My new advisor, Keith Runcorn, was so astonished when I told him that he practically fell off his chair. The insight helped me to my Ph.D. and I was able to accept the invitation of Subrahmanyan Chandrasekhar to become his postdoc at the Yerkes Observatory, where I learned a great deal before having to return to the U.K. to fulfill my national service obligations.

My interest in the geodynamo was rekindled in 1965 when news reached the west of the powerful advances being made by Stanislav Braginsky. It was an exciting time for the subject. Steenbeck, Krause, and Radler were revolutionizing ideas about solar, stellar, and galactic magnetism, and I began to develop a passion for the subject and, as Hegel wrote, 'Nothing great in this world has been accomplished without passion.' Passion, then, is a necessary state, as many of us know well, and it certainly made me an expert in what everyone else was doing and, very occasionally, like Robert Burton's 'dwarf standing on the shoulders of a giant,' I could see something a giant could not.

My move to UCLA in 1986 gave me the privilege and pleasure (which I treasure beyond measure) of working with Stanislav Braginsky and Gary Glatzmaier. I am grateful to them and to many colleagues and students (too many to name here) for contributing to my success as a geophysicist. The geodynamo has come remarkably far since I first encountered it almost 50 years ago, although there is (as I hope the funding agencies have noticed) still much to be done. To compare the subject today with where it stood nearly half a century ago is like comparing the Kitty Hawk with a Boeing 747. And, like the Boeing, there is room for many passengers, though what we need is not passengers but, rather, more hustlers and zealots to propel the subject onward. To them, 'welcome aboard!'"

—*Paul H. Roberts, University of California, Los Angeles, USA*

Copyright 1963. University of Chicago.
Reprinted from *Proc. R. Soc. Lond.*, ***300***, pp. 94–107. Copyright 1967, The Royal Society.
Reproduced with permission from *Mathematika*. Copyright 1970.
Copyright 1970. The University of Chicago.
Copyright 1971. Gordon & Breach.
Reproduced with permission from *Mathematika*. Copyright 1972.
Reprinted with permission of Wiley-VCH Verlag. Copyright 1975.
Reproduced with permission of the University of Washington. *Geofluiddynamical Wave Mechanics Researched*, ed. W. Criminale. Publ. of Appl. Math. Dept. Univ. of Wash. Seattle, pp. 218–225. Copyright 1977.
Reproduced with permission from SAACM. Copyright 1991.
Copyright 1992 by the American Physical Society.
Copyright 1993. Gordon & Breach.
Reprinted from *Physics Letters A,* Vol 213, pp. 59–64. Copyright 1996 with permission from Elsevier Science.
Reprinted with permission of the American Institute of Physics. Copyright 1997.
Reprinted with permission of the American Institute of Physics. Copyright 1998.
Reprinted from *Physics of the Earth and Planetary Interiors*, Vol 30 A, pp. 468–69. Copyright 1969 with permission from Elsevier Science.
Reprinted with permission from APS: R.J. Donnelly, P.H. Roberts, *Physics Review Letters*, 23, 1969, pp. 1491–1495. Copyright 1969 by the American Physical Society.
Reprinted with permission from APS: P.H. Roberts and R.J. Donnelly, *Physics Review Letters*, 1970, pp. 367–371. Copyright 1970 by the American Physical Society.
Reprinted with permission from APS: D.M. Strayer, R.J. Donnelly and P.H. Roberts, *Physics Reviews Letters*, 1979, pp. 725–728. Copyright 1971 by the American Physical Society.
Reprinted from *Physics Letters*, Vol. 43A, pp. 1–2. Copyright 1973 with permission from Elsevier Science.
Reprinted from *Physics Letters*, Vol. 55a, pp. 443–445, 1976. Copyright 1976 with permission from Elsevier Science.
Provided through the courtesy of the Jet Propulsion Laboratory, California; Institute of Technology, Pasadena, California (Proc. Int. Coll. on Drops and Bubbles. Their science and the systems they model. I JPL Publication 1974).
Reprinted from *Physics Letters*, Vol. 70A, pp. 437–440. Copyright 1979 with permission from Elsevier Science.
Reprinted with permission from APS: R.J. Donnelly, P.H. Roberts, *Physics Review Letters*, 1979, pp. 725–728. Copyright 1979 by the American Physical Society.
Reprinted from *Physics Letters*, Vol 89A, pp. 444–447. Copyright 1982 with permission from Elsevier Science.
Reprinted from 75th Jubilee Conference on Helium-4, pp. 113–114, World Scientific Press. Copyright 1983.
Reprinted with permission from SAACM. Copyright 1991.
Reprinted with permission from SAACM. Copyright 1992.

Part 1

GEOMAGNETISM AND DYNAMOS

Extrait des ANNALES DE GÉOPHYSIQUE
Tome 15, N° 1. — Janvier-Mars 1959.

PROPAGATION OF INDUCED FIELDS THROUGH THE CORE

by P. H. ROBERTS
(University of Durham, King's College, Newcastle upon Tyne)

RÉSUMÉ. — *On étudie la propagation hydromagnétique de petites perturbations magnétiques à partir d'une source située dans un fluide de conductivité finie jusqu'à la frontière de celui-ci. On applique les résultats au problème du taux de croissance du champ magnétique dû aux centres de variation séculaire.*

ABSTRACT. — *A study is made of the hydromagnetic propagation of small magnetic disturbances from a source within a finitely conducting fluid to its boundary. The results are applied to the problem of the rate of growth of the magnetic field due to the centres of secular change.*

1. INTRODUCTION.

The first applications of electrodynamics to astrophysical problems ignored the fluid nature of cosmical matter and simply carried over the notions of electromagnetic theory as applied to rigid conductors which were the result of rather extensive developments by electrical engineers. For example, the longest period associated with the decay currents in a rigid sphere of radius R cm, conductivity σ e. m. u. is (LAMB [1])

$$\tau = \frac{4\sigma R^2}{\pi} \text{ sec.}$$

This led astrophysicists to believe that this decay time of 10^{10} years for the Sun implied that no mechanism was required to maintain a solar magnetic field against the effects of free decay. However, since the inception of hydromagnetics, the study of the flow of conducting fluids in the presence of magnetic fields has made it evident that the transmission of energy by ALFVÉN waves which arises in fluids is much more significant than the transmission by electromagnetic diffusion which arises in both fluid and solid conductors. Thus the decay time based on the assumption of a rigid conductor can be, in some cases, almost meaningless.

The work described in this paper was commenced in 1953 at a time when some of the results of LOWES and HERZENBERG [2] presented an interesting dilemma. The investigations of these authors had then established that, if an eddy model were advanced to account for the non-dipole field, the eddies themselves must be required to possess radii of several hundred kilometres. The similar regional character of non-dipole and secular variation fields strongly suggested that the secular variation is generated by the changing rotation of these eddies

However, the observed rate of growth of the secular variation field is so rapid that, assuming any reasonable value for the electrical conductivity of the core, the skin-depth d_0 is but a small fraction of the eddies' radii. In other words, the final magnitude and character of the non-dipole field could be understood on the eddy hypothesis but the singularly rapid rise of the secular variation field could not.

It occurred to the present author that there was one serious flaw in this argument. The skin-depth d_0 had been calculated on the assumption that the core was solid and therefore that disturbances in it were propagated by (electromagnetic) diffusion only. However, the core is fluid and can therefore transmit disturbances as (hydromagnetic) waves. Now, disturbances are propagated to large distances much more effectively by waves than by diffusion and, indeed, for the time-scale of the secular variation, the hydromagnetic skin-depth $\mathcal{D}$ greatly exceeds the diameter of the core itself (see § 2 below). This seemed a promising start to the removal of the dilemma posed by the work of LOWES and HERZENBERG. However, it remained to answer certain possible objections and it is the purpose of this paper to do so. These require the resolution of the following difficulties :

(*a*) Although it is not unlikely that the disturbances are propagated *along* the excitation field in the manner of hydromagnetic waves, it may be expected that in perpendicular directions they can be transmitted by a diffusion mechanism only. It is thought that the toroidal field is much greater than the poloidal field in the core. The lines of magnetic force are therefore almost tangential to its boundary. Thus the disturbances could apparently only reach the mantle by travelling across the lines of force by the diffusion mechanism and that the dilemma has not, after all, been resolved. This point is discussed in § 2.

(*b*) The hydromagnetic skin depth argument of § 2 concerns a simple disturbance in which the lines of force are distorted but not altered in number. If, however, we consider disturbances which generate lines of force, it is clear that the greater the conductivity, the more the fluid clings to its old lines of force and the more difficult it finds it to accept the new ones. This suggests that it could accept the new lines of force effectively only within a solid conductor skin depth d_e from the source of disturbance (see paper [2], p. 543). If this is so, the dilemma might still have to be faced for this particular class of disturbances. This question is discussed in § 3.

(*c*) There are similar difficulties concerning the transmission of lines of force from the core into the mantle. The most serious of these concerns the case in which lines of force of the undisturbed field lie tangentially to the boundary separating a conducting fluid from an insulator. If the conductivity of the fluid is infinite, no disturbance can separate the lines of force from the fluid particles to which they are attached and none can be pushed into the insulator. It is therefore

clear that the transmission of lines of force through the boundary is in this case essentially a finite conductivity effect, apparently unchanged by the presence of wave motion. If, as in the Earth, the field has a component across the boundary, the difficulty is not so great since, even in the case of infinite conductivity, the disturbance can appear in the insulator by the redistribution of lines of force by the fluid motions. These problems are investigated in § 4. (The boundary is, for simplicity, assumed to be a plane).

The mathematical basis of the following discussion will be largely omitted since the interested reader will be able to find it at length in two papers already published [3,4] and in another forthcoming, [5].

II. Basic equations and some simple solutions.

We consider that there is initially a uniform prevailing field $\mathbf{H}_0$ (e. m. u.) in the z-direction. We study small perturbations $\boldsymbol{h}$ to this field. It is found convenient to measure length and time in units of

$$l = \frac{V}{\omega_0} \qquad \tau = \frac{1}{\omega_0}$$

respectively,
where V is the Alfvén velocity

$$V = \frac{H_0}{\sqrt{(4\pi\rho)}}$$

and ω_0 is a frequency defined by

$$\omega_0 = \frac{\sigma H_0^2}{\rho}$$

(ρ is the density of the fluid and σ its electrical conductivity in e. m. u.). If we take the geophysically plausible values

$$(1)\quad \begin{cases} \rho = 10 \text{ gm.cm}^{-3} \qquad \sigma = 3{,}000 \text{ mho.cm}^{-1} \qquad H_0 = 4 \text{ gauss (considering only the poloidal field)} \\ \text{we find} \quad l = 0.7 \text{ km} \qquad \tau = 2.4 \text{ days} \qquad \text{approximately} \end{cases}$$

If we measure $\boldsymbol{h}$ in units of H_0 and the fluid velocity $\boldsymbol{u}$ in units of V, we are led to the following basic equations

$$(2)\quad \begin{cases} \left(\dfrac{\partial}{\partial t} - \nabla^2\right)\boldsymbol{h} = \dfrac{\partial \boldsymbol{u}}{\partial z} \\ \dfrac{\partial \boldsymbol{u}}{\partial t} = \dfrac{\partial \boldsymbol{h}}{\partial z} - \text{grad } \tilde{\omega} \\ \text{div } \boldsymbol{h} = 0 \qquad \text{div } \boldsymbol{u} = 0 \end{cases}$$

where $\tilde{\omega}$ is related to the total (hydrostatic and magnetic) pressure in the fluid. If the fluid is unbounded and the field quantities have no singularities, we have

$$\nabla^2 \tilde{\omega} = 0$$

which implies that ω is constant everywhere. Thus both $\boldsymbol{u}$ and $\boldsymbol{h}$ satisfy the equations

$$\frac{\partial}{\partial t}\left[\frac{\partial}{\partial t} - \nabla^2\right] \boldsymbol{a} = \frac{\partial^2 \boldsymbol{a}}{\partial z^2} \qquad \operatorname{div} \boldsymbol{a} = 0$$

Exhibited in this form, the phenomenon clearly shows its combination of wave-like and diffusion-like properties (see paper [3]).

We can now consider the simple skin-depth argument referred to in the introduction. Suppose that, *on the plane* z = 0, the disturbance is specified as

$$\boldsymbol{a} = \boldsymbol{a}_0 e^{-i\omega t} \qquad (\boldsymbol{a}_0 \text{ constant, } a_{0z} \text{ zero, } \omega \text{ real})$$

we find that, at a general point of space and time, $\boldsymbol{a}$ is given by

$$\boldsymbol{a} = \boldsymbol{a}_0 \exp\left\{-\frac{|z|}{\mathcal{D}} + i(\alpha_1 |z| - \omega t)\right\} \qquad (\alpha_1 \text{ real})$$

where $\mathcal{D}$, the hydromagnetic skin-depth, is given by

$$\mathcal{D} = d_0 \left[\frac{1+\omega^2}{\omega^2}\right]^{1/4} \left[1 - \frac{1}{\sqrt{(1+\omega^2)}}\right]^{-1/2}$$

where d_0 is the skin depth for a *solid* conductor (in the present units $d_0^2 = 2/\omega$). The hydromagnetic skin depth $\mathcal{D}$ exceeds d_0 for all frequencies and in geophysical applications does so by an extremely large factor. For example, for frequencies of 2π/(one year), d_0 is barely 5 km but $\mathcal{D}$ is nearly 900 km. using the values (1), and becomes even larger for larger values for H_0 (1). However, the propagation in perpendicular directions is very different. Suppose that, *on the plane* x = 0, the disturbance is specified as

$$\boldsymbol{a} = \boldsymbol{a}_0 e^{-i\omega t} \qquad (\boldsymbol{a}_0 \text{ constant, } a_{0x} \text{ zero, } \omega \text{ real})$$

we find that $\boldsymbol{a}$ is given generally by the form

$$\boldsymbol{a} = \boldsymbol{a}_0 \exp\left\{-\frac{|x|}{d_0} + i(\alpha_2 |x| - \omega t)\right\} \qquad (\alpha_2 \text{ real}).$$

Thus, perpendicular to the excitation field $\mathbf{H}_0$, the attenuation is much greater than along that field and is, in fact, just the same as for a solid conductor.

We now discuss the point raised in (*a*) of § 1. Suppose the poloidal field at the core boundary is 4 gauss and that the tangential field (poloidal and toroidal), averaged over the top 100 km. of core, is 40 gauss. The lines of $\mathbf{H}_0$ at the surface of the fluid makes an angle of $\tan^{-1} 0.1$ with the surface. Consider a disturbance centred at a depth of 100 km. Two mechanisms help to transmit this disturbance to the surface of the core : wave motion along the lines of force and diffusion

(1) The frequency of 2π / (one year) is an overestimate, the field H_0 of 4 gauss is an underestimate, of typical values appropriate to the core. Thus, the conclusions drawn underestimate hydromagnetic effects.

perpendicular to them. The former is incomparably more effective even though the wave must travel ove 1,000 km. along the lines of force before reaching the surface. This is because the hydromagnetic skin depth $\mathcal{D}$ for the total field of over 40 gauss and frequency of 2π/(one year) is nearly 10^6 kilometres. Thus although the distance the wave must travel to the surface increases by a factor of over 10 the distance it can effectively travel increases by a factor of over 1 000 (Note : Once H_0 exceeds a certain limit, of the order of $(\rho\,\omega/\sigma)^{1/2}$, about 0,7 gauss for this frequency, the skin depth $\mathcal{D}$ is proportional to H_0^3).

All these considerations apply equally for aperiodic disturbances. The author has shown [3] that if at $t = 0$ the field

$$\boldsymbol{a} = \boldsymbol{a}_0 \qquad (\boldsymbol{a}_0 \text{ constant, } \boldsymbol{a}_{0z} \text{ zero})$$

is suddenly applied *on the plane* z $= 0$, $\boldsymbol{a}$ is given generally by

$$\boldsymbol{a} = \boldsymbol{a}_0 f(z, t)$$

where

$$(3) \quad f(z, t) = \int_0^t dt \frac{e^{-t}}{2\sqrt{(\pi t^3)}} \int_0^\infty \left[\frac{(\xi + z)^2}{2t} - 1\right] \exp\left[-\frac{(\xi + z)^2}{4t}\right] I_0(2\sqrt{(\xi z)})\, d\xi.$$

We return to this function in § 4 ; it is important to note that it approaches unity as t becomes large and that it attains a high percentage of this value at the point z by a time $t = z$; that is, when the hydromagnetic wave reaches this point.

III. Disturbances which generate lines of force.

As mentioned in the introduction, the question of how rapidly a source submerged in a fluid can produce lines of force must be faced. Suppose an unbounded fluid, initially undisturbed, in which a distribution of current $\mathbf{J}$ is suddenly introduced at time $t = 0$; this is mathematically the simplest case, but there is no difficulty, in principle, in postulating instead a rotating eddy whose interaction with $\mathbf{H}_0$ provides the source of the disturbance. The problem is most conveniently discussed by introducing $\mathbf{J}$ into equations (2) as extra terms thus :

$$(4) \quad \begin{cases} \left(\dfrac{\partial}{\partial t} - \nabla^2\right) \boldsymbol{h} = \dfrac{\partial \boldsymbol{u}}{\partial z} + \operatorname{curl} \mathbf{J} \\ \dfrac{\partial \boldsymbol{u}}{\partial t} = \dfrac{\partial \boldsymbol{h}}{\partial z} + \mathbf{1}_z \times \mathbf{J} - \operatorname{grad} \tilde{\omega} \\ \operatorname{div} \mathbf{J} = 0 \qquad \operatorname{div} \boldsymbol{h} = 0 \qquad \operatorname{div} \boldsymbol{u} = 0 \end{cases}$$

where $\mathbf{1}_z$ is a unit vector in the direction of z.

We then find that $\boldsymbol{h}$, $\boldsymbol{u}$, $\tilde{\omega}$ can be written as

$$(5)\qquad \begin{cases} \boldsymbol{h}(\boldsymbol{r},t) = \dfrac{1}{4\pi}\,\mathrm{curl}\left\{\displaystyle\int \frac{\mathbf{J}(\boldsymbol{r}')\,d\boldsymbol{r}'}{|\boldsymbol{r}-\boldsymbol{r}'|} + \frac{\partial}{\partial l}\int \mathbf{J}(\boldsymbol{r}')\,\mathrm{G}(\boldsymbol{r}-\boldsymbol{r}',t)\,d\boldsymbol{r}'\right\} \\ \boldsymbol{u}(\boldsymbol{r},t) = \dfrac{1}{4\pi}\,\mathrm{curl}\left\{\dfrac{\partial}{\partial z}\displaystyle\int \mathbf{J}(\boldsymbol{r}')\,\mathrm{G}(\boldsymbol{r}-\boldsymbol{r}',t)\,d\boldsymbol{r}'\right\} \\ \tilde{\omega}(\boldsymbol{r},t) = \dfrac{1}{4\pi}\,\mathrm{div}\displaystyle\int \frac{(\mathbf{1}_z\times\mathbf{J})'}{|\boldsymbol{r}-\boldsymbol{r}'|}\,d\boldsymbol{r}' \end{cases}$$

where the integrations are conducted round the distribution of current $\mathbf{J}$ and the Green's function $\mathrm{G}(\boldsymbol{r}, t)$ is given by

$$(6)\qquad \mathrm{G}(\boldsymbol{r},t) = -\frac{1}{2\pi^2}\iiint \frac{(e^{p_1 t}-e^{p_2 t})}{(p_1^2-p_2^2)}\exp\left[-i(lx+my+nz)\right]dl\,dm\,dn$$

and

$$(7)\qquad p_1, p_2 = -\frac{1}{2}(l^2+m^2+n^2) \pm i\sqrt{\left[n^2-\frac{1}{4}(l^2+m^2+n^2)^2\right]}.$$

The first term in the expression (5) for $\boldsymbol{h}$ is the familiar Biot-Savart law for the steady state field set up by an arbitrary current distribution and it is the form to which $\boldsymbol{h}$ tends for very large times.

It is convenient at this stage to introduce an approximation due to WALÉN [6]. According to this, it is adequate to replace equation (7) by

$$(8)\qquad p_1, p_2 = -\frac{1}{2}(l^2+m^2+n^2) \pm in.$$

It may be shown that this approximation asymptotically produces the correct form for $\mathrm{G}(\boldsymbol{r}, t)$ as t approaches infinity. Its use enables $\mathrm{G}(\boldsymbol{r}, t)$ to be expressed in closed form

$$\mathrm{G}(\boldsymbol{r},t) = -\int_t^\infty \frac{1}{4\mathrm{T}}\exp\left[-\frac{(x^2+y^2)}{2\mathrm{T}}\right]\left(\mathrm{erf}\,\frac{t-z}{\sqrt{(2\mathrm{T})}}+\mathrm{erf}\,\frac{t+z}{\sqrt{(2\mathrm{T})}}\right)d\mathrm{T}$$

so that, for example, the solutions corresponding to a dipole $\mathbf{M}$ placed at the origin are

$$\boldsymbol{h}(\boldsymbol{r},t) = \mathrm{curl}^2\,\mathbf{M}\left\{\frac{1}{\mathrm{R}}+\frac{1}{4t}e^{-\frac{1}{2t}(x^2+y^2)}\left(\mathrm{erf}\,\frac{t-z}{\sqrt{(2t)}}+\mathrm{erf}\,\frac{t+z}{\sqrt{(2t)}}\right)-\frac{1}{2\mathrm{R}_+}\mathrm{erf}\,\frac{\mathrm{R}_+}{\sqrt{(2t)}}-\frac{1}{2\mathrm{R}_-}\mathrm{erf}\,\frac{\mathrm{R}_-}{\sqrt{(2t)}}\right\}$$

$$\boldsymbol{u}(\boldsymbol{r},t) = \mathrm{curl}^2\,\mathbf{M}\left\{\frac{1}{2\mathrm{R}_+}\mathrm{erf}\,\frac{\mathrm{R}_+}{\sqrt{(2t)}}-\frac{1}{2\mathrm{R}_-}\mathrm{erf}\,\frac{\mathrm{R}_-}{\sqrt{(2t)}}\right\}$$

where

$$\mathrm{R} = \sqrt{[x^2+y^2+z^2]}$$
$$\mathrm{R}_+ = \sqrt{[x^2+y^2+(z+t)^2]}$$
$$\mathrm{R}_- = \sqrt{[x^2+y^2+(z+t)^2]}.$$

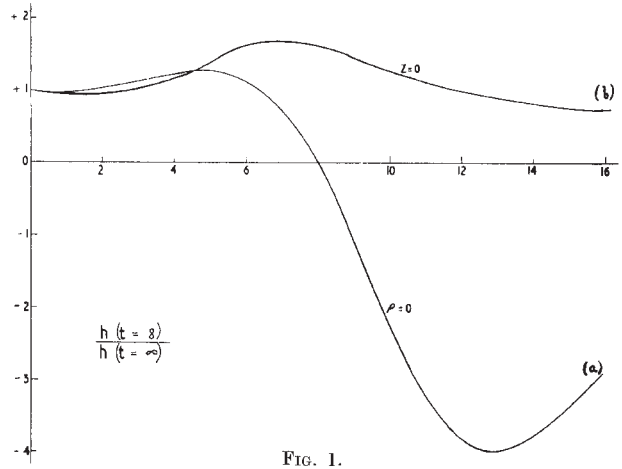


FIG. 1.

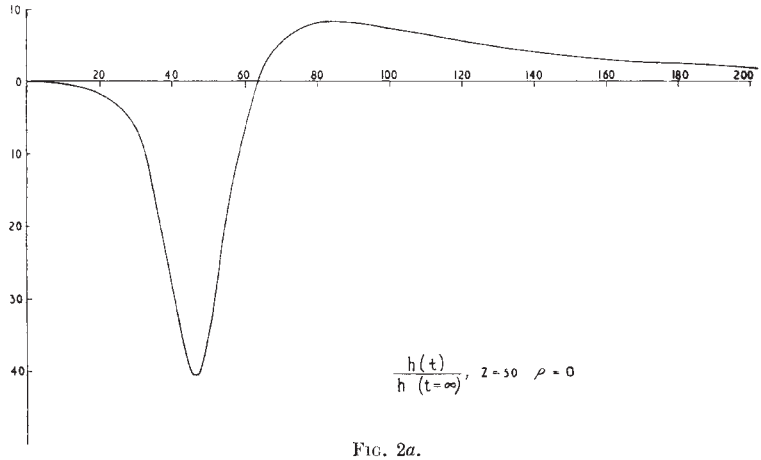


FIG. 2*a*.

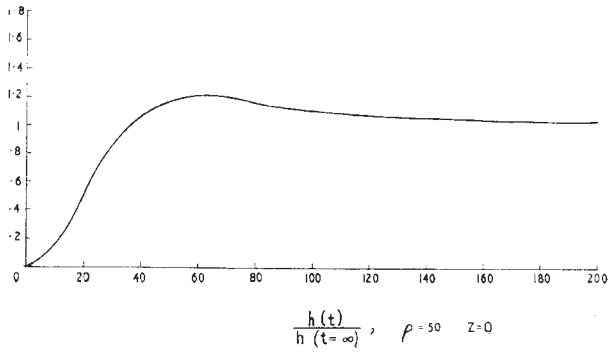


FIG. 2*b*.

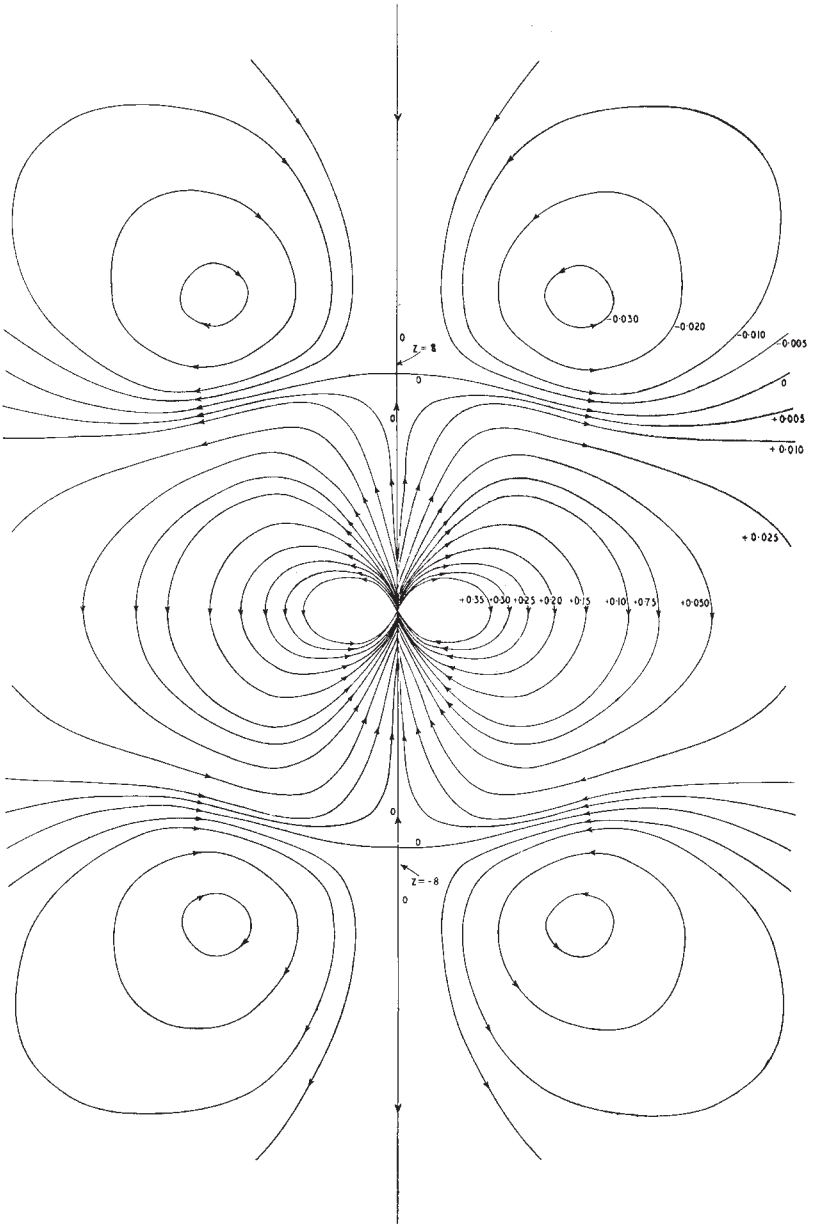

-0·030
-0·020
-0·010
-0·005
0
+0·005
+0·010
+0·025
+0·035
+0·030
+0·025
+0·020
+0·015
+0·010
+0·075
+0·050
z = 8
z = -8
0

FIG. 3a_8

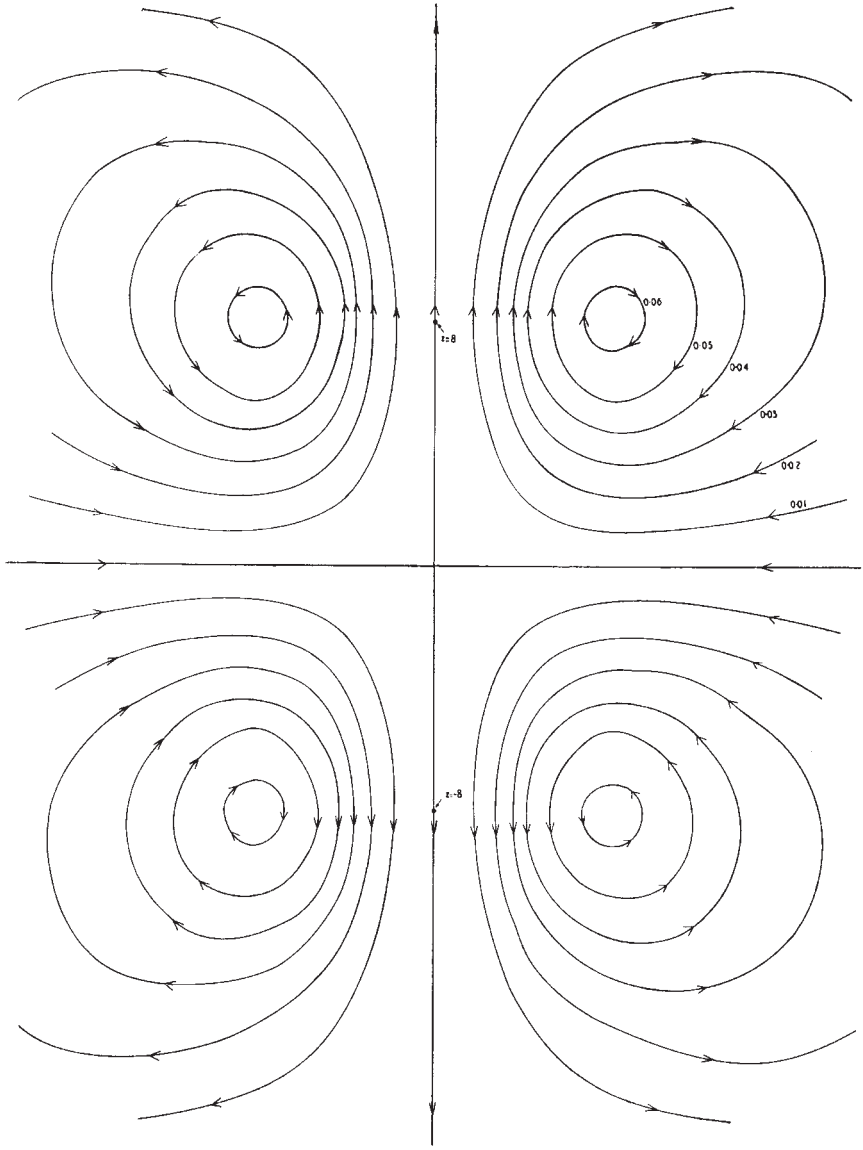


FIG. 3*b*.

In the case of a dipole pointing along the z-axis, figures 1 show the ratio

$$\frac{h_z(t = 8)}{h_z(t = \infty)}$$

(a) for zero x, y plotted as a function of z and (b) for zero z plotted as a function of $\rho\left(= \sqrt{(x^2 + y^2)}\right)$. Figures 2 exhibit the ratio

$$\frac{h_z(t)}{h_z(\infty)}$$

plotted as a function of t, (a) for the point (0,0,50) and (b) for the points $\rho = 50$, $z = 0$. Figures 3 depict (a) the lines of magnetic force and (b) the streamlines of the fluid motion at the instant $t = 8$. The striking feature of these diagrams is not so much the rapid propagation of the disturbance along the z-axis (for this was not unexpected in view of the earlier discussion) but the fact that the disturbance travels nearly as effectively in perpendicular directions. We may explain this in the following way. The dipole may be thought of as a current loop which, until it has acquired its quota of lines of force threading itself, draws in the lines of force from its neighbourhood and at the same time carries the fluid that is loosely coupled to these lines with them. Once the lines of force have crossed into the loop, they are held there while the fluid particles attached to them are first driven along the z-axis and afterwards pushed away. The fluid velocities in these returning streams opposes the redistribution of the lines of force at the current loop but are so small that their drag on the lines of force is not important. The argument also shows that the rapid increase in the field on the axis of the loop must be accompanied by an almost equally rapid change in directions from which the lines of force are drawn.

IV. Boundary effects.

We now briefly discuss progress so far made in the investigation of the transmission of disturbances from the dipole (M_x, M_y, M_z) in the core to the mantle. Due to the linear nature of the problem, we may consider two cases separately :

(a) $\mathbf{H}_0$ is perpendicular to the boundary separating core and mantle and

(b) $\mathbf{H}_0$ is parallel to the boundary.

The continuity conditions that must be applied at the (rigid) interface S between fluid and insulator may be reduced to five independent ones :

(1, 2) tangential components of the electric field $\mathbf{E}$ are continuous

(3, 4) tangential components of $\boldsymbol{h}$ are continuous

(5) the normal component of $\boldsymbol{u}$ is zero.

If we are content to leave the electric field incompletely specified, we may determine $\boldsymbol{h}$ and $\boldsymbol{u}$ uniquely by the following equivalent conditions :

(i) h_n, the normal component of $\boldsymbol{h}$, is continuous (from (1) and (2))

(ii) $\partial h_n/\partial n$ is continuous (from (3), (4) and equation of continuity of $\boldsymbol{h}$)

(iii) j_n, the normal component of current, is zero (from (3) and (4))

(iv) u_n, the normal component of $\boldsymbol{u}$, is zero. (see (5))

In the case (*a*), all the conditions with the exception of (i) can be satisfied without the introduction of an external field by supposing a reversed and reflected image dipole ($-\mathrm{M}_x$, $-\mathrm{M}_y$, M_z) at the image point. To satisfy condition (i) we must introduce an external field and an internal hydromagnetic disturbance without upsetting the condition (ii) — (iv) already satisfied. Fortunately this is possible. Details will appear in paper [5]. It is sufficient to give here the field at the point P where the axis of a vertical dipole meets the core boundary. We find

$$\bar{h}_z(\mathrm{P}) = 2\mathrm{M}\int_0^\infty \frac{(p^2 e^{-\mathrm{Q}d} - \xi \mathrm{Q} e^{-\xi d})}{p^2\mathrm{Q} + p^2\xi - 2\xi^2\mathrm{Q}} \frac{\xi^3}{\xi}\,\mathrm{d}\xi \tag{9}$$

where the bar denotes the Heaviside operational transform that replaces differentiations with respect to time by the algebraic operator p and where

$$\mathrm{Q} = \sqrt{\left[\frac{p(\xi^2 + p)}{1 + p}\right]}$$

and d is the depth of the dipole below the surface of the core. For *large* times (small p) it seems permissible to replace the denominator in the integral (9) by its most effective term, namely $-2\xi^2\mathrm{Q}$. We then find that

$$h_z(\mathrm{P}) = \frac{2\mathrm{M}}{d^3} f(z, t)$$

where $f(z, t)$ has been defined in equation (3). Thus at least for large times ($t \gg 1$) the presence of the boundary does not affect the skin depth argument for disturbances of this type. It may easily be shown that, to this approximation, the same result is substantially true for other points of the core boundary although there is evidence that diffusion plays an important role at great distances. Again, the fluid flow along the core boundary opposes the redistribution of lines of force taking place at the dipole but the effect is too small to seriously modify the results. (See also a paper by COULOMB [7] who shows that unreasonably large velocities are necessary at the core boundary to produce by themselves important changes in the pattern of magnetic force there.)

In the case (*b*) suppose the normal to S is along the x-axis. All the conditions, with the exception of (ii) can be satisfied without the introduction of an external field. All that is required is the introduction of a reversed and reflected image dipole ($-\mathrm{M}_x$, $-\mathrm{M}_y$, $-\mathrm{M}_z$) at the image point together with another rather more complicated part. Again an external field and an internal hydromagnetic disturbance can be added to remedy this situation without upsetting the conditions

already satisfied. Calculations have not, as yet, been completed in these cases and no further comment will be made here.

V. CONCLUSIONS AND ACKNOWLEDGEMENTS.

Assuming the core to be a rigid conductor, several authors have argued that the disturbances responsible for the secular variation arise in a thin layer at the core boundary. However, in a conducting fluid, the diffusion of electromagnetic energy is generally much less rapid than its transmission by ALFVEN waves and, in consequence, the foregoing conclusion is incorrect. This has been illustrated in this paper by several examples in one of which the effect of a boundary has been considered.

The author wishes to make a full acknowledgement to Mr. M. G. EL MOHANDIS of this department for helping with the discussion and numerical calculations of § 3 (the figures 2 are taken from his work). He has developed independently other methods of investigating the problems of § 4 and these he plans to publish at a latter date. The author is grateful to Dr. R. HIDE for constructive criticisms and to Dr. F. J. LOWES for his consistent refusal to believe that the arguments of § 2 were adequate without further amplification, thereby fostering an interest in the problem.

Manuscrit reçu à Cambridge, pour regroupement, le 1er *octobre* 1957.
Manuscrit reçu à Paris, pour impression, le 17 *avril* 1958.

RÉFÉRENCES

[1] H. LAMB, *Proc. London Math. Soc.*, **13**, 1881, 51.
[2] A. HERZENBERG and F. J. LOWES, *Phil. Trans. Roy. Soc. A.*, **249**, 1957, 507.
[3] P. H. ROBERTS, *Ap. J.*, **122**, 1955, 315.
[4] P. H. ROBERTS, *Ap. J.*, **126**, 1957, 418.
[5] M. G. EL MOHANDIS, and P. H. ROBERTS, to be submitted to the *Ap. J.*
[6] C. WALÉN, *Ark. f. Mat., Astr. o. Fys.*, **30** A, 1944, nº 15.
[7] J. COULOMB, *Ann. de Géophys.*, **11**, 1955, 80.

Imprimerie TAFFIN-LEFORT à Lille (France). — *Published in France*

On Analysis of the Secular Variation

1. A Hydromagnetic Constraint: Theory

P.H. ROBERTS
Department of Mathematics, University of Newcastle upon Tyne, England
S. SCOTT
Rutherford College of Technology, Newcastle upon Tyne, England.

(Received April 25, 1965)

Abstract

It is noted that the time-scale of the secular variation of the geomagnetic field is rather short compared with the electromagnetic diffusion times appropriate to the earth's core. It is therefore suggested that the secular variation is primarily due to the rearrangement of pre-existing lines of force emanating from the core, and not due to the creation of new (or destruction of old) flux tubes by electromagnetic diffusion. The theoretical consequences of this idea are fully examined.

1. Introduction

In 1838 Gauss showed in a celebrated paper that, to a high degree of approximation, the main geomagnetic field is a potential field due to sources within the earth. Since his time, as better and more extensive data became available, there have been many repetitions of his analysis in terms of spherical harmonics. One motivation has been a desire to possess maps that exhibit the world-wide structure of the field, and not merely its values at a number of isolated data points. This, however, has *not* been the only objective. Indeed, if it were, it might have been argued (particularly recently) that, from the standpoint of *numerical* accuracy, spherical harmonics are not the best choice for the representation series. There has been another* objective: that of understanding the nature of the sources of field. Here spherical harmonics have two important *physical* advantages. First, they facilitate the separation of the field due to external sources from that due to internal sources. Second, they allow, *in principle*, the field at the earth's surface ($r=a$) to be readily extrapolated to the core boundary ($r=b$): each term of the representation series is multiplied by $(a/b)^n$, where n is its harmonic number**. *In practice*, the data have proved insufficient in quantity

* There have also, of course, been many other objectives, e.g. to be able to compute the trajectories of charged particles in the earth's environment.

** Provided, that is, ferromagnetic sources in the mantle are absent. To the extent that the earth is spherically symmetric, *steady* currents flowing in the mantle are toroidal and therefore produce no detectable magnetic fields at $r=a$. It is therefore permissible to assume, as we have done here, that a geomagnetostatic potential exists throughout the mantle and which satisfies Laplaces' equation. This conclusion is not correct for the secular variation since unsteady currents in the core can induce toroidal currents in the mantle if conducting. However, this effect may be small, and will be neglected here (see below).

and quality, to permit meaningful extrapolation. More precisely, when the representation series is truncated as late as the data can possibly warrant, it shows no very obvious signs of convergence when extrapolated term by term to the core surface. In principle, other methods of extrapolation are possible. For example, after the analysis of the field into physical sources, such as dipoles, (cf. Alldredge and Hurwitz, 1964a; Lowes and Russell, 1965), the field of each could be computed over the core boundary, and the resulting terms be resummed. Another avenue has been opened by the integral methods of Vestine et al (1963). It seems, however, that, whatever method is employed, severe uncertainties remain in the values calculated at the core surface.

The impracticability of extrapolation appears in a much more acute form when the secular variation field is considered. Here the time scale of the phenomenon so strongly indicates the field is internal in origin that only terms in ascending powers of (a/r) are retained in the harmonic series. Even so, the data itself is of such poor quality that direct extrapolation to the core is virtually meaningless, and it is difficult to believe that further progress is possible without some additional physical postulate. In this paper, we propose the introduction of a very simple result from hydromagnetics; the normal flux of magnetic field from any element on the surface of a body of perfectly conducting fluid remains constant during any motions of the fluid (even though they convect, dilate and distort this element). Or, as it is sometimes restated, the magnetic field is "frozen in" the fluid as it moves. In § 2, we discuss how far this result is applicable to the geomagnetic field and its secular variation. We should observe here that the introduction of a physical hypothesis in the analysis of geomagnetic data is by no means new. For example*, prior to the division of the geomagnetic field into harmonics, it is often (e.g. Vestine et al, 1947; Fougere and Mc Clay, 1957) smoothed in such a way that the line integral of the field round any horizontal curve is zero. In other words, the hypothesis is made that ground to air currents are absent. Now this assumption is not literally and precisely correct, and were data of unlimited accuracy available, this could be demonstrated. But, in making the hypothesis, the belief and claim is that deviations from the hypothesis are so small that they are lost in the general noise, at the present level of data quality. In precisely the same way, we do not regard our hydromagnetic model as being literally and precisely obeyed at the core surface. Rather, we believe that deviations from it are likely to be small or absent in the truncated representation series at the earth's surface. As a corollary it follows that our confidence in the orthodox secular variation analysis is weakened.

The hydromagnetic assumption clearly introduces a link between the main field and its secular variation, and in principle, therefore, its use prohibits an analysis of either separately. However, since the data for the main field are so much more reliable than for its secular

* Other examples from geomagnetism are (a) the analyses by Fougere (1964) and Fougere and McClay (1957) of the main field in which the pythagorean relationship $F^2=X^2+Y^2+Z^2$ of the X, Y and Z components of total intensity F were used as a constraint, (b) the analysis by Scott (1965a) of the field using the time derivative of the pythagorean relationship $F\dot{F}=X\dot{X}+X\dot{X}+Z\dot{Z}$ as constraint. Other applications of the idea are common in data processing, cf.e.g. "Lectures on numerical weather prediction" by Berkofsky (1962).

variation, it is likely that, when appropriate statistical weights are allotted to allow for this fact, the main field analysis is almost independent of that of the secular variation. The converse is not true, and in this paper we examine the effect of the hydromagnetic hypothesis on the analysis of the secular variation, assuming that the main field is known. It might be objected that we have merely transformed the problem of finding secular variation coefficients to one of finding like coefficients for a harmonic expansion of the horizontal velocity field on the core surface. This is not correct. In principle, given a secular variation field, it is generally *not* possible to find a velocity field that will generate it from the main field by electromagnetic induction. What our hydromagnetic hypothesis does is to introduce a constraint on the analysis, albeit a constraint of a somewhat unusual nature. It should be noted that, although the explicit connection of main field and secular variation is made at the core surface, the constraint is applied at the earth's surface. Thus our analysis avoids explicit extrapolation of fields to the core, and the concommitant growth of errors. This point is discussed further in § 4.

2. The Hydromagnetic Assumption

If the values ($\sigma \simeq 3.10^5$ mho/m) often quoted for the conductivity of the core are correct, it is, by laboratory standards, a rather poor conductor. However, from the standpoint of the hydromagnetics of the secular variation, it is an extremely good conductor, as we shall now demonstrate.

The change in magnetic field at a fixed point due to the motion of an electrically conducting fluid is often visualized in two parts. First, there is a change in the field due to the convection of the magnetic flux tubes by the motion. Second, there is a diffusion of the field lines relative to the motions due to ohmic dissipation. The relative importance of these two contributions is gauged by the magnetic Reynolds number, R_m. This is defined by

$$R_m = \frac{UL}{\eta}, \qquad (1)$$

where L is a length characteristic of the magnetic field, U is a velocity characteristic of the flow, and η (cm^2/sec) is the magnetic diffusivity which, in electromagnetic units, is defined by

$$\eta = \frac{1}{4\pi\sigma\mu}, \qquad (2)$$

where σ is electrical conductivity and μ is permeability (e.m.u.). When $R_m \gg 1$, the change in magnetic field is primarily due to the convection of flux tubes by the motions. This may readily be seen if we re-write the condition $R_m \gg 1$ as

$$\tau_\sigma \equiv \frac{L^2}{\eta} \gg \frac{L}{U} \equiv \tau_u . \qquad (3)$$

According to (3) the electromagnetic decay time τ_σ, during which sources can create new flux tubes and ohmic decay can destory old, is large compared with the dynamical time-

scale τ_u, during which the fluid motions can substantially alter the field by redistributing the lines of force*. If we take $L \approx 10^8$cm and $U \approx 1$mm/s, as characteristic of the secular variation, we find $\tau_u \approx 30$ years. But if we accept the estimate $\sigma = 3.10^{-6}$ e.m.u. mentioned earlier, and set here (and henceforth) $\mu = 1$, we obtain $\eta = 2.4\ 10^4$ cm^2/sec; whence $\tau_\sigma \approx 12{,}000$ yr. Thus (3) holds for the secular variation.

It seems to us inevitable, then that the main body of the core may be treated as a perfectly conducting fluid when the secular variation is discussed. Allan and Bullard (1965) have, however, urged that there is a loophole in the argument just given. While not disputing that τ_σ governs the rate of separation of field lines from fluid, they point out that this would still imply a large secular change when the field involved is large. Now, it is generally agreed that the toroidal fields, though weak at the surface of the core (because of the high resistivity of the mantle), are large at great depths (because of the shearing of poloidal field by fluid motions). Any sufficiently rapid upwelling of fluid will convect the toroidal field to the core surface and, inevitably, some will separate from the fluid by ohmic diffusion, and will spread into the mantle. To estimate the consequent secular change, Allan and Bullard have constructed models in which an upwelling motion is suddenly "switched on" in a half-space of fluid which contains a field initially tangential to the (plane rigid) boundary, and increasing in strength linearly with distance from it. Using the equations of electromagnetic induction, they computed the subsequent growth of field in the (insulating) exterior of the fluid. They found that the dominant harmonic generated by the upwelling gives a field of the form

$$Z = Z_\infty f(\eta t / L^2), \qquad (4)$$

where L is a length characteristic of the applied field gradient or of the depth the upwelling reaches, whichever is less. The constant Z_∞ is the value attained by Z in the final steady state attained when the diffusion of field through the moving fluid is everywhere equal and opposite to the convective motions; it is proportional to the applied field gradient, the velocity of upwelling, and the conductivity; f approaches 1 monotically as its argument increases. The field normal to the surface is positive on one side of the upwelling, negative on the other, and might be roughly described by a dipole placed beneath the surface and directed antiparallel to the applied field.

The interpretation of Allan and Bullard's model which seems most natural is to suppose Z_∞ is of the same order as an anomaly of the main field, say 0.1 gauss at the earth's surface, at most (cf. e.g. Runcorn, 1956). This is compatible with the observation (cf. Runcorn, loc. cit., p. 511) that the lifetime of a centre of secular change is roughly equal to the magnitude of a non-dipole anomaly divided by a typical secular change rate. Quite elementary magnetohydrodynamic arguments (see boundary layer analysis below) establish that the toroidal

* The estimate L^2/η for τ_σ is not, we believe, in conflict with the findings of Chandrasekhar (1946). Chandrasekhar conidered the influence of axisymmetric motions on the decay times for a fluid sphere. He showed that suitably directed motions of sufficient vigour could reduce these times far below L^2/η, where L is the radius of the sphere. However, we believe this is essentially because these motions convected the regions of strong flux into length scales small compared with L.

field gradients are not abnormally large at the core surface, nor is its characteristic scale length small thereat. It is not unreasonable, then, to take $L=500$km as typical for a large centre of change (cf. Herzenberg and Lowes, 1958). We then obtain (using (4) and assuming $f'=0(1)$)

$$\dot{Z} \lesssim 0\left(\frac{\eta Z_\infty}{L^2}\right) \sim 3\gamma/\text{year}\,, \qquad (5)$$

which is rather small to be of interest. Allan and Bullard point out that the *initial* growth of the function f appearing in (4) is rapid ($f'(\eta)>1$, small η). But this is partially a consequence of their model, and if, instead of supposing the velocity is suddenly switched on, we take the more reasonable (though still extreme) view that the body force creating the motion is suddenly switched on, the initial rise of f is more gradual by a factor of $\eta t/L^2$ (see Hide and Roberts, 1961, pp. 92–95). Moreover the position is not substantially improved when Allan and Bullard's model is modified by the inclusion of the Navier-Stokes equation to a fully hydromagnetic (perturbation) model; cf. Hide and Roberts, loc. cit.

The value of $\dot{Z}$ obtained in (5) above rests squarely on the hypothesis that Z_∞ is comparable in magnitude with an anomaly of the non-dipole field. If one is willing to abandon this idea, it is possible to increase $\dot{Z}$ to reasonable values ($\sim 100\gamma/yr$) by increasing U and, therefore, Z_∞. But one must then consider the upwelling is shortlived, compared with the lifetime of the centres themselves; otherwise Z would reach values comparable with Z_∞. This complication reduces the simplicity and appeal of the model. Moreover, the empirical observation concerning the lifetime of secular variation centres mentioned above is difficult to explain.

One other slight obstacle presented by Allan and Bullard's models should be noted. They would suggest that secular change fields observed at the earth's surface would be best represented by horizontal dipoles positioned beneath the core. However, the opposite view (i.e. vertical dipoles) has been strongly championed by Lowes and Runcorn, 1951.

For all these reasons, we prefer to abandon Allan and Bullard's mechanism in favour of the hydromagnetic assumption; viz., that we may assume $\eta=0$ except, of course, in boundary layers. In the main body of the core, therefore, the equation of electromagnetic induction becomes

$$\frac{\partial \boldsymbol{B}}{\partial t} = \text{curl}\,(\boldsymbol{u}\times\boldsymbol{B})\,, \qquad (6)$$

where $\boldsymbol{u}$ is the fluid velocity. We may note here a simple consequence of (6) which will be useful later. Taking spherical polar coordinates (r, θ, ϕ) based on the geocentre and fixed relative to the mantle, the r component of (4) gives

$$\frac{\partial B_r}{\partial t} = u_r\left[\frac{1}{r\sin\theta}\frac{\partial}{\partial\theta}(B_\theta\sin\theta)+\frac{1}{r\sin\theta}\frac{\partial B_\phi}{\partial\phi}\right]$$
$$-B_r\left[\frac{1}{r\sin\theta}\frac{\partial}{\partial\theta}(u_\theta\sin\theta)+\frac{1}{r\sin\theta}\frac{\partial u_\phi}{\partial\phi}\right]$$

$$+\left(\frac{B_\theta}{r}\frac{\partial}{\partial\theta}+\frac{B_\phi}{r\sin\theta}\frac{\partial}{\partial\phi}\right)u_r-\left(\frac{u_\theta}{r}\frac{\partial}{\partial\theta}+\frac{u_\phi}{r\sin\theta}\frac{\partial}{\partial\phi}\right)B_r. \qquad (7)$$

Equations (6) and (7) hold in the main body of the core but *not* in a thin boundary layer in contact with the core surface, $\mathscr{S}$. Since we are vitally interested in the behaviour of the field as it leaves the core, it is absolutely essential to consider the structure of the boundary layer in some detail. First we note that the equation

$$\text{div}\,\boldsymbol{B}=0 \qquad (8)$$

ensures that as we descend from any point of $\mathscr{S}$, assumed spherical, through the boundary layer, B_r does not change in the first approximation*, i.e.

$$[B_r]_i^0=0. \qquad (9)$$

Similarly, since compressibility is unimportant at secular variation frequencies, we have

$$[u_r]_i^0=0. \qquad (10)$$

But at the outer edge $\mathscr{S}_0$ of the boundary layer, the fluid is in contact with a rigid solid. Thus

$$u_r(\mathscr{S}_0)=0. \qquad (11)$$

From (10) and (11) we have

$$u_i(\mathscr{S}_i)=0, \qquad (12)$$

from which it follows, since $\mathscr{S}_i$ is spherical, that

$$\frac{\partial u_r}{\partial\theta}=\frac{\partial u_r}{\partial\phi}=0 \text{ on } \mathscr{S}_i. \qquad 13)$$

Since $\mathscr{S}_i$ is the junction of main stream and boundary layer, (7) also holds thereat. Thus, by (7), (12), (13) we see that *on* $\mathscr{S}_i$

$$\frac{\partial B_r}{\partial t}=-B_r\left[\frac{1}{r\sin\theta}\frac{\partial}{\partial\theta}(u_\theta\sin\theta)+\frac{1}{r\sin\theta}\frac{\partial u_\phi}{\partial\phi}\right]-\left(\frac{u_\theta}{r}\frac{\partial}{\partial\theta}+\frac{u_\phi}{r\sin\theta}\frac{\partial}{\partial\phi}\right)B_r. \qquad (14)$$

As a result of (9), we see that (14) holds not only on $\mathscr{S}$ *but also on* $\mathscr{S}_0$, i.e. at the base of the mantle. This is the result we will make use of in § 4. We conclude this section by showing, through further consideration of the boundary layer at $\mathscr{S}$, that not only is the radial component (6), but also the θ and ϕ components of equation (14), are obeyed at $\mathscr{S}_0$. This result, in principle, contains information about $\partial\boldsymbol{B}/\partial r$ on $\mathscr{S}_i$, i.e. about the toroidal currents flowing near the core boundary.

* The notation $[....]_i^0$ used in (9) denotes the leap in the quantity...as we pass from the outer edge, $\mathscr{S}_0$, of the boundary layer to its inner edge, $\mathscr{S}_i$. The reversal of the usual boundary layer terminology has been deemed advisable here, since "inner" and "outer" have obvious meanings in relation to the geocentre. In common with all boundary layer analyses we are in reality considering the leading terms of an asymptotic solution in the limit $\eta\to 0$. However, the present "engineering" picture of a layer of finite physical extent has seemed clearer than a strict mathematical description.

The estimate made of η made earlier exceeds the values of the kinematic viscosity $\nu(10^{-3}$ to 10^{3} cm^2/sec) which are often quoted (see, for example, the discussion of Hide and Roberts, 1961). It is likely, then, that the magnetic Prandtl number,

$$p_m=\frac{\nu}{\eta}, \tag{15}$$

is small. And, since R_m is large, so is the hydrodynamic Reynolds number

$$R_e=\frac{UL}{\nu}. \tag{16}$$

Thus, in the bulk of the core, or what in the language of boundary layer theory might be termed 'the main stream,' we may take $\nu=\eta=0$. The consequent lowering of the order of the governing differential equations means that it is impossible, in general, for the main stream to satisfy the no slip conditions on $\boldsymbol{u}$

$$\boldsymbol{u}=0 \text{ on } \mathscr{S}_0, \tag{17}$$

and the conditions

$$\boldsymbol{B} \text{ continuous across } \mathscr{S}_0, \tag{18}$$

which are required of a physically reasonable solution at the base $\mathscr{S}_0$ of the mantle. And, of course, it is the boundary layer which must effect the necessary adjustments.

In the absence of magnetic fields, the neglect neglect of ν reduces the order of the governing differential equation of hydrodynamics by 2. Thus the number of boundary demands on the main stream at $\mathscr{S}_i$ must be reduced by 2. It is, in fact, only required, in the first approximations, to satisfy one of the conditions (17), namely $u_r=0$, and $\boldsymbol{u}$ is then determined completely. In general, the solution obtained will not satisfy $u_\theta=u_\phi=0$. This situation is corrected by a boundary layer, of thickness $(\nu\tau)^{1/2}$, in which the diffusive term $\nu\nabla^2$ is crucial, which matches with the main stream it its inner edge, and which obeys the no slip conditions on $\mathscr{S}_0$. (Here τ denotes the time scale of the variations.)

In the absence of fluid motions, the neglect of η reduces the order of the governing differential equations of electrodynamics by 2. Thus the number of boundary demands on the "main stream" at $\mathscr{S}_i$ must also be reduced by 2. It is, in fact, only required, in the first approximation, to satisfy one of the conditions (18) on $\boldsymbol{B}$, namely that on B_r, and $\boldsymbol{B}$ is then determined completely. In general, the solution obtained in this way will not satisfy the conditions on B_θ and B_ϕ. This situation is corrected by a boundary layer, of thickness $(\eta\tau)^{1/2}$, in which the diffusive term $\eta\nabla^2$ is crucial, which matches with the main solution at its inner edge and with all the continuity conditions on $\mathscr{S}_0$.

In the presence of both fluid motions and magnetic fields, the approximation $\nu=\eta=0$ reduces the order of the governing differential equations of magnetohydrodynamics by 2, *and not by* 4. In general (more precsely, when $B_r\neq 0$), the number of boundary demands must be reduced by 2, *and not by* 4. The main stream is, in fact, required, in the first approximation, to satisfy, not only the conditions on u_r and B_r, but also two other conditions. This raises the interesting dilemma, "Do we keep the conditions on u_θ and u_ϕ, or those on B_θ and

B_ϕ, or do we replace these four conditions by two others?" Stewartson (1957, 1960a, b) first answered this question. He showed that a boundary layer, whose thickness is $(\nu\eta)^{1/2}/V_A$ (where V_A is Alfven velocity based on B_r; i.e. $V_A=B_r/\sqrt{4\pi\mu\rho}$), exists in which both the diffusive terms νV^2 and ηV^2 are crucial. In the first approximation, not only are u_r and B_r continuous across this layer, but also

$$\left[u_{\theta,\phi}-\frac{B_{\theta,\phi}}{\sqrt{4\pi\mu\rho}}\left(\frac{\eta}{\nu}\right)^{1/2}\right]_i^0=0\,, \tag{19}$$

where ρ is the density of the fluid. In other words, if we know the fields $B_{\theta,\phi}(\mathscr{S}_0)$ at the base of the mantle, we must require the main stream solutions to obey

$$p_m{}^{1/2}u_{\theta,\phi}(\mathscr{S}_i)-\frac{1}{\sqrt{4\pi\mu\rho}}B_{\theta,\phi}(\mathscr{S}_i)=-\frac{1}{\sqrt{4\pi\mu\rho}}B_{\theta,\phi}(\mathscr{S}_0)\,. \tag{20}$$

Only when this has been achieved is it possible to construct a boundary layer which satisfies (17) and (18) at $\mathscr{S}_0$.

It is, perhaps, curious that, even though the main stream solutions are based on the assumptions $\nu\to0$, $\eta\to0$, a knowledge of their ratio p_m is required before they can be determined correctly.

In the geophysical case ($p_m\ll1$), (20) reduces simply to

$$[B_\theta]^0_i=[B_\phi]^0_i=0,$$

i.e. by (9)

$$[\boldsymbol{B}]^0_i=0\,. \tag{21}$$

This has three consequences of interest. First, there are no appreciable surface currents in the boundary layer.* Second, the fields we observe at the earth's surface are linked directly to the fields at the edge of the main stream, and are uninfluenced by conditions in the boundary layer. Third, no demands are made of u_θ and u_ϕ for the main stream at its junction $\mathscr{S}_i$ with the boundary layer. It is certainly not to be expected that the no slip conditions will be obeyed.

It may be objected that strictly (19) should only be applied to a steady state. However, as Stewartson (loc. cit.) demonstrated, if condition (19) is violated, an Alfvén wave is emitted by the boundary layer as a kind of tangential shock discontiouity. This passes into the main stream leaving (19) satisfied behind it. In fact, (21) is restored in a time of order $\tau_A=\rho/\sigma B_r^2$

* Were the core is solid, the electromagnetic skin-depth $(\eta\tau)^{1/2}$ for a secular variation time scale τ of 40 years is only 4.10^6cm. In many early discussion (e.g. Lowes and Runcorn, 1951, § 6) of the secular variation it was erroniously concluded that this result has some bearing on the electrodynamics of the *fluid* core, and that it therefore followed that the secular variation is produced by intense currents flowing within a depth of 4.10^6 cm. of $\mathscr{S}$. The fallacy was first pointed out by Roberts (1954, 1955, 1959; reported by Runcorn, 1954); for further corroboration, see Herzenberg and Lowes (1958). It is curious how, in this geophysical context, allowance for the mobility of the core utterly reverses the conclusions of the skin-depth argument for solid conductors. (Tamao (1959) and Rochester (1960) have recently sought to maintain the earlier conclusion through the analysis of the attenuation of particular Alfvén waves in the presence of Coriolis forces. However, the model chosen is rather special and, in fact, others could equally well be exhibited in which Coriolis forces have no effect at all on the conclusions detailed above.)

$=\eta/V_A^2$. Thus, if we consider a phenomenon whose time scale τ is large compard with τ_A, the boundary layer can continuously adjust itself so that (21) is always satisfied. In that present geophysical context, $\rho=10$ gm/cc, and taking $B_r=1G$, (a conservative estimate), we obtain $\tau_A=6$ weeks. Since this is short compared with the secular variations time scale, we may accept (21) without reservation.

As a result of (21) we see that (6) holds not only on $\mathscr{S}$ but also on $\mathscr{S}_0$, i.e. at the base of the mantle.

3. The Nature of the Constraint Imposed by the Hydromagnetic Assumption

The aim of this section is to discuss the bearing of (14), applied to the main field $\boldsymbol{B}$ and its secular variation $\partial\boldsymbol{B}/\partial t$ at the core surface $r=b$, to the relationship of these fields at the earth's surface $r=a$. The connection can only be established if some physical assumptions concerning the intervening mantle $a>r>b$, are made. The electromagnetic decay time of the mantle has been estimated to be less than four years (cf. Hide and Roberts, 1964, § IV), a time which is rather small compared with that charateristic of the secular variation. It may be concluded that the effects of currents induced in the mantle on the relationship between the field at $r=a$ and that at $r=b$, are fairly small. In this paper, we neglect them entirely, and treat the mantle as an insulator.* The connection between $r=a$ and $r=b$ is then unique (assuming, as is reasonable, no important current sources). To prove this, we may argue as follows:

Since there are no currents outside $r=b$, there exists a geomagnetic potential $V(t)$ such that

$$\boldsymbol{B}=-\mathrm{grad}V, \quad (r>b), \tag{22}$$

where

$$\nabla^2 V=0, \tag{23}$$

and

$$V=0(r^{-2}), \quad r\to\infty. \tag{24}$$

By (23), (24) and a well known theorem of potential theory, the specification of $B_r(=-\partial V/\partial r)$ on $\mathscr{S}$ determines $\boldsymbol{B}$ uniquely in $r\geqslant b$.

In principle, from a knowledge of $\boldsymbol{B}$ on the earth's surface ($r=a$), we may use (22) to (24) to find B_r on $\mathscr{S}$. If we know $\boldsymbol{u}$, we may then use (14) to compute $\partial B_r/\partial t$ on $\mathscr{S}$, and then use (22) to (24) again to extrapolate the result back to $r=a$, thereby determining $\partial\boldsymbol{B}/\partial t$ on the earth's surface. Even if we do not know $\boldsymbol{u}$, the very postulate that it exists, and satisfies (12), is enough to establish a constraint between $\boldsymbol{B}$ and $\partial\boldsymbol{B}/\partial t$ on $r=a$. It is, in fact, only in

* Intuitively, it is clear that the observed time-scale of the secular variation is lengthened by the electromagnetic screening of the mantle. Indeed, it is through a study of this lengthening that the estimate of four years above was reached. For this reason the present authors believe that the neglect of conductivity in the mantle has more serious consequences on their conclusions than, for instance, the possible partial violation of the hydromagnetic assumption discussed in § 3. They hope to return to this question in a later paper.

this sense that we propose to use (14): we postulate no particular form for $\boldsymbol{u}$. The analysis does cast up, as a by-product, some information about $\boldsymbol{u}$, although not enough to determine it uniquely.

It follows from (14) that the flux of $\boldsymbol{B}$ through a surface element of $\mathscr{S}$ as it is convected by $\boldsymbol{u}$ remains constant. It is, as a matter of fact, always possible to define a ficticious "velocity" $\boldsymbol{U}$ whose only purpose is to convect flux tubes. And it might be thought that the existence of such a velocity field contradicts our claim that the hydromagnetic hypothesis is a constraint to our analysis. To examine this point, we observe that, the existence of a vector field $\boldsymbol{U}$ which convects the magnetic field implies

$$\frac{\partial \boldsymbol{B}}{\partial t}=\mathrm{curl}(\boldsymbol{U}\times\boldsymbol{B}); \tag{25}$$

whence, by Faraday's law of induction

$$\mathrm{curl}\boldsymbol{E}=-\frac{\partial \boldsymbol{B}}{\partial t} \tag{26}$$

for the electric field $\boldsymbol{E}$, we have

$$\mathrm{curl}(\boldsymbol{E}+\boldsymbol{U}\times\boldsymbol{B})=0. \tag{27}$$

Thus there exists a single-valued scalar field ϕ such that

$$\boldsymbol{E}+\boldsymbol{U}\times\boldsymbol{B}=-\mathrm{grad}\ \phi. \tag{28}$$

It follows from this equation that

$$\phi=-\int \boldsymbol{E}\cdot\boldsymbol{B}\frac{ds}{B}, \tag{29}$$

where the integral is taken along a line of force. Since there can be no closed lines of force in the exterior of $\mathscr{S}$, (since it is insulating), ϕ, as defined by (29) is single-valued therein, as demanded above. Now by (28) we have

$$\boldsymbol{U}=(\boldsymbol{E}+\mathrm{grad}\ \phi)\times\frac{\boldsymbol{B}}{B^2}+\chi\boldsymbol{B}, \tag{30}$$

where χ is an arbitrary scalar field. If χ is bounded, the velocity (30) is also bounded everywhere on and outside $\mathscr{S}$, since, by Earnshaw's theorem for a potential field, $\boldsymbol{B}$ can have no zeros in an insulator. It would be possible, if we wished, to choose χ so that div $\boldsymbol{U}=0$ outside $\mathscr{S}$; to do this we require that

$$\chi=-\int\frac{ds}{B}\,\mathrm{div}\left[(\boldsymbol{E}+\mathrm{grad}\ \phi)\times\frac{\boldsymbol{B}}{B^2}\right], \tag{31}$$

where the integral is taken along a line of force. However, even if we chose χ at one of the intersections of the line of force with $\mathscr{S}$ in such a way that $U_r=0$, the value at the other intersection, which is then uniquely determined by (31), could *not* generally be such that $U_r=0$. *Moreover,* even if the condition div $\boldsymbol{U}=0$ is entirely abandoned, it is still not generally possible to make $U_r=0$ on $\mathscr{S}$, without making χ, and therefore $\boldsymbol{U}$, infinite thereon. (cf. eq.

(30); in general, there is no reason why $r\times\boldsymbol{B}$ and $\boldsymbol{E}+\text{grad}\ \psi$ should be perpendicular at the curves on $\mathscr{S}$ for which $B_r=0$.) Thus, our hypothesis of a *bounded* velocity field $\boldsymbol{u}$ convecting field on is indeed a constraint.

This may also be very easily appreciated in the case in which $\boldsymbol{u}$ is purely toroidal. For then (14) reduces to

$$\frac{\partial B_r}{\partial t}=-\left[\frac{u_\theta}{b}\frac{\partial}{\partial\theta}+\frac{u_\phi}{b\sin\theta}\frac{\partial}{\partial\phi}\right]B_r \text{ on } \mathscr{S}, \tag{32}$$

which implies that the foci of the main field on the core surface (i.e. the points of $\mathscr{S}$ at which B_r achieves maximum or minimum values) must lie on the curves of zero $\dot{B}_r$. It is also clear that, even when the main field and secular variation fields are compatible so that (32) possesses a solution for $\boldsymbol{u}$, that solution is not unique. For we can add to $\boldsymbol{u}$ any flow round the curves of equal B_r drawn on the core surface. (Again, if B_r and $\dot{B}_r$ are axisymmetric, as they are in some of the examples we have treated, the component of $\boldsymbol{u}$ along lines of longitude is uniquely determined by (14). But is readily seen that *any* flow round lines of latitude may be added without disturbing the relationship between B_r and $\dot{B}_r$.)

Some other properties of the hydromagnetic assumption should be noticed. Since the flux of field out of each element of $\mathscr{S}$ is conserved, it follows that the integrated flux, $\mathscr{F}$, out of the core (which, of course, is equal and opposite to the net flux, $-\mathscr{F}$, *into* the core) is constant in time. This does not mean that any of the multipolar moments are constant in time. Indeed, the axial dipole moment (for example) can take any value between $3b^3\mathscr{F}/2$ and $-3b^3\mathscr{F}/2$, the extreme values being taken when the incoming and outgoing lines of force are bunched tightly round each pole. Thus, the theory has no bearing at all on the speculation that the earth's dipole moment is changing (cf. Bullard, 1953).

The fact that $\mathscr{F}$ is conserved on the *core* surface in no way implies that the net positive (or negative) flux is conserved over the *earth's* surface. This is made intuitively clear by figure 1. It can also be demonstrated analytically by a simple example. Suppose the

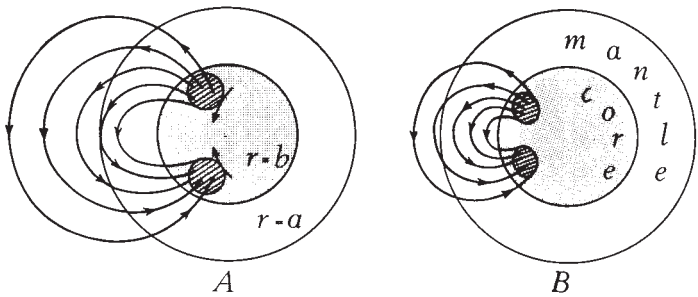


Fig. 1. This schematic diagram compares the disposition of a particular flux tube leaving the core, when the two areas of its intersection with the core surface are moved together by fluid motions (A). It is clear, from the Faraday-Maxwell picture of lines of force, that the tube will shorten and move closer to the core (B). Thus the flux crossing the earth's surface in the outward sense is reduced. The inward flux is, of course, reduced by the same amount.

magnetic potential V is axisymmetric and given by

$$V(r, \mu, t)=-k(t)\left[\left(\frac{b}{r}\right)^2 P_1(\mu)+\frac{2}{3}c(t)\left(\frac{b}{r}\right)^3 P_2(\mu)\right], \tag{33}$$

whence $\mu=\cos\theta$, $\theta=$colatitude, $c<1$ and k is chosen to be

$$k=\frac{27c^2\mathscr{F}}{4\pi b[(9c^2-1)+(1+3c^2)^{3/2}]} \tag{34}$$

By (34), the net positive flux over the core surface is $\mathscr{F}$, which we suppose is constant. The net positive flux, f, over the earth's surface is then

$$f=\frac{\mathscr{F}}{ba^2}\left[\frac{a(9b^2c^2-a^2)+(a^2+3b^2c^2)^{3/2}}{(9c^2-1)+(1+3c^2)^{3/2}}-\right], \tag{35}$$

and this certainly varies as $c=c(t)$ varies. (It is also interesting to observe that, if we apply (14) to the field (33), the fluid velocities required to give a secular variation of 200γ/year at colatitude 60° from a field of $20{,}000\gamma$ are only of the order of 3 mm/sec, which is not unreasonably large, geophysically.)

It is appropriate here to comment on the relationship of Yukutake's study of the westward drift (Yukutake, 1962) to our analysis. Yukutake introduces a drift velocity $\boldsymbol{v}_0{}^1$ of field in the mantle which is, in some senses, analagous to our $\boldsymbol{U}$. Strictly, however, his velocity field $\boldsymbol{v}_0{}^1$ cannot convect field lines over the earth's surface unless it is toroidal thereon (compare his eq. (1-4) with eqs. (14) and (32) above). That it should be mainly toroidal on the core surface we have some reason to expect from the westward drift; but that it should also be true on the earth's surface is certainly arguable (cf. Fig. 1 above where an inward component of $\boldsymbol{U}$ would be required on $r=a$ to produce the right-hand diagram from the left.) It should also be noted that the induction equation (14) is linear but *not* homogeneous in $\boldsymbol{u}$. It is therefore strictly not permissible to superpose solutions obtained from different $\boldsymbol{u}$, as Yukutake has, in essence, done.

4. Reduction to Matrix Equations

Let us introduce the customary symbol Z for $-B_r$. By (22) to (24) we have

$$\nabla^2(rZ)=0 \text{ for } r>b, \tag{36}$$

where

$$Z=0(r^{-3}),\ r\to\infty. \tag{37}$$

Thus, for $r>b$, we may expand Z in a series of the form

$$Z=\sum_{m=1}^{\infty}\left(\frac{a}{r}\right)^{m+2}\left\{A_m P_m(\theta)+\sum_{v=1}^{n}(A_{m,v}\cos v\phi+B_{m,v}\sin v\phi)P_m^v(\theta)\right\}. \tag{38}$$

We will adopt the convention

$$A_{m,v}=B_{m,v}=0, \text{ if } v\leq 0. \tag{39}$$

In virtue of (12) we may express $\boldsymbol{u}$ on $\mathscr{S}$ by

$$\boldsymbol{u}=(0,\ u_\theta(\theta,\phi),\ u_\phi(\theta,\phi))\,. \tag{40}$$

However, it does not follow from (12) that $\partial u_r/\partial r$ vanishes on $\mathscr{S}$; indeed, over an upwelling (cf. e.g. Coulomb, 1955) this quantity will be large in magnitude. We find it is convenient to separate $\boldsymbol{u}$ (assumed solenoidal near $\mathscr{S}$) into a toroidal part (T), and a poloidal part (S):

$$\boldsymbol{u}=\left(0,\ \frac{\partial S}{\partial\theta}+\frac{1}{\sin\theta}\frac{\partial T}{\partial\phi},\ \frac{1}{\sin\theta}\frac{\partial S}{\partial\phi}-\frac{\partial T}{\partial\theta}\right). \tag{41}$$

Then the scalar L^2S takes the value of $b\partial u_r/\partial r_r$ on $\mathscr{S}$. Here L^2 is the operator defined by

$$L^2=-\left[\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}+\frac{1}{\sin^2\theta}\frac{\partial^2}{\partial\phi^2}\right]\right]. \tag{42}$$

Using (41), equation (14) may be written as

$$b\frac{\partial Z}{\partial t}=ZL^2S-\left(\frac{\partial S}{\partial\theta}\frac{\partial Z}{\partial\theta}+\frac{1}{\sin^2\theta}\frac{\partial S}{\partial\phi}\frac{\partial Z}{\partial\phi}\right)+\frac{1}{\sin\theta}\left(\frac{\partial T}{\partial\theta}\frac{\partial Z}{\partial\phi}-\frac{\partial T}{\partial\phi}\frac{\partial Z}{\partial\theta}\right). \tag{43}$$

Let us expand T and S in harmonic series

$$T=\sum_{n=1}^{\infty}\left[C_n P_n(\theta)+\sum_{w=1}^{n}(C_{n,w}\cos w\phi+D_{n,w}\sin w\phi)P_n^{\,w}(\theta)\right], \tag{44}$$

$$S=\sum_{n=1}^{\infty}\left[E_n P_n(\theta)+\sum_{w=1}^{n}(E_{n,w}\cos w\phi+F_{n,w}\sin w\phi)P_n^{\,w}(\theta)\right]. \tag{45}$$

Multiplying equation (43) by $2P_l^{\,u}(\theta)\cos u\phi\sin\theta d\theta d\phi/w$ and integrating over all permissible θ and ϕ, we obtain, after integration by parts, (adopting Schmidt's normalization of Legendre functions)

$$\begin{aligned}
\frac{8b}{(2l+1)}\frac{dA_{l,u}}{dt}=&\sum_{m=1}^{\infty}\sum_{n=1}^{\infty}\left(\frac{a}{b}\right)^{m-l}\left\{\delta_u{}^0A_{m,0}E_{n,0}\int_{-1}^{1}(1-\mu^2)\frac{dP_l^{\,0}}{d\mu}\frac{dP_n^{\,0}}{d\mu}P_m^{\,0}d\mu\right.\\
&+\sum_{w=0}^{n}\left[(A_{m,u+w}D_{n,w}-B_{m,u+w}C_{n,w})\int_{-1}^{1}\left(uP_l^{\,u}\frac{dP_n^{\,w}}{d\mu}-wP_n^{\,w}\frac{dP_l^{\,u}}{d\mu}\right)P_m^{\,u+w}d\mu\right.\\
&+(A_{m,u+w}E_{n,w}+B_{m,u+w}F_{n,w})\int_{-1}^{1}\left((1-\mu^2)\frac{dP_l^{\,u}}{d\mu}\frac{dP_n^{\,w}}{d\mu}-\frac{uw}{1-\mu^2}P_l^{\,u}P_n^{\,w}\right)P_m^{\,u+w}d\mu\\
&-(A_{m,u-w}D_{n,w}+B_{m,u-w}C_{n,w})\int_{-1}^{1}\left(uP_l^{\,u}\frac{dP_n^{\,w}}{d\mu}+wP_n^{\,w}\frac{dP_l^{\,u}}{d\mu}\right)P_m^{\,u-w}d\mu\\
&+(A_{m,u-w}E_{n,w}-B_{m,u-w}F_{n,w})\int_{-1}^{1}\left((1-\mu^2)\frac{dP_l^{\,u}}{d\mu}\frac{dP_n^{\,w}}{d\mu}+\frac{uw}{1-\mu^2}P_l^{\,u}P_n^{\,w}\right)P_m^{\,u-w}d\mu\\
&-(A_{m,w-u}D_{n,w}-B_{m,w-u}C_{n,w})\int_{-1}^{1}\left(uP_l^{\,u}\frac{dP_n^{\,w}}{d\mu}+wP_n^{\,w}\frac{dP_l^{\,u}}{d\mu}\right)P_m^{\,w-u}d\mu\\
&+(A_{m,w-u}E_{n,w}+B_{m,w-u}F_{n,w})\int_{-1}^{1}\left((1-\mu^2)\frac{dP_l^{\,u}}{d\mu}\frac{dP_n^{\,w}}{d\mu}\right.\\
&\left.\left.\left.+\frac{uw}{1-\mu^2}P_l^{\,u}P_n^{\,w}\right)P_m^{\,w-u}d\mu\right]\right\}.
\end{aligned} \tag{46}$$

Similarly, by multiplication by $2P_l^{\,u}(\theta)\sin u\phi\sin\theta d\theta d\phi/\pi$ and integration, we obtain

$$\frac{8b}{(2l+1)}\frac{dB_{l,u}}{dt}=\sum_{m=1}^{\infty}\sum_{n=1}^{\infty}\sum_{w=0}^{n}\left(\frac{a}{b}\right)^{m-l}\Big\{(A_{m,u+w}C_{n,w}+B_{m,u+w}D_{n,w})\int_{-1}^{1}uP_l^u\frac{dP_n^w}{d\mu}$$
$$-wP_n^w\frac{dP_l^u}{d\mu}\Big)d\mu$$
$$-(A_{m,u+w}F_{n,w}-B_{m,u+w}E_{n,w})\int_{-1}^{1}\left((1-\mu^2)\frac{dP_n^w}{d\mu}\frac{dP_l^u}{d\mu}-\frac{uw}{1-\mu^2}P_l^uP_n^w\right)P_m^{u+w}d\mu$$
$$+(A_{m,u-w}C_{n,w}-B_{m,u-w}D_{n,w})\int_{-1}^{1}\left(uP_l^u\frac{dP_n^w}{d\mu}+wP_n^w\frac{dP_l^u}{d\mu}\right)P_m^{u-w}d\mu$$
$$+(A_{m,u-w}F_{n,w}+B_{m,u-w}E_{n,w})\int_{-1}^{1}\left((1-\mu^2)\frac{dP_n^w}{d\mu}\frac{dP_l^u}{d\mu}+\frac{uw}{1-\mu^2}\right)P_m^{u-w}d\mu$$
$$+(A_{m,w-u}C_{n,w}+B_{m,w-u}D_{n,w})\int_{-1}^{1}\left(uP_l^u\frac{dP_n^w}{d\mu}+wP_n^w\frac{dP_l^u}{d\mu}\right)P_m^{w-u}d\mu$$
$$+(A_{m,w-u}F_{n,w}-B_{m,w-u}E_{n,w})\int_{-1}^{1}\left((1-\mu^2)\frac{dP_n^w}{d\mu}\frac{dP_l^u}{d\mu}+\frac{uw}{1-\mu^2}P_l^uP_n^w\right)P_m^{w-u}d\mu\Big\}. \tag{47}$$

The factor $(a/b)^{m-l}$ in (46) and (47) arises from the fact that (43) is applied on $\mathcal{S}$, and that (38) and its time derivative must therefore be evaluated at $r=b$. It is the presence of this factor which precludes the introduction of a non singular velocity $\boldsymbol{U}$ directly onto $r=a$. A large number of the "coupling integrals" appearing on the right of (46) and (47) have been evaluated (Scott, 1965 b).

Let us introduce column vectors of the form

$$\mathscr{X}=\begin{pmatrix}A_{l,u}\\B_{l,u}\end{pmatrix},\quad \mathscr{U}=\begin{pmatrix}C_{n,w}\\D_{n,w}\\E_{n,w}\\F_{n,w}\end{pmatrix}, \tag{48}$$

(listing A_l as $A_{l,0}$ etc). Then equations (46) and (47) may be written as

$$\sum_{n,w}\mathscr{C}_{(l,u)(n,w)}\,\mathscr{U}_{(n,w)}=\dot{\mathscr{X}}_{(l,u)}, \tag{49}$$

where $\mathscr{C}$ denotes "the coupling matrix". This matrix is linear in $\mathscr{X}$. Since the hydromagnetic approximation is a constraint, the matrix $\mathscr{C}$ is singular for all $\mathscr{X}$; solutions to (49) are only possible when the (secular variation) column vector $\dot{\mathscr{X}}$ on the right is compatible with $\mathscr{C}$, and the solution then obtained for the velocity coefficients is not then unique.

The structure of (49) will be considered in more detail in the second paper of this series. We offer here only a few remarks. The coupling integrals appearing in (46) and (47) vanish unless the so-called triangle inequalities are satisfied (cf. e.g. Bullard and Gellman, 1954). This has the effect of limiting the exponent of a/b in the $\mathscr{C}_{(l,u)(n,w)}$ element of the coupling matrix by the value n. Remembering that $a/b\approx 2$, this has two important effects on (49). First, since $\dot{\mathscr{X}}$ is not dominated by high harmonics (despite the large size of the $\mathscr{C}$ entries having large n), we must conclude that the high harmonics of $\mathscr{U}$ are small,* i.e. the series

* We are here excluding from the discussion the singular part of $\mathscr{U}$ i.e. the part obtained by solving (49) for $\dot{\mathscr{X}}=0$. This singular part could clearly have any magnitude, and we would not know about it at the earth's surface. But this seems rather unlikely.

we obtain for $\boldsymbol{u}$ shculd be rapidly convergent. Second, small changes in $\mathscr{X}$ generate large changes in the elements of $\mathscr{C}$ of large n, and will therefore affect its compatibility properties significantly, but the corresponding changes in $\mathscr{Y}$ will be small. In other words, $\boldsymbol{u}$ is insensitive to appreciable (compatible) changes in $\mathscr{X}$. We return to these matters in the second paper of this series.

Brief reports of this work has already appeared (Yerkes Observatory report in Astrophysical Journal for 1964), and in a thesis (Scott, 1964). We should thank Dr. Vestine for his interest in our ideas when one of us (P.H.R.) explained them to him at the Conference on geophysical data handling held at Yorktown Heights, N.Y. in October 1962.

References

Allan, D.W. and Bullard, E.C. Rev. Mod. Phys. **30**, 1087. 1958.
Allan, D.W. and Bullard, E.C. Unpublished. 1965.
Alldredge, L.R. and Hurwitz, L. J. Geophys. Res. **69**, 2631. 1964.
Berkofsky, L. Report AFCRL-62-684. 1962.
Bullard, E.C. 1953. J. Geophys. Res. **58**, 277.
Bullard, E.C. and Gellman, H. Phil. Trans. Roy. Soc. Lond. **A, 247**, 213. 1954.
Chandrasekhar, S. Astrophys. J. **124**, 232-265. 1956.
Coulomb, J. Ann. Geophys, **11**, 80. 1955.
Fougere, P.F. J. Geophys. Res. **69**, 2641. 1964.
Fougere, P.F. and McClay, J.F. Geophys. Res. Paper 55. AFCRC, ASTIA document ADII7268 1957.
Hide, R. and Roberts, P.H. Chapter 1 of "Physics and Chemistry of the Earth," Vol. 4. Pergamon. 1961.
Lowes, F.J. and Herzenberg, A. Ann. Geophys. **14**, 526. 1958.
Lowes, F.J. and Runcorn, S.K. Phil. Trans. Roy. Soc. Lond. **A, 243**, 525. 1951.
Lowes, F.J. and Russell, L.M. to apear. 1965.
Roberts, P.H. unpublished Ph. D. Dissertation, Cambridge. 1954.
Roberts, P.H. Astrophys. J. **122**, 315. 1955.
Roberts, P.H. Ann. Geophys. **15**, 75. 1959.
Rochester, M.G. Phil. Trans. Roy. Soc. London. **A, 252**, 431. 1960.
Runcorn, S.K. Handbuch der Physik, 47, 498. 1956.
Scott, S. unpublished dissertation, Newcastle University. 1964.
Scott, S. to appear. 1965a.
Scott, S. to appear. 1965b.
Stewartson, K. Proc. Camb. Phil. Soc. **53**, 774. 1957.
Stewartson, K. J. Fl. Mech. 8, 82. 1960a.
Stewartson, K. Rev. Mod. Phys. **32**, 855. 1960b.
Tamao, T. Sc. Rep. Tâhuko Univ. (Ser. 5), Geophysics, **10**, 137. 1959.
Vestine, E.H. et al. Carnegie Institute Rublications, Washington, No.578. 1947.
Vestine, E.H. et al, J. Geomag. and Geoelec. **15**, 47-89. 1963.
Yukutake, T. Bull. Earthquake Res. Inst. **40**, 1. 1962.

SOME COMMENTS ON THE THEORY OF HOMOGENEOUS DYNAMOS

R. D. GIBSON AND P. H. ROBERTS
University of Newcastle upon Tyne

1
THE DYNAMO PROBLEM

'The dynamo problem' is the search for answers to the questions, "Can a steady magnetic field be entirely maintained by a (fixed) volume, V, of conducting fluid through motions taking place in its interior?" and, "If so, how?" The geophysical and astrophysical interest in these questions is considerable. For example, if the answer to the former is "Yes", it provides the most satisfactory explanation of the geomagnetic field. The largest free decay time of a current system in a conducting sphere of radius a (m) surrounded by an insulator is $\tau_\eta = a^2/\pi^2\eta$, where

$$\eta = \frac{1}{\mu\sigma} \qquad \ldots [1]$$

is the magnetic diffusivity (m^2/s), μ is the magnetic permeability (henries/m) and σ is the electrical conductivity (mho/m). Taking this as a model of the earth's core ($a = 3{\cdot}48 \times 10^6$ m), and adopting the values $\mu = 4\pi \times 10^{-7}$ H/m and $\sigma = 3{\cdot}10^5$ mho/m, we obtain $\tau_\eta = 15{,}000$ years. It is doubtful whether this estimate is in error by more than a factor of 10. Paleomagnetism has shown, however, that the geomagnetic field has persisted at much the same strength for millions of years. There must, therefore, be some agency operating which maintains the field against ohmic losses. Now, it has been shown that quite small concentrations of dissolved radioactivity in the earth's core would suffice to set it in motion, and it is natural to wonder whether the resulting flow would act as the moving parts of a dynamo which maintains the geomagnetic field over times long compared with τ_η.

Physically a dynamo process may be described by the following sequence: the steady flow, u, of fluid across a field, B, sets up volume charges, ρ, in V, surface charges, χ, on S, the surface of V, and hence an electrostatic field $E = -\text{grad}\ \Phi$, where Φ is the (single-valued) electrostatic

potential. This electric field together with the induced electric field, $u \times B$, creates volume currents

$$j = \sigma(E + u \times B) = \sigma(-\text{grad } \Phi + u \times B), \qquad \ldots [2]$$

by Ohm's law (in the form appropriate to moving conductors). The question we ask is whether circumstances exist in which the current j produced from B by this induction process can generate, by Ampère's law, the very field B which gave it birth. In this light, the process may seem suspiciously like 'hoisting oneself up by his own bootstraps'. It must be emphasized, nevertheless, that the scheme is, in no sense whatever, a perpetual motion machine: all energy dissipated ohmically by the fluid is made good by the working of u against the Lorentz force $j \times B$. We shall, however, not consider the fluid dynamics of the situation at all, and will consider u to be *given*. In defence of this extreme attitude, we urge that the difficulty in answering the electrodynamic problems raised are so severe that we must not allow ourselves to be fettered by hydrodynamic misgivings. We will, however, make some concessions to the fluid dynamics. We will require that the velocity gradients $\partial u_i/\partial x_j$ are everywhere finite,† and we will exclude flows which, for finite mass density, necessarily require sources or sinks. For the same reason, we insist that

$$u_n = 0 \text{ on } S. \qquad \ldots [3]$$

Some simplifications can occur if u obeys the no-slip conditions on S, and we will therefore sometimes use the condition

$$u = 0 \text{ on } S \qquad \ldots [4]$$

in place of [3].

By [2] and Ampère's law

$$\mu j = \text{curl } B, \qquad \ldots [5]$$

we have

$$\eta \text{ curl } B = E + u \times B = -\text{grad } \Phi + u \times B, \text{ in } V, \qquad \ldots [6]$$

whence, using

$$\text{div } B = 0, \qquad \ldots [7]$$

we obtain

$$\eta \nabla^2 B = -\text{curl } (u \times B), \text{ in } V. \qquad \ldots [8]$$

(In [5] and elsewhere in this paper, we assume that μ has the same value everywhere.) We will suppose that the exterior, $\hat{V}$, of the moving region,

† For this reason, we exclude dynamos of the type proposed by Davis (1958).

V, is insulating, and we will use $\hat{}$ to distinguish quantities in $\hat{V}$ from their corresponding values in V. By [5] we have

$$\text{curl } \hat{B}=\hat{j}=0, \text{ in } \hat{V}, \qquad \ldots [9]$$

$$\nabla^2\hat{\Phi}=-\text{div } E=-\hat{\rho}/\hat{\varepsilon}=0, \text{ in } \hat{V}. \qquad \ldots [10]$$

It should be clear, from the earlier discussion, that we are interested only in self-excited dynamos, i.e. those in which the currents generating B (and the charges generating Φ) lie entirely in V. It then follows that

$$\hat{B}=0(r^{-3}),\ \Phi=0(r^{-1}),\ r\to\infty \text{ in } \hat{V}, \qquad \ldots [11a, b]$$

where r denotes distance from some origin in V. At the interface S we require that

$$\langle B_n\rangle=\langle n\times B\rangle=\langle\Phi\rangle=0, \qquad \ldots [12a, b, c]$$

where $\langle\Phi\rangle$ denotes $\Phi-\hat{\Phi}$ evaluated on S. The second of these, together with [5], ensures that $j_n=0$ on S. When Φ has been determined it is always possible, by the uniqueness theorem of potential theory, to find a $\hat{\Phi}$ satisfying [10], [11b] and the last of [12]. In other words, the last of [12] is not a constraint on the solution: we can approach the dynamo problem from B alone confident of being able to return to [6], [10], [11b] and [12c] later, to determine Φ and $\hat{\Phi}$.

According to [7], [9] and [11a] there exists an $\hat{\Omega}$ such that

$$\hat{B}=\text{grad } \hat{\Omega};\ \nabla^2\hat{\Omega}=0;\ \hat{\Omega}=0(r^{-2}),\ r\to\infty. \qquad \ldots [13]$$

When B has been determined it is possible (again by the uniqueness theorem) to find an $\hat{\Omega}$ satisfying [13] and [12a] *or* [12b]. Whichever of these two conditions then remains unsatisfied acts as a constraint to the solution of [7] and [8] in V. If we insist, as we must, that B is non-singular in V, this constraint cannot in general be satisfied, except by the trivial solution $B=0$. In fact, the dynamo problem has an eigenvalue character, as we shall be better able to appreciate presently.

The present paper is intended to review briefly some of the theoretical progress which has been made in answering the questions posed in this section, and to report a few minor points of interest which the present authors noted in the course of their study of them. Further discussion will be found in a book by Roberts (1967, Chapter 3).

2

Results by Cowling, Bullard and Gellman, Backus, and Herzenberg

The first major advance in understanding of fluid dynamos came from the pen of Cowling. In a celebrated paper (Cowling 1934) he showed

that there could be no axisymmetric solutions, B, of the dynamo equations. He also proved that, even in non-axisymmetric cases, some B structures could not arise in steady dynamos. His arguments are so well known that we feel it is pointless to repeat them here.

The next major step was taken by Bullard and Gellman (1954) who systematized dynamo theory for the case in which V is a sphere (radius a, say). They divided B into toroidal and poloidal parts:

$$B = \text{curl } T\mathbf{r} + \text{curl}^2 S\mathbf{r}, \qquad \ldots [14]$$

where $\mathbf{r}$ is the radius vector from the centre of the sphere, and further subdivided the toroidal and poloidal scalars (T and S) into their spherical harmonic components:

$$T = \sum_{n=1}^{\infty} \sum_{m=0}^{n} [T_n^{mc} r \cos m\phi + T_n^{ms} r \sin m\phi] P_n^m(\theta), \qquad \ldots [15]$$

$$S = \sum_{n=1}^{\infty} \sum_{m=0}^{n} [S_n^{mc} r \cos m\phi + S_n^{ms} r \sin m\phi] P_n^m(\theta). \qquad \ldots [16]$$

Assuming the fluid in V is of uniform density

$$\text{div } u = 0 \qquad \ldots [17]$$

they likewise divided u into toroidal and poloidal harmonics, t_n^{mc}, t_n^{ms}, s_n^{mc} and s_n^{ms}. Equation [8] was then replaced by an infinite set of ordinary, second order, differential equations for $T_n^{mc,s}$ and $S_n^{mc,s}$ or, abbreviating the notation, T_γ and S_γ. The term on the right of [8] gives rise to complications. The $u \times B$ interaction of an s_α motion (say) with an S_β field may give rise to an S_γ contribution to B provided a certain coupling integral, denoted by $(s_\alpha, S_\beta, S_\gamma)$, is non-zero, and similarly with other interactions $(t_\alpha, S_\beta, S_\gamma)$, etc. Bullard and Gellman gave selection rules by means of which $(s_\alpha, S_\beta, S_\gamma)$ etc. can be evaluated. These rules show, however, an unfortunate feature: starting with any harmonic S_α and any fluid motion, an infinite sequence of other field harmonics will be excited, in general. Bullard and Gellman prove that, if u is everywhere toroidal ($u_r \equiv 0$), the equations have no non-trivial solution, i.e. dynamo action is impossible. In other cases they could make no general statement, and had to resort to numerical methods. They concentrated their search on dynamos which might be supported by a $t_1 + s_2^{2c}$ flow, a motion which seemed not unrealistic geophysically. Their approach was to start from one harmonic of the field, S_1, and construct a diagram giving the interactions between this, the t_1 and s_2^{2c} motions, and the harmonics they induce. All of those fields with a harmonic number greater than N (say) were set zero, and a solution was sought for the resulting truncated equations.

They hoped to show that, as N was increased, these solutions would settle down to a definite structure and the corresponding eigenvalue would show every indication of converging to a fixed value. On both these counts, their results were unconvincing. For example, using dimensionless units in which the eigenvalue is the magnetic Reynolds number,

$$Rm = \frac{aU}{\eta}, \qquad \ldots [18]$$

where U is a typical fluid velocity, and scaling so that the surface of the sphere is $r=1$, they found that, when

$$s_2^{2c} = Rm\, r^2(1-r)^2,\ t_1 = 10Rm\, r^2 \qquad \ldots [19]$$

(a flow which satisfies [3]), the eigenvalue Rm was 22·06 provided the S_1, T_2, T_2^{2c} and T_2^{2s} harmonics alone were retained. When, however, they included all harmonics to degree 4, they obtained $Rm = 67{\cdot}4$.

No more convincing results have been obtained in an even simpler example, namely the effect of the s_2^{2c} motion of [19] alone. Commencing with a T_2 field, Bullard and Gellman found that, if the T_2 and T_2^{2c} fields alone are retained, $Rm = 124$. We have recently repeated Bullard and Gellman's calculations and have obtained $Rm = 125{\cdot}152$. On adding the S_3^{2s} component of B, we obtained $Rm = 502$, and keeping the next harmonics excited, viz. T_4, T_4^{2c} and T_4^{4c}, we found no eigenvalue between $Rm=0$ and $Rm=3000$. Moreover, the overall form of the solution, B, seemed to depend fundamentally on the level N of approximation selected. Bullard and Gellman had already noted that, for $N=2$, the solutions had a pathological structure which required careful handling. We, too, using Chebyshev expansions for the radial factors T_2, T_2^{2c}, etc., found that we required many terms in the series to represent the solutions adequately. For example, to obtain the value $Rm = 125{\cdot}152$ quoted above, we took series for T_2 and T_2^{2c} of 26 terms each. We were interested to note, however, that, even though many terms were needed to represent T_2 and T_2^{2c} accurately, it was not necessary to include so many if the eigenvalue Rm was all that was required. For example, taking only 16 terms in the series for T_2 and T_2^{2c} we obtained a value of 124 for Rm. (The Appendix below suggests a plausible reason for this insensitivity of Rm.)

From Bullard and Gellman's work it became clear that the outstanding analytical difficulty of proving that dynamo action is possible in a fluid sphere lay in the slow convergence of the solution on the level of truncation of the equations and the difficulty in handling more than a few of the terms of its harmonic expansion. What was required was a means of majorizing the dissipation produced by the higher harmonics in the

series, and of proving that, despite this ohmic drag on the dynamo, it would function on its lower harmonics (provided u was sensibly chosen).

Two very different and very ingenious methods of achieving this were devised independently and almost simultaneously by Herzenberg (1958) and Backus (1958). Each proved that fluid dynamos are possible. Backus's model arose from the observation that, in the absence of fluid motion, the (unwanted) high harmonics of B decay more rapidly than the low harmonics. If, then, u is assumed to be time-dependent, he could arrange that the 'contamination' of the high harmonics were reduced below any pre-assigned level by including sufficiently long periods of rest ($u = 0$) between sufficiently violent periods of regeneration ($u \neq 0$). Herzenberg's model arose from the observation that, when a volume L_1 (say) of moving fluid is surrounded by a large volume of stationary fluid of the same conductivity, the (unwanted) high harmonics of the field, B_1, it induces from an *applied* external field die away with distance from L_1 more rapidly than do the low harmonics. If, now, the applied field, B_2, is produced by a similar volume, L_2, at a great distance R from L_1, (and is induced by the motions in L_2 across the field B_1), it is conceivable that, if the flow in L_1 and L_2 is sufficiently vigorous and R is sufficiently great, the effect of the high harmonics can be minimized, and the dynamo will function essentially on its low harmonics alone. We discuss Herzenberg's dynamo further in the next section.

Unlike Bullard and Gellman's model, neither the mechanism of Backus nor that of Herzenberg can be regarded as plausible geophysically or astrophysically, but they provide existence proofs very fundamental and necessary to the subject.

3

Herzenberg's Dynamo and some Generalizations

In the first instance, it is perhaps best to think of the Herzenberg dynamo as consisting of an unbounded rigid conductor, V, having two spherical rotors L_1 and L_2, radii a_1 and a_2, centres O_1 and O_2 embedded within it, where $\overrightarrow{O_1O_2} = \boldsymbol{R}$, $(R > a_1 + a_2)$. The angular velocities of the rotors are ω_1 and ω_2, say. In order to make $\partial u_i/\partial x_j$ finite (see Section 1) we may imagine that there is a thin, highly sheared, film of fluid of the same conductivity between each rotor and its surroundings. The object is to produce (by the rotation of L_2 in a field B_1) a field B_2, and to arrange matters so that L_1 (by its rotation in the field B_2) creates the very field B_1 that L_2 requires. We shall find that the condition required for the system to act as a dynamo is that a magnetic Reynolds number defined by

$$Rm = \sqrt{\left[\left(\frac{\omega_1 a_1^2}{\eta}\right)\left(\frac{a_1}{R}\right)^3\left(\frac{\omega_2 a_2^2}{\eta}\right)\left(\frac{a_2}{R}\right)^3\right]} \qquad \ldots [20]$$

attains a certain critical value. In arranging this, we may suppose that

$$R >> a_1,\ a_2 \text{ and } \eta << a_1^2\omega_1,\ a_2^2\omega_2 \qquad \ldots [21a, b]$$

but there should be no implication that a limit is actually taken. By assuming [*21*], the effects of the high harmonics can be made arbitrarily small.

Let B_2^a be that part of B_2 which at O_1 is asymmetric with respect to the direction ω_1. Due to the rotation ω_1, the rotor L_1 will induce from B_2^a a field B_1^* (say). Now in a frame fixed to the surface of L_1 as it rotates, B_2^a appears as a field oscillating with frequency $2\pi/\omega_1$. Then, by the usual skin depth arguments for solid conductors, the currents induced by B_2^a will penetrate into L_1 no further than a depth of order $(\eta/\omega_1)^{\frac{1}{2}}$, a distance small compared with a_1 (cf. [*21b*]). At depths greater than this, L_1 will be shielded from B_2^a, i.e. the net field $B_2^a + B_1^*$ will be essentially zero. Thus $|B_2^a|$ and $|B_1^*|$ will be comparable on and near L_1. This strongly indicates that (for given B_2^a) B_1^* tends to a limit as $\omega_1 \to \infty$, and indeed this can be proved to be the case (Bullard, 1949). On the other hand, B_2^s, the symmetric part of B_2 at O_2 with respect to the direction ω_1 induces a field $B_1^\dagger$ which is proportional to ω_1, for all ω_1. Thus, for fixed $|B_2^a| / |B_2^s|$, we can, by making ω_1 sufficiently great (cf. [*21b*]), make $|B_1^*| / |B_1^\dagger|$ arbitrarily small. It is, therefore, necessary to consider only the parts of B_1 and B_2 which are symmetric with respect to ω_2 and ω_1, respectively. We now evaluate the field $B_1^\dagger$ created by L_1 from B_2^s.

Let us expand B_2 in a Taylor series about O_1:

$$\begin{aligned} B_{2i}(\boldsymbol{r}) &= B_{2i}(O_1) + x_j\left(\frac{\partial B_{2i}}{\partial x_j}\right)_{O_1} + \ldots \\ &= B_{2i}(O_1) + \tfrac{1}{2}\left(\frac{\partial B_{2i}}{\partial x_j} + \frac{\partial B_{2j}}{\partial x_i}\right)_{O_1} x_j + \tfrac{1}{2}\left(\frac{\partial B_{2i}}{\partial x_j} - \frac{\partial B_{2j}}{\partial x_i}\right)_{O_1} x_j + \ldots \\ &= \frac{\partial}{\partial x_i}\left[B_{2j}(O_1)x_j + \tfrac{1}{2}\left(\frac{\partial B_{2j}}{\partial x_k}\right)_{O_1} x_j x_k + \ldots\right] \\ &\qquad + \mu[\{j_2(O_1) + \ldots\} \times \boldsymbol{r}]_i, \end{aligned} \qquad \ldots [22]$$

using [*5*]. In [*22*], $\boldsymbol{r}(= x_i)$ denotes the position vector from O_1. The fields induced by the asymmetric parts of B_2 are, as we have already seen, negligible in a first approximation. We need, therefore, include only the symmetric parts of j_2 in [*22*]. This, however, lies at $\boldsymbol{r}$ in the plane of $\boldsymbol{r}$ and $\boldsymbol{\omega}_1$, and its associated magnetic field is therefore (as [*22*] confirms) in the azimuthal direction with respect to $\boldsymbol{\omega}_1$, i.e. parallel to the motion

of the rotor at that point. It therefore produces no inductive effect whatever. We may, then, replace [22] by

$$B_2(r) = \text{grad}\left[B_{2j}(O_1)x_j + \tfrac{1}{2}\left(\frac{\partial B_{2j}}{\partial x_k}\right)_{O_1} x_j x_k + \ldots\right]. \qquad \ldots [23]$$

This may be written alternatively as

$$B_2(r) = a_1 \,\text{grad} \sum_{m=1}^{\infty} A_m^{(2)} \left(\frac{r}{a_1}\right)^m P_m(\cos\theta) \qquad \ldots [24]$$

$$+ \text{ asymmetric terms},$$

where (r, θ, ϕ) are spherical polar coordinates, centre O_1, axis ω_1. Comparing coefficients in [23] and [24] we obtain

$$A_1^{(2)} = \frac{\omega_{1i}}{\omega_1} B_{2i}(O_1),$$

$$A_2^{(2)} = a_1 \frac{\omega_{1i}\omega_{1j}}{2\omega_1^2} \frac{\partial B_{2i}(O_1)}{\partial x_j}, \qquad \ldots [25]$$

. . . . . .

Having decomposed B_2^s into individual harmonics as in [24], we next have to find the field induced by L_1 from each. This matter is dealt with by Herzenberg (1958; see also Bullard 1949) who shows that the field $B_{1m}^{\dagger}$ induced from the $A_m^{(2)}$ term of $B_2^{s\dagger}$ is in the ϕ-direction and is given by

$$B_{1m\phi}^{\dagger} =$$

$$-\frac{mA_m^{(2)}}{(2m+1)}\left(\frac{\omega_1 a_1^2}{\eta}\right)\left[\frac{1}{(2m-1)}\left(\frac{a_1}{r}\right)^m \frac{\partial P_{m-1}}{\partial\theta} - \frac{1}{(2m+3)}\left(\frac{a_1}{r}\right)^{m+2} \frac{\partial P_{m+1}}{\partial\theta}\right]. \qquad \ldots [26]$$

For $m \neq 1$, only the term in P_{m-1} is significant at large distances $r >> a_1$, and we have

$$B_{11\phi}^{\dagger} = -\frac{1}{5} A_1^{(2)} \left(\frac{a_1^2\omega_1}{\eta}\right)\left(\frac{a_1}{r}\right)^3 \sin\theta\cos\theta,$$

$$B_{12\phi}^{\dagger} \approx \frac{2}{15} A_2^{(2)} \left(\frac{a_1^2\omega_1}{\eta}\right)\left(\frac{a_1}{r}\right)^2 \sin\theta, \qquad \ldots [27]$$

. . . . . .

These results must now be returned to vector form independent of our choice of axes:

$$B_1^{\dagger} = -\frac{1}{5}a_1^3 A_1^{(2)}\left(\frac{a_1^2\omega_1}{\eta}\right)\frac{(\boldsymbol{\omega}_1 . \boldsymbol{r})(\boldsymbol{\omega}_1 \times \boldsymbol{r})}{\omega_1^2 r^5}$$

$$+\frac{2}{15}a_1^2 A_2^{(2)}\left(\frac{a_1^2\omega_1}{\eta}\right)\frac{(\boldsymbol{\omega}_1 \times \boldsymbol{r})}{\omega_1 r^3} + \ldots \qquad \ldots [28]$$

We may now use [25], with the rotor suffices exchanged, and [28] at $\boldsymbol{r} = \boldsymbol{R}$, to evaluate $A_m^{(1)}$:

$$A_1^{(1)} = \left(\frac{a_1^2\omega_1}{\eta}\right)\left(\frac{a_1}{R}\right)^3 \frac{\boldsymbol{R}.(\boldsymbol{\omega}_1 \times \boldsymbol{\omega}_2)}{5\omega_1\omega_2 R}\left[\frac{(\boldsymbol{\omega}_1 . \boldsymbol{R})}{\omega_1 R}A_1^{(2)} - \frac{2R}{3a_1}A_2^{(2)}\right], \quad \ldots [29]$$

$$A_2^{(1)} = \frac{a_2}{a_1}\left(\frac{a_1^2\omega_1}{\eta}\right)\left(\frac{a_1}{R}\right)^3 \frac{\boldsymbol{R}.(\boldsymbol{\omega}_1 \times \boldsymbol{\omega}_2)}{5\omega_1\omega_2 R}\left[\frac{1}{2}\left\{\frac{\boldsymbol{\omega}_1 . \boldsymbol{\omega}_2}{\omega_1\omega_2} - \right.\right.$$

$$\left.\left.\frac{5(\boldsymbol{\omega}_1 . \boldsymbol{R})(\boldsymbol{\omega}_2 . \boldsymbol{R})}{\omega_1\omega_2 R^2}\right\}\frac{a_1}{R}A_1^{(2)} + \frac{(\boldsymbol{\omega}_2 . \boldsymbol{R})}{\omega_2 R}A_2^{(2)}\right]. \qquad \ldots [30]$$

In deriving [29] and [30] we have ignored the $m > 2$ contributions of $A_m^{(2)}$. These may readily be shown to be small compared with the terms retained in [29] and [30].

Let Θ_1 and Θ_2 be the angles between $\boldsymbol{\omega}_1$ and $\boldsymbol{R}$, and between $\boldsymbol{\omega}_2$ and $\boldsymbol{R}$, respectively. Let Ψ be the angle between the planes defined by $\boldsymbol{\omega}_1$ and $\boldsymbol{R}$ and by $\boldsymbol{\omega}_2$ and $\boldsymbol{R}$. Let

$$\Lambda_1 = \left(\frac{a_1^2\omega_1}{\eta}\right)\left(\frac{a_1}{R}\right)^3 \frac{\boldsymbol{R}.(\boldsymbol{\omega}_1 \times \boldsymbol{\omega}_2)}{5\omega_1\omega_2 R}, \quad \Lambda_2 = \left(\frac{a_2^2\omega_2}{\eta}\right)\left(\frac{a_2}{R}\right)^3 \frac{\boldsymbol{R}.(\boldsymbol{\omega}_1 \times \boldsymbol{\omega}_2)}{5\omega_1\omega_2 R},$$

$$\ldots [31]$$

$$p = \sin\Theta_1 \sin\Theta_2 \cos\Psi - 4\cos\Theta_1 \cos\Theta_2. \qquad \ldots [32]$$

Then [29] and [30] become

$$A_1^{(1)} = \Lambda_1\left[A_1^{(2)}\cos\Theta_1 - \frac{2R}{3a_1}A_2^{(2)}\right], \qquad \ldots [33]$$

$$\frac{2R}{a_2}A_2^{(1)} = \Lambda_1\left[pA_1^{(2)} + \frac{2R}{a_1}A_2^{(2)}\cos\Theta_2\right]. \qquad \ldots [34]$$

Similarly, interchanging rotor suffixes and the direction of $\boldsymbol{R}$,

$$A_1^{(2)} = \Lambda_2\left[-A_1^{(1)}\cos\Theta_2 - \frac{2R}{3a_2}A_2^{(1)}\right], \qquad \ldots [35]$$

$$\frac{2R}{a_1}A_2^{(2)} = \Lambda_2\left[pA_1^{(1)} - \frac{2R}{a_2}A_2^{(1)}\cos\Theta_1\right]. \qquad \ldots [36]$$

In matrix form [33]–[36] are

$$\begin{pmatrix} 1 & 0 & \Lambda_1 \cos\Theta_1 & -\frac{1}{3}\Lambda_1 \\ 0 & 1 & p\Lambda_1 & \Lambda_1 \cos\Theta_2 \\ -\Lambda_2 \cos\Theta_2 & -\frac{1}{3}\Lambda_2 & 1 & 0 \\ p\Lambda_2 & -\Lambda_2 \cos\Theta_1 & 0 & 1 \end{pmatrix} \begin{pmatrix} A_1^{(1)} \\ 2RA_2^{(2)}/a_2 \\ A_1^{(2)} \\ 2RA_2^{(2)}/a_1 \end{pmatrix} = 0. \quad \ldots [37]$$

For a non-trivial solution, the determinant of the 4×4 matrix appearing in [37] must vanish, i.e. multiplying out,

$$[\Lambda_1\Lambda_2 (\cos\Theta_1 \cos\Theta_2 - \sin\Theta_1 \sin\Theta_2 \cos\Psi) - 3]^2 = 0. \quad \ldots [38]$$

If we include the effects of higher harmonics of the field, [38] is replaced by

$$[\Lambda_1\Lambda_2 (\cos\Theta_1 \cos\Theta_2 - \sin\Theta_1 \sin\Theta_2 \cos\Phi) - 3]^2 = 0\left(\left(\frac{a_1}{R}\right)^2, \left(\frac{a_2}{R}\right)^2\right). \quad \ldots [39]$$

Before it can be concluded that dynamo action is possible if

$$\Lambda_1\Lambda_2 (\cos\Theta_1 \cos\Theta_2 - \sin\Theta_1 \sin\Theta_2 \cos\Psi) = 3,$$

i.e. if

$$Rm^2 \sin^2\Theta_1 \sin^2\Theta_2 \sin^2\Phi (\cos\Theta_1 \cos\Theta_2 - \sin\Theta_1 \sin\Theta_2 \cos\Psi) = 75, \quad \ldots [40]$$

it must be shown, as Herzenberg pointed out, that the term on the right of [39] is not negative definite. There are two main avenues of approach. The first, adopted by Herzenberg, is to remove the restriction that the conductor is unbounded, and to suppose instead that L_1 and L_2 are embedded in (say) a large sphere outside which an insulator lies. When allowance is made for the currents reflected from the surface of the sphere, Herzenberg found that, in some circumstances at least, the right of [39] could be made positive.

The second avenue of approach is to retain the infinite conductor but include the effects of more terms in the harmonic expansions of B_1 and B_2. This method has been adopted by Gibson (unpublished) for the case $a_1 = a_2 = a$ and $\omega_1 = \omega_2 = \omega$. Computing the right of [39] successively to order $(a/R)^2$, $(a/R)^4$ and $(a/R)^5$ he has shown that, at each level of approximation, the conditions for dynamo action may (as in [38]) be expressed as the vanishing of a perfect square. One is left with the distinct impression that Herzenberg's dynamo may not function in an infinite medium.

REFERENCES

BACKUS, G. 1958. *Ann. Phys.* **4**, 372.
BULLARD, E. C. 1949. *Proc. R. Soc. A*, **199**, 413.
BULLARD, E. C. and GELLMAN, H. 1954. *Phil. Trans. R. Soc. A*, **217**, 213.
COWLING, T. G. 1934. *Mon. Not. R. astr. Soc.* **94**, 39.
DAVIS, L. 1958. In *Electromagnetic Phenomena in Cosmical Physics*, ed. B. Lehnert. Cambridge University Press.
GIBSON, R. D. 1967. (To appear.)
HERZENBERG, A. 1958. *Phil. Trans. R. Soc. A*, **250**, 543.
INCE, E. L. 1926. *Ordinary Differential Equations*. Longmans.
ROBERTS, P. H. 1960. *J. math. Analysis Applic.* **1**, 195.
—— 1967. *An Introduction to Magnetohydrodynamics*. Longmans.

APPENDIX

THE ADJOINT DYNAMO PROBLEM

Associated with any characteristic value problem there is, as is well known, an adjoint system having the same characteristic values and having characteristic solutions orthogonal to those of the given system. The adjoint system and the given system together give rise to a variational method of determining the characteristic values (see Ince 1926, §9.31). In a previous paper (Roberts 1960) the adjoint of the dynamo problem has been given for a case involving a conductor infinite in spatial extent. In this appendix we give the adjoint in more general cases.

Let V be a region of fluid (diffusivity η), and let its exterior, $\hat{V}$, be a fixed solid (diffusivity $\hat{\eta}$). We suppose the fluid velocity, u, in V vanishes on S, the interface between V and $\hat{V}$. In $\hat{V}$, the field, $\hat{B}$, obeys [7] and Ohm's law for a stationary conductor:

$$\hat{\eta}\ \text{curl}\ \hat{B} = -\text{grad}\ \hat{\Phi}, \text{ in } \hat{V}. \qquad \ldots [A1]$$

Across S, [12] holds.

Take the scalar product of [6] with an arbitrary solenoidal vector $B^\dagger$ and integrate over V:

$$\begin{aligned} 0 &= \int B^\dagger \,.\, \{\eta\ \text{curl}\ B - u\times B + \text{grad}\ \Phi\} dV \\ &= \int B \,.\, \{\eta\ \text{curl}\ B^\dagger + u\times B^\dagger + \text{grad}\ \Phi^\dagger\} dV \\ &\quad + \int [\eta(B\times B^\dagger) + \Phi B^\dagger - \Phi^\dagger B] \,.\, dS. \end{aligned} \qquad \ldots [A2]$$

where $\Phi^\dagger$ is arbitrary. In a similar way, assuming (cf. [11]) that

$$\hat{B}^\dagger = 0(r^{-2});\ \hat{\Phi}^\dagger = 0(r^{-1}),\ r\to\infty, \text{ in } \hat{V}, \qquad \ldots [A3]$$

we find from [A1] that

$$\begin{aligned} 0 &= \int \hat{B} \,.\, \{\hat{\eta}\ \text{curl}\ \hat{B}^\dagger + \text{grad}\ \Phi^\dagger\} d\hat{V} \\ &\quad - \int [\hat{\eta}(\hat{B}\times\hat{B}^\dagger) + \hat{\Phi}\hat{B}^\dagger - \hat{\Phi}^\dagger \hat{B}] \,.\, dS, \end{aligned} \qquad \ldots [A4]$$

the minus sign arising since the normal to S is drawn outwards from V. Adding [*A2*] and [*A3*] we have

$$\int B \,.\, \{\eta \text{ curl } B^\dagger + u \times B^\dagger + \text{grad } \Phi^\dagger\} dV$$
$$+ \int \hat{B} \,.\, \{\hat{\eta} \text{ curl } \hat{B}^\dagger + \text{grad } \hat{\Phi}^\dagger\} d\hat{V}$$
$$= -\int \langle \eta(B \times B^\dagger) + \Phi B^\dagger - \Phi^\dagger B \rangle \,.\, dS. \qquad \ldots [A5]$$

We now demand that $B_n^\dagger$, $\eta(n \times B^\dagger)$ and $\Phi^\dagger$ are continuous across S. The surface integral on the right of [*A5*] then vanishes. The volume integral on the left vanishes for every eigenvector B. Assuming these form a complete set, we must therefore have

$$\eta \text{ curl } B^\dagger = -\text{grad } \Phi^\dagger - u \times B^\dagger, \text{ in } V, \qquad \ldots [A6]$$

$$^{\eta} \text{ curl } \hat{B}^\dagger = -\text{grad } \hat{\Phi}^\dagger, \text{ in } \hat{V}. \qquad \ldots [A7]$$

This, together with

$$\text{div } B^\dagger = 0 \qquad \ldots [A8]$$

and

$$\langle B_n^\dagger \rangle = \langle \eta(n \times B^\dagger) \rangle = \langle \hat{\Phi} \rangle = 0, \text{ on } S, \qquad \ldots [A9]$$

determines the adjoint differential system. Whenever the dynamo problem possesses a solution, so must its adjoint system [*A6*]–[*A9*]. It may be observed that the differential equations of the adjoint system are essentially the same as those of the given system, but with the velocity u reversed. The boundary conditions are, however, quite different.

When $\hat{V}$ is an insulator ($\hat{\eta} = \infty$), the analysis just given does not apply. Indeed, the cases $\hat{\eta} \to \infty$ and $\hat{\eta} = \infty$ are mathematically distinct. In the former, the equations (curl $^2\hat{B} = \text{div } \hat{B} = 0$) governing $\hat{B}$ are of higher order than those (curl $\hat{B} = \text{div } \hat{B} = 0$) obtaining in the latter. Correspondingly, $\hat{B}$ must, in the case $\hat{\eta} \neq \infty$, satisfy one more independent continuity condition (viz. four) on S than are necessary in the case $\hat{\eta} = \infty$. (Another minor difference may be noted: in the case $\hat{\eta} = \infty$, $\hat{B} = 0(r^{-3})$ at great distances, r, from V; in the case $\hat{\eta} \neq \infty$, on the other hand, $\hat{B} = 0(r^{-2})$, as $r \to \infty$.)

Unfortunately, we have not yet succeeded in finding the general form of the conditions which replace [*A9*] in the case $\hat{\eta} = \infty$. We have, however, obtained them in the special case in which S is a sphere ($r = a$). When $\hat{B}$ is divided into toroidal and poloidal components, as described in [*14*] to [*16*] above, it is easily shown from [*12a, b*] and [*13*] that

$$T_n^m = \frac{dS_n^m}{dr} + \frac{(n+1)}{a} S_n^m = 0, \text{ or } r = a. \qquad \ldots [A10]$$

If we introduce defining scalars $T_n^{m\dagger}$ and $S_n^{m\dagger}$ for the adjoint field $B^\dagger$, we

readily find, by arguments similar to those given above, that $B^\dagger$ obeys [*A6*] and that

$$S_n^{m\dagger} = \frac{dT_n^{m\dagger}}{dr} + \frac{(n+1)}{a} T_n^{m\dagger} = 0, \text{ on } r = a. \qquad \ldots [A11]$$

These equations bear a remarkable resemblance to [*A10*]. Indeed, [*A10*] expresses the fact that

$$\langle T \rangle = \langle S \rangle = \left\langle \frac{\partial S}{\partial r} \right\rangle = 0, \text{ on } S, \qquad \ldots [A12]$$

where

$$\hat{T} = \nabla^2 \hat{S} = 0, \text{ in } \hat{V}. \qquad \ldots [A13]$$

In a similar way [*A11*] shows that

$$\langle S^\dagger \rangle = \langle T^\dagger \rangle = \left\langle \frac{\partial T^\dagger}{\partial r} \right\rangle = 0, \text{ on } S, \qquad \ldots [A14]$$

where

$$\hat{S}^\dagger = \nabla^2 \hat{T}^\dagger = 0, \text{ in } \hat{V}. \qquad \ldots [A15]$$

Finally, we should note that, writing $\mathbf{u} = U\mathbf{v}$, where $\mathbf{v}$ is dimensionless, we have

$$Rm = \frac{Ua}{\eta} = \frac{\int B^\dagger . \operatorname{curl} B \, dV}{\int B^\dagger . (v \times B) \, dV}. \qquad \ldots [A16]$$

This expression has a variational significance. If first order errors δB and $\delta B^\dagger$ are made in B and $B^\dagger$, the corresponding first order error δRm in Rm is zero. This, in fact, appears to be the reason why the values of Rm obtained in Section 2 were, for any given N, relatively insensitive to the number of terms included in the representation series.

Reprinted from:
1968, *Phys. Earth Planet. Interiors* **1**, 288–296. North-Holland Publishing Company, Amsterdam
Printed in the Netherlands

NEARLY SYMMETRIC HYDROMAGNETIC DYNAMOS

J. G. TOUGH and P. H. ROBERTS

School of Mathematics,
University of Newcastle upon Tyne, England

Received 16 November 1967

Most theoretical studies of homogeneous dynamos are concerned with the kinematic problem. This consists in selecting a flow for the body of conducting fluid concerned (e.g. the Earth's core), and attempting to determine whether it can maintain a magnetic field indefinitely against ohmic losses. In this paper, the (more difficult but more realistic) hydromagnetic dynamo is studied, i.e. starting from a selected body force, an attempt is made to solve the equations of fluid motion and electromagnetic induction simultaneously and obtain a permanent magnetic field. It is shown that certain remarkable simplifications, which Braginskiĭ discovered for a class of kinematic dynamos in which the flow is almost axially symmetric, occur also for a class of hydromagnetic dynamos in which the fluid body is rapidly rotating.

1. Introduction: review of Braginskiĭ's work

The Earth has succeeded in maintaining its magnetic field over geological times, despite the fact that its high internal temperature rules out sufficient permanent magnetism within it, and despite the fact that the free decay times of current systems, $\boldsymbol{j}$, within it are probably less than 10^4–10^5 years. The most plausible explanation of this permanence of field is provided by the dynamo theory, in which it is supposed that motions in the Earth's core regenerate the geomagnetic field by electromagnetic induction in much the same way as a commercial self-excited dynamo does. The man-made dynamo differs, however, from the Earth in that it is deliberately constructed in a multiply-connected and highly asymmetric way. The Earth, on the other hand, is, through the action of gravity, simply-connected and almost spherically symmetric. These differences are so important that, in fact, they rule out simple dynamos entirely, and greatly hamper the search for a plausible working dynamo of any description.

The dynamo problem exists at several levels of difficulty. The simplest is *the kinematic dynamo problem*. Here we specify the fluid velocity, $\boldsymbol{u}$, in some reasonable way, and ask whether it can support a magnetic field, $\boldsymbol{B}$, which does not tend to zero as the time, t, tends to infinity. This procedure has the mathematical advantage of leading to a linear differential system for $\boldsymbol{B}$. It has the disadvantage of leaving out the reaction of the field on the flow. This is brought out clearly by the observation that, if we find a flow $\boldsymbol{u}$ which maintains a field $\boldsymbol{B}$ of constant amplitude, then by increasing $\boldsymbol{u}$ by some constant factor we could, presumably, cause $\boldsymbol{B}$ to grow without limit! This absurdity highlights the importance of *the hydromagnetic dynamo problem*. In this, we specify $\boldsymbol{F}$, the body force per unit mass acting on the fluid, in some reasonable way, and ask whether we can solve simultaneously the induction equation and the Navier-Stokes equation to obtain a flow $\boldsymbol{u}$ and a field $\boldsymbol{B}$ which do not tend to zero as t tends to infinity. Starting from a state of rest, we might expect that, as $\boldsymbol{u}$ increased under the action of $\boldsymbol{F}$, a stray field $\boldsymbol{B}$ might grow. Now, however, the quadratically increasing Lorentz force, $\boldsymbol{j}\times\boldsymbol{B}$, would (Lenz's law) tend to reduce $\boldsymbol{u}$ and eventually bring the growth of $\boldsymbol{B}$ to a halt. Although this is far more realistic than the kinematic approach, it is also far more formidable, for one is required to solve a pair of coupled nonlinear vector equations, viz. the induction equation and the Navier-Stokes equation. The objective of this paper is to demonstrate, however, that recent methods, devised for the kinematic dynamo by BRAGINSKIĬ (1964, a, b), can be successfully generalized to the hydromagnetic dynamo.

Braginskiĭ's approach to the kinematic dynamo might be summarized in the following way. It is well

known that axisymmetric dynamos do not exist, i.e. it is impossible to maintain an axially symmetric $\boldsymbol{B}$ indefinitely against ohmic decay, (COWLING, 1933). Indeed BACKUS (1957) has shown that one cannot, by any physically possible fluid motions, lengthen the free decay time in a sphere by even a factor of 4. Nevertheless, there is no reason why dynamos should not exist which are "nearly" axisymmetric and one would expect that, as the degree of axisymmetry is increased, the vigour of the fluid motions, as measured by the magnetic Reynolds number, R, required to maintain the dynamo would have to be increased also. Braginskiĭ was led therefore to consider a model in which the degree of asymmetry became small and R simultaneously became large. The mathematical advantage of considering the case $R \to \infty$ is, of course, that the problem is then of a boundary layer type. That is, the field $\boldsymbol{B}$ may be thought of as consisting of (i) an interior field far from the boundary S, of the fluid in which $\nabla = \mathrm{O}(1)$ and in which the effects of ohmic dissipation can, in the first approximation, be neglected ($R = \infty$), and (ii) a boundary layer (electromagnetic skin) on S in which the interior field adjusts to the potential field in the surrounding insulator.

Braginskiĭ discovered that the approach outlined above led to a remarkably simple theory for the symmetric components of field (*see also* TOUGH, 1967). To appreciate this simplification, it is necessary to state the details of his model with more precision (at the same time taking the opportunity to develop a notation). The primary flow pattern he assumed was of the form

$$\boldsymbol{u} = \boldsymbol{u}_0 + \boldsymbol{u}_1, \tag{1}$$

where $\boldsymbol{u}_1$ was purely asymmetric and $\boldsymbol{u}_0$ purely axisymmetric and, in fact, in the azimuthal (ϕ) direction. [(ϖ, ϕ, z) will be cylindrical coordinates.] The magnitude of u_0 is used to define R, and it is supposed that $u_1/u_0 \to 0$, as $R \to \infty$. Indeed, the relevant order of magnitude is shown, by Braginskiĭ, to be $u_1 = \mathrm{O}(R^{-\frac{1}{2}}u_0)$. To leading order, the field far from the boundary layers, is

$$\boldsymbol{B} = \boldsymbol{B}_0 + \boldsymbol{B}_1, \tag{2}$$

where $\boldsymbol{B}_1$ is purely asymmetric and $\boldsymbol{B}_0$ is purely azimuthal. This field satisfies the induction equation far from the boundary layers by being *aligned*, i.e.

$$(\boldsymbol{u}_0 + \boldsymbol{u}_1) \times (\boldsymbol{B}_0 + \boldsymbol{B}_1) = 0, \tag{3}$$

which, to the order $R^{-\frac{1}{2}}$ to which we are working, reduces to

$$\boldsymbol{u}_0 \times \boldsymbol{B}_1 + \boldsymbol{u}_1 \times \boldsymbol{B}_0 = 0, \tag{4}$$

or

$$\boldsymbol{B}_{1p} = \frac{B_0}{u_0}\boldsymbol{u}_{1p}, \tag{5}$$

where the suffix p denotes meridional components. The ϕ-component of $\boldsymbol{B}_1$ can be determined from eq. (5) and

$$\nabla \cdot \boldsymbol{B} = 0, \tag{6}$$

giving

$$B_{1\phi} = -\varpi \nabla \cdot \hat{\boldsymbol{B}}'_{1p} = -\varpi \nabla \cdot (B_0 \hat{\boldsymbol{v}}'_{1p}), \tag{7}$$

where

$$\boldsymbol{v}' = \frac{\boldsymbol{u}'}{u_0}. \tag{8}$$

The superfix $\wedge$ in eq. (7) denotes the operation of integration with respect to ϕ holding the unit vectors $\boldsymbol{I}_\varpi$ and $\boldsymbol{I}_\phi$ constant, the prime indicating there, and later, that the quantity concerned is purely asymmetric. It is clear that $B_1 = \mathrm{O}(R^{-\frac{1}{2}}B_0)$. The result of substituting the flow (1) into the induction equation

$$\frac{\partial \boldsymbol{B}}{\partial t} = R \operatorname{curl}(\boldsymbol{u} \times \boldsymbol{B}) + \nabla^2 \boldsymbol{B} \tag{9}$$

is, of course, to generate an infinite series expansion for $\boldsymbol{B}$

$$\boldsymbol{B} = \sum_{n=0}^{\infty} \boldsymbol{B}_n, \tag{10}$$

where $B_n = \mathrm{O}(R^{-n/2}B_0)$. As far as the first two terms (given in eq. (2)) in this expansion are concerned, there is no dynamo problem, since eq. (9) reduces (in essence) to the form (3), and ohmic effects do not enter. The problem of regeneration arises first for the term $\boldsymbol{B}_2$, which, it turns out, has an axisymmetric part in meridian planes.

It is convenient to re-order the terms in eq. (10) as

$$\boldsymbol{B} = B_\phi \boldsymbol{I}_\phi + \boldsymbol{B}' + \boldsymbol{B}_p, \tag{11}$$

where B_ϕ and $\boldsymbol{B}_p$ contain all axisymmetric contributions, according as to whether they are azimuthal or meridional, and $\boldsymbol{B}'$ contains all asymmetric terms, headed by $\boldsymbol{B}'_1$.

The term $B_\phi \boldsymbol{I}_\phi$ is headed by $\boldsymbol{B}_0$; the term $\boldsymbol{B}_p$ is headed by the symmetric part of $\boldsymbol{B}_2$ and is therefore $O(R^{-1}B_0)$. Since $\boldsymbol{B}_p$ is axisymmetric and meridional, it may be written in cylindrical polar components as

$$\boldsymbol{B}_p = \left[-\frac{\partial A}{\partial z}, 0, \frac{1}{\varpi}\frac{\partial}{\partial \varpi}(\varpi A) \right]. \tag{12}$$

Before proceeding we may remark that no difficulty arises in generalizing $\boldsymbol{u}$ from the form (1) to

$$\boldsymbol{u} = u_\phi \boldsymbol{I}_\phi + \boldsymbol{u}' + \boldsymbol{u}_p, \tag{13}$$

where $u' = O(R^{-\frac{1}{2}}u_0)$ and $u_p = O(R^{-1}u_0)$. Similarly we can write

$$\boldsymbol{u}_p = \left[-\frac{\partial a}{\partial z}, 0, \frac{1}{\varpi}\frac{\partial}{\partial \varpi}(\varpi a) \right]. \tag{14}$$

On substituting the forms (11) and (12) into equation (9), remarkable simplifications result. Working to leading order, Braginskiĭ showed that new "effective" fields A_e and a_e arise naturally which, like the fields A and a they replace, are symmetric: they are given to leading order by

$$A_e = A + wB_0, \quad a_e = a + wu_0, \tag{15}$$

where

$$w = \frac{\varpi}{2}(\boldsymbol{v}_p' \times \hat{\boldsymbol{v}}_p')_\phi^A \tag{16}$$

and the superscript A denotes the average over ϕ. The equations governing the symmetric parts of the field now reduce to strikingly simple forms:

$$\frac{\partial A_e}{\partial t} + \frac{R}{\varpi}\boldsymbol{u}_{ep} \cdot \nabla(\varpi A_e) = \left(\nabla^2 - \frac{1}{\varpi^2}\right) A_e + B_\phi \Gamma, \tag{17}$$

$$\frac{\partial B_\phi}{\partial t} + R\varpi \boldsymbol{u}_{ep} \cdot \nabla\left(\frac{B_\phi}{\varpi}\right) = \left(\nabla^2 - \frac{1}{\varpi^2}\right) B_\phi + R\left[\nabla\left(\frac{u_\phi}{\varpi}\right) \times \nabla(\varpi A_e)\right]_\phi, \tag{18}$$

where the source Γ is, to leading order, given by

$$\Gamma = \frac{1}{\varpi}\left[\boldsymbol{v}_p' \times \hat{\boldsymbol{v}}_p' + \boldsymbol{v}_p' \times \frac{\partial_1 \boldsymbol{v}_p'}{\partial \phi}\right]_\phi^A + 2[\nabla_p(\varpi v_\varpi' + z v_z') \cdot \nabla_p v_z']^A. \tag{19}$$

Tough (1967) has shown that, provided A_e, a_e and Γ are suitably redefined, this simplification persists to the next order beyond the leading terms considered by Braginskiĭ. It may be noted that, in eqs. (17) and (18), in which t is scaled by the free decay time of the fluid mass, $\partial/\partial t$ is assumed to be O(1).

In this paper, we extend Braginskiĭ's methods to the hydromagnetic dynamo, on the assumption that the dynamics of the fluid are (like the Earth's core) dominated by a primary balance between Coriolis and Lorentz forces. We find that the effective quantities (15) are again relevant, and lead to considerable simplifications of the equations governing the axisymmetric fields.

2. The equations of motion

We suppose the fluid is incompressible and, for simplicity, suppose it is of uniform density, ρ, so that

$$\nabla \cdot \boldsymbol{u} = 0. \tag{20}$$

We refer all fields to a frame fixed in the fluid container, which is supposed to be rapidly rotating (angular velocity $\Omega \boldsymbol{I}_z$). The Navier-Stokes equations may then be written in the dimensionless form

$$R_d\left[\frac{\partial \boldsymbol{u}}{\partial t} + R\boldsymbol{u} \cdot \nabla \boldsymbol{u}\right] + \boldsymbol{I}_z \times \boldsymbol{u} = = -\nabla p + \boldsymbol{j} \times \boldsymbol{B} + E\nabla^2 \boldsymbol{u} + R^{-\frac{1}{2}}\boldsymbol{F}. \tag{21}$$

Here $\boldsymbol{j} = \nabla \times \boldsymbol{B}$, E is the Ekman number, and R_d is a Rossby number based on the "velocity of drift", $\eta/\mathscr{L}$, of lines of force:

$$E = \frac{\nu}{2\Omega\mathscr{L}^2}, \quad R_d = \frac{(\eta/\mathscr{L})}{2\Omega\mathscr{L}} = \frac{1}{R}\left(\frac{\mathscr{U}}{2\Omega\mathscr{L}}\right). \tag{22}$$

Here $\mathscr{L}$ is a typical dimension of the fluid, $\mathscr{U}$ is a typical value of u_ϕ, ν is the kinematic viscosity of the fluid and η is its magnetic diffusivity. As for the induction equation (9), length is measured in units of $\mathscr{L}$ and time in units of $\mathscr{L}^2/\eta$. Also we introduce $\mathscr{F}$ as a typical magnitude of the body force per unit mass, $\boldsymbol{F}$, and $\mathscr{B}$ as a typical magnitude of B_ϕ. For $\mathscr{U}$ and $\mathscr{B}$, we take

$$\mathscr{U} = \frac{\mathscr{L}\mathscr{F}^2}{4\Omega^2\eta}, \quad \mathscr{B} = \mathscr{L}\mathscr{F}\left(\frac{\mu\rho}{2\Omega\eta}\right)^{\frac{1}{2}}, \quad [\mu = \text{permeability}], \tag{23}$$

so that the magnetic Reynolds number, R, is

$$R = \frac{\mathscr{U}\mathscr{L}}{\eta} = \left(\frac{\mathscr{L}\mathscr{F}}{2\Omega\eta}\right)^2. \tag{24}$$

In the case of the Earth, R_d is probably of order 10^{-8} so that, even if R were as large as 10^3–10^4, R_d and RR_d would both be small. We will therefore neglect the term in square brackets in eq. (21), writing

$$\boldsymbol{I}_z \times \boldsymbol{u} = -\nabla p + \boldsymbol{j} \times \boldsymbol{B} + E\nabla^2 \boldsymbol{u} + R^{-\frac{1}{2}} \boldsymbol{F}. \qquad (25)$$

This approximation of slow "quasi-steady" motion implies that, on our $\mathscr{L}^2/\eta$ time-scale, the flow responds "instantaneously" to any change in the forces to which it is subjected.

It is very probably true that E is small in the core. This leads us to seek flows of boundary layer type consisting of, (i) an interior flow far from S in which $\mathbf{V} = \mathrm{O}(1)$ and in which the effects of viscous dissipation can, in the first few approximations to eq. (25), be neglected, and (ii) a boundary (Ekman) layer on S in which the interior flow adjusts to the no-slip conditions on S. We will, in fact, presently (§3) assume that $E = \mathrm{O}(R^{-2})$. The term $E\nabla^2 \boldsymbol{u}$ in eq. (25) will therefore not contribute to the first four terms in the expansion of $\boldsymbol{u}$ and $\boldsymbol{B}$ in the interior. In the first approximation, the Coriolis and Lorentz forces ($\boldsymbol{I}_z \times \boldsymbol{u}_0$ and $\boldsymbol{j}_0 \times \boldsymbol{B}_0$) are axisymmetric and, since they lie in meridian planes (i.e. since they are perpendicular to $\boldsymbol{u}_0$), they do no work. Any O(1) part of $R^{-\frac{1}{2}}\boldsymbol{F}$ is therefore conservative, and could be absorbed into the ∇p_0 term. There is, then, no loss of generality in supposing that $R^{-\frac{1}{2}}\boldsymbol{F}$ is $\mathrm{O}(R^{-\frac{1}{2}})$, and we have in fact chosen our scaling above in a "natural" way to bring this fact out. We may also note that the $\mathrm{O}(R^{-\frac{1}{2}})$ contribution to the Coriolis and Lorentz forces are purely asymmetric. Thus any O(1) axisymmetric part of $\boldsymbol{F}$ must again be conservative, and may therefore again be neglected without loss of generality. We now see that the simplest non-trivial assumption is that $\boldsymbol{F}$ is O(1) and is purely asymmetric:

$$\boldsymbol{F} = \boldsymbol{F}'. \qquad (26)$$

We now treat eq. (25) after the manner of Braginskiĭ. The technique is, by averaging over ϕ, to divide any equation considered into an asymmetric part and into an axisymmetric part which can itself be separated into meridional and azimuthal components. In the case of eq. (25) these symmetric parts are

$$-u_\phi \boldsymbol{I}_\varpi = -\nabla(p + \tfrac{1}{2}B^{A2}) - \frac{B_\phi^2}{\varpi}\boldsymbol{I}_\varpi + \boldsymbol{B}_p^A \cdot \nabla \boldsymbol{B}_p^A + \boldsymbol{S}_p^A + E(\nabla^2 \boldsymbol{u})_p, \qquad (27)$$

$$u_\varpi = \frac{1}{\varpi}\boldsymbol{B}_p^A \cdot \nabla(\varpi B_\phi) + S_\phi^A + E(\nabla^2 \boldsymbol{u})_\phi. \qquad (28)$$

Here $\boldsymbol{S}^A$ is the axisymmetric part of the Lorentz force arising from the asymmetric magnetic field:

$$\boldsymbol{S}^A = [\boldsymbol{j}' \times \boldsymbol{B}']^A = [(\nabla \times \boldsymbol{B}') \times \boldsymbol{B}']^A. \qquad (29)$$

The next step consists in approximating to eqs. (27) and (28) by retaining only the leading terms in their expansion in powers of $R^{-\frac{1}{2}}$. It can easily be seen that the first three terms appearing in eq. (27) are $\mathrm{O}(R^0)$, and dominate the remainder. It is therefore unnecessary in the first order to compute $\boldsymbol{S}_p^A$ which is $\mathrm{O}(R^{-1})$. We simply have

$$\left(u_\phi - \frac{B_\phi^2}{\varpi}\right)\boldsymbol{I}_\varpi = \nabla(p + \tfrac{1}{2}B_\phi^2). \qquad (30)$$

Since the gradient on the right has no z-component, it must depend on ϖ and t alone. Thus we recover the well known result

$$u_\phi = \frac{B_\phi^2}{\varpi} + \varpi g(\varpi, t), \qquad (31)$$

where g is, at this stage, arbitrary. Looking next at eq. (28), we see that all terms are $\mathrm{O}(R^{-1})$ except the last, which is smaller. It is therefore necessary to compute S_ϕ^A to order R^{-1}. To do so, we must first solve the equation for the asymmetric components obtained from eq. (25) by removing terms (27) and (28), giving

$$\boldsymbol{I}_z \times \boldsymbol{u}' = -\nabla p' + (\nabla \times \boldsymbol{B}') \times \boldsymbol{B}^A + (\nabla \times \boldsymbol{B}^A) \times \boldsymbol{B}' + \boldsymbol{S}' + E\nabla^2 \boldsymbol{u}' + R^{-\frac{1}{2}}\boldsymbol{F}', \qquad (32)$$

where $\boldsymbol{S}'$ is the asymmetric part of the Lorentz force:

$$\boldsymbol{S}' = (\nabla \times \boldsymbol{B}') \times \boldsymbol{B}' - \boldsymbol{S}^A. \qquad (33)$$

As with the axisymmetric equations (27) and (28), we can obtain an infinite sequence of equations from eq. (32) by comparing terms of equal order in $R^{-\frac{1}{2}}$ on each side. To obtain S_ϕ^A, we require the first of these, viz. that obtained from the terms of order $R^{-\frac{1}{2}}$:

$$\boldsymbol{I}_z \times \boldsymbol{u}_1' = -\nabla p_1' + (\nabla \times \boldsymbol{B}_1') \times \boldsymbol{B}_0 + (\nabla \times \boldsymbol{B}_0) \times \boldsymbol{B}_1' + R^{-\frac{1}{2}}\boldsymbol{F}', \qquad (34)$$

which, on eliminating p_1', using Braginskiĭ's first approximation to $\boldsymbol{B}_1'$ (cf. eq. (5) above) and eq. (31), ultimately reduces to the equations

$$\frac{1}{\varpi}\frac{\partial Q'}{\partial \varpi} + \varpi g L\left(1 + \frac{\partial^2}{\partial \phi^2}\right) v'_\varpi - \varpi^2 g \frac{\partial v'_z}{\partial z} = \frac{\varpi}{R^{\frac{1}{2}}}(\nabla \times \boldsymbol{F}')_z, \tag{35}$$

$$\frac{1}{\varpi}\frac{\partial Q'}{\partial z} + \varpi g L \frac{\partial^2}{\partial \phi^2} v'_z + \varpi^2 g \frac{\partial v'_\varpi}{\partial z} = -\frac{\varpi}{R^{\frac{1}{2}}}(\nabla \times \boldsymbol{F}')_\varpi, \tag{36}$$

where

$$Q' = \varpi^4 g L \nabla \cdot \boldsymbol{v}'_p, \quad L = -1 + \frac{u_\phi}{\varpi g}. \tag{37}$$

One of the main practical difficulties of the present theory is that of solving eqs. (35) and (36). We may observe, however, that, if we expand $\boldsymbol{F}'$ as

$$\boldsymbol{F}'(\varpi, \phi, z, t) = \sum_{m=1}^{\infty} \boldsymbol{F}'_m(\varpi, z, t)\, e^{im\phi}, \tag{38}$$

the $\boldsymbol{F}'_m$ component of $\boldsymbol{F}'$ is solely responsible for the corresponding $\boldsymbol{v}'_m$ component in the expansion of $\boldsymbol{v}'$. Considering one such m component alone, and omitting the suffix m for simplicity, we observe (after some reductions; see Appendix) that eqs. (35) and (36) may be solved in the form

$$v'_\varpi = \frac{L}{MR^{\frac{1}{2}}}\left[\frac{\partial D'}{\partial \varpi} + \left(\frac{M}{\varpi L} - \frac{(m^2-1)L}{\varpi}\right) D' - \frac{1}{g}(\nabla \times \boldsymbol{F}')_z\right], \tag{39}$$

$$v'_z = \frac{1}{m^2 L R^{\frac{1}{2}}}\left[\frac{\partial D'}{\partial z} + \frac{1}{g}(\nabla \times \boldsymbol{F}')_\varpi\right], \tag{40}$$

where

$$M = (m^2-1)L^2 - 1 + \frac{L}{\varpi g}\frac{\mathrm{d}}{\mathrm{d}\varpi}(\varpi^2 g), \tag{41}$$

and D' obeys the second order equation

$$\frac{1}{\varpi^3 g}\frac{\partial}{\partial \varpi}\left[\frac{\varpi^3 g L}{M}\frac{\partial D'}{\partial \varpi}\right] + \frac{1}{m^2}\frac{\partial}{\partial z}\left[\frac{1}{L}\frac{\partial D'}{\partial z}\right] - \frac{(m^2-1)}{\varpi^2}\left[\frac{L(1+L)}{M} + \varpi^2 \frac{\partial}{\partial \varpi}\left(\frac{L^2}{\varpi M}\right)\right] D' = \left[\frac{\partial}{\partial \varpi} + \frac{1+L}{\varpi L}\right]\left[\frac{L}{Mg}(\nabla \times \boldsymbol{F}')_z\right] - \frac{1}{m^2 g}\frac{\partial}{\partial z}\left[\frac{1}{L}(\nabla \times \boldsymbol{F}')_\varpi\right]. \tag{42}$$

It may be noted that

$$D' = R^{\frac{1}{2}}\varpi(v'_\varpi + \varpi L \nabla \cdot \boldsymbol{v}'_p),$$

though this fact is not directly useful in practice.

Clearly, once v'_ϖ and v'_z have been obtained from eqs. (39) to (42), we may, if we so desire, determine v'_ϕ from eq. (20) in the same way that $B_{1\phi}$ was obtained from eq. (6); cf. eq. (7).

We now return to the evaluation of S^A_ϕ. It may easily be shown, from eq. (29), that

$$S^A_\phi = \frac{1}{\varpi}[\hat{\boldsymbol{B}}' \cdot \nabla(\varpi^2 \nabla \cdot \boldsymbol{B}'_p)]^A. \tag{43}$$

In evaluating this to order R^{-1}, we may use eq. (5), giving

$$S^A_\phi = \frac{1}{\varpi}[\nabla \times (wB_\phi \boldsymbol{I}_\phi)] \cdot \nabla(\varpi B_\phi) + \frac{1}{\varpi^2}\left[\hat{\boldsymbol{v}}'_p \cdot \nabla\left(\varpi^4 B^2_\phi \nabla \cdot \left(\frac{\boldsymbol{v}'_p}{\varpi}\right)\right)\right]^A. \tag{44}$$

The last term on the right can, after considerable reductions (see Appendix), be transformed to give

$$S^A_\phi = \frac{1}{\varpi}[\nabla \times (wB_\phi \boldsymbol{I}_\phi)] \cdot \nabla(\varpi B_\phi) + \frac{\partial}{\partial z}(wu_\phi) + \Lambda, \tag{45}$$

where Λ is an energy source term given by

$$\Lambda = \frac{\varpi}{R^{\frac{1}{2}}}[\boldsymbol{v}' \times (\nabla \times \hat{\boldsymbol{F}}')]^A_\phi. \tag{46}$$

It is now apparent that the same effective quantities (15) which resulted in the remarkable simplifications to the induction equation discovered by Braginskiĭ are of equal relevance here for, on using eq. (45) to replace S^A_ϕ in eq. (28), we have

$$u_{e\varpi} = \frac{1}{\varpi}\boldsymbol{B}_{ep} \cdot \nabla(\varpi B_\phi) + \Lambda. \tag{47}$$

3. The boundary layers

We have already observed in §1 that, to leading order, the velocity and magnetic fields in the interior region far from the boundary layers are

$$\boldsymbol{u} = u_0 \boldsymbol{I}_\phi + \boldsymbol{u}_1, \quad \boldsymbol{B} = B_0 \boldsymbol{I}_\phi + \boldsymbol{B}_1, \tag{48}$$

and that, to order $R^{-\frac{1}{2}}$, this is an aligned flow which, when eq. (31) holds, obeys the steady equations of motion and electromagnetic induction for a highly-rotating *inviscid perfectly-conducting* fluid. It does not, however, in general satisfy the boundary conditions on

S, viz. the no-slip condition on $\boldsymbol{u}$ and the continuity of $\boldsymbol{B}$ to a suitably constructed potential field in the exterior insulator. A boundary layer, $\mathscr{S}$, therefore exists on S in which the adjustment of the interior flow to these conditions is effected.

The exact nature of $\mathscr{S}$ is difficult to assess immediately. Because of the differential rotation, the layer will have some similarity to an Ekman layer, familiar in the hydrodynamics of rotating systems. The role of the magnetic field is, however, more equivocal. We are familiar in magnetohydrodynamics with boundary layers of two main types. The first of these is the layer of the Hartmann type, associated particularly with crossed flows ($\boldsymbol{u}\times\boldsymbol{B}\neq 0$) in ducts. In this $\boldsymbol{n}\cdot\boldsymbol{B}$ is non-zero (except usually at isolated points on S) as the limit $\nu\to 0$, $\eta\to 0$ is taken. Like the Ekman layer, the Hartmann layer is local in the sense that the jump conditions across it are determined by the adjacent interior flow and not by the previous "history" of the flow. Also like the Ekman layer, the Hartmann layer controls the interior flow since it selects which, of an infinite sequence of possible flows satisfying the interior equations, is actually realized. The second type of magnetohydrodynamic boundary layer is the magnetic and viscous diffusion layer associated (mainly) with aligned flows: the magnetohydrodynamic Blasius flow is a familiar example (cf. e.g. ROBERTS, 1967, §6.3). This layer is historical, in the sense that local conditions in the interior are not sufficient to determine its structure; the conditions previously encountered by the flow enters the boundary layer description. It is also passive, in the sense that the interior flow can be determined uniquely as a first step, and the boundary layer may be matched to it later. In the present situation, it is not clear with which (if either!) of these cases we are dealing. For, when ν and η are finite, $\boldsymbol{n}\cdot\boldsymbol{B}$ $[=\boldsymbol{n}\cdot(\boldsymbol{B}'+\boldsymbol{B}_p)]$ is non-zero; but, as ν and η tend to zero, so does $\boldsymbol{n}\cdot\boldsymbol{B}$. In some sense, we are dealing with a triple limiting process, viz. $\nu\to 0$, $\eta\to 0$, $\boldsymbol{n}\cdot\boldsymbol{B}\to 0$!

Let $\boldsymbol{n}$ be the inward normal at a point, P, on S and let ϕ and θ be tangential coordinates at P, θ being the angle between $\boldsymbol{1}_z$ and $-\boldsymbol{n}$. Use n, ϕ and θ to distinguish components, of vectors and also of vector equations (see fig. 1).

By eq. (25, n) the total pressure, magnetic and mechanical, is constant across the boundary layer in the first approximation; also, we note that eq. (9, n) is essentially a consequence of eqs. (9, θ, ϕ) and the fact that $\boldsymbol{\nabla}\cdot\boldsymbol{B}=0$. We may therefore concentrate on eqs. (25, θ, ϕ) and eqs. (9, θ, ϕ). We observe that the $B_n\partial/\partial n$ part of the $\boldsymbol{B}\cdot\boldsymbol{\nabla}$ operator in these equations is, at most $O(R^{-\frac{1}{2}}B_0)$ times* its operand, since $\boldsymbol{n}\cdot\boldsymbol{B}=0$ to leading order. In fact, if the boundary layer is of

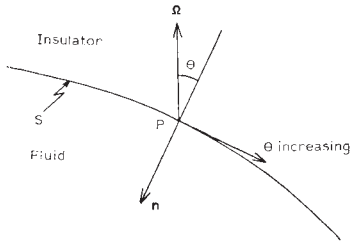


Fig. 1. The boundary layer coordinates.

Hartmann character, the $B_n\partial/\partial n$ terms in eq. (25, ϕ) and eq. (9, ϕ) must be of the same order as the diffusion terms, requiring

$$\frac{B_n}{\delta}B_\phi\approx\frac{E}{\delta^2}u_\phi,\quad \frac{B_n}{\delta}u_\phi\approx\frac{1}{R\delta^2}B_\phi, \tag{49}$$

where δ is the boundary layer thickness. This gives

$$\delta\approx\left(\frac{E}{RB_n^2}\right)^{\frac{1}{2}}, \tag{50}$$

i.e. if $B_n=O(R^{-\frac{1}{2}})$ on S, then $\delta=O(E^{\frac{1}{2}})$, and is greater otherwise. The thickness of this layer is of the same order as the hydrodynamic Ekman layer, in which the flow adjusts to the no-slip condition on S; this is seen by comparing the terms $\boldsymbol{1}_z\times\boldsymbol{u}$ and $E\nabla^2\boldsymbol{u}$ in eq. (25). Thus, if $B_n=O(R^{-\frac{1}{2}})$, the boundary layer must really be a composite Ekman-Hartmann layer. If, however, $B_n=O(R^{-1})$, for example, there must be a second (inner) boundary layer of Ekman character within the (outer) boundary layer just described.

The radial flow in or out of the Ekman layer is $O(u_\phi\delta)=O(E^{\frac{1}{2}}u_\phi)$, which will, in general, have an axisymmetric part. In the Braginskiĭ theory as we have

* A referee has asked us to emphasize that R has the same meaning as before, i.e. it is a magnetic Reynolds number based on u_ϕ, the azimuthal velocity, and $\mathscr{L}$, a typical dimension of the container.

developed it, however, the largest, symmetric, radial flow is $O(R^{-1}u_\phi)$. It follows that†

$$\frac{1}{RE^{\frac{1}{2}}} \to \text{finite limit, } \lambda, \text{ say,} \tag{51}$$

as $E \to 0$ and $R \to \infty$. In general one does not expect such a restriction to be imposed upon the analysis, i.e. it should be possible to let $\nu \to 0$, $\eta \to 0$ independently. We conclude, therefore, that if, for example, $ER \to$ non-zero limit as $E \to 0$ and $R \to \infty$, the resulting dynamo, if one exists, will not be of the Braginskiĭ type.

We can show that B_n' is not as large as $O(R^{-\frac{1}{2}})$ on S; for suppose that $B_n' = O(R^{-\frac{1}{2}})$ on S. Then by eq. (49) the jump in B_ϕ across $\mathscr{S}$ is $O(R^{-\frac{1}{2}})$. In other words, B_ϕ would in the interior region satisfy on S*, the edge of the boundary layer; the boundary condition required on S, viz. $B_\phi = 0$. In other words B_0/u_0 would be zero on S*. We have already seen, however, that $u_n' = O(R^{-1})$, at most, giving $B_{1n}'/u_{1n}' \to \infty$ on S*. This contradicts eq. (5). Thus B_n' vanishes to order R^{-1} or higher on S. In fact, we can now see that $\mathscr{S}$ has the double structure mentioned above, the inner layer being, to the first approximation, an *axisymmetric* Ekman layer. The radial influx or efflux created by u_ϕ is therefore axisymmetric also and supplies $\boldsymbol{u}_p$ in part or whole. The first non-vanishing component of $\boldsymbol{u}'$ on S* will be $\boldsymbol{u}_3'$ [$= O(R^{-\frac{1}{2}})$], at most. This supports Braginskiĭ's contention that $\boldsymbol{B}_3'$ is the largest asymmetric field escaping from the fluid.

Thus to summarize our findings: there is a double boundary layer on S consisting of (i) an inner layer, of thickness R^{-1} ($= E^{\frac{1}{2}}$), which is, to a first approximation, an axisymmetric Ekman layer. It is local and controls the flow in the outer boundary layer and interior region. The no-slip conditions are satisfied on S, its outer edge, but not on its inner edge. There is also (ii) a magnetic diffusion layer in which the aligned flow in the interior region adjusts to the potential field in the surrounding insulator. In this layer, which is passive and historical, the flow and field become, to some extent at least, crossed in direction. If we examine eq. (9, ϕ), however, we soon convince ourselves that there is no such boundary layer for B_ϕ; for the notion that the layer thickens by diffusion in the direction of field and flow (as in the magnetohydrodynamic Blasius situation) is inconsistent with its necessary periodicity in ϕ. The outer boundary is, then, associated with $\boldsymbol{B}'$ only. It is given presumably by a theory akin to that developed by Braginskiĭ [1964a, p. 733].

Since the inner boundary layer is an axisymmetric Ekman layer, we may apply the usual jump conditions on $\boldsymbol{n} \cdot \boldsymbol{u}$ across it, and thereby relate the axisymmetric u_n to the unknown function g:

$$u_n(\mathrm{S}^*) = \pm \frac{E^{\frac{1}{2}}}{r_l r_p} \frac{\partial}{\partial \theta} \left[\frac{r_l}{\sqrt{(\pm 2 \cos \theta)}} (u_\phi \pm u_\theta) \right]. \tag{52}$$

Here $r_l(\theta)$ and $r_p(\theta)$ are the radii of curvature of the line of latitude and the meridian section at the point of S concerned and upper signs are taken for $\cos\theta > 0$ and lower for $\cos\theta < 0$. In our application, we have $u_\theta = 0$ and

$$u_\phi = \varpi g(\varpi), \quad \text{on} \quad \mathrm{S}^*, \tag{53}$$

which leads to

$$u_n(\mathrm{S}^*) = \pm \frac{1}{\lambda R r_l r_p} \frac{\partial}{\partial \theta} \left[\frac{r_l \varpi g(\varpi)}{\sqrt{(\pm 2 \cos \theta)}} \right]. \tag{54}$$

For the particular case of a unit sphere, $\varpi = r_l = \sin\theta$ and $r_p = 1$, so that eq. (54) becomes

$$u_{cr}(\mathrm{S}^*) = \pm \frac{1}{\lambda R \sin\theta} \frac{\partial}{\partial \theta} \left[\frac{\sin^2 \theta g(\sin\theta)}{\sqrt{(\pm 2 \cos \theta)}} \right]. \tag{55}$$

(This result, and the boundary layer theory which give rise to it, break down of course near the equator. One may expect, however, that the boundary layer singularities in these regions will be passive.) Since $w = 0$ on S*, there is no difference between $\boldsymbol{u}_p$ and $\boldsymbol{u}_{cp}$ thereon. We have, therefore, used u_{cr} in (55) rather than u_r; here r denotes the radial component.

4. Conclusions

We now show that, with the derivation of condition (55), we have closed the equations governing the dynamo. At the same time, we outline a method for integrating them. We suppose A_e and B_ϕ are known at one time-step. If, then, we can show how to compute Γ, u_ϕ, and $\boldsymbol{u}_p$ from $\boldsymbol{F}'$, we can clearly determine A_e and B_ϕ at the next time-step by using eqs. (17) and (18).

† If we take $R = 10^4$ and $\lambda = 1$ this gives $\nu = 1 \cdot 8 \times 10^3$ m²/s. for the Earth's core ($\mathscr{L} \approx 3.5.10^6$ m, $2\Omega = 1.4.10^{-4}$ s^{-1}). This is rather large by usual geophysical reckoning, though well within the generous uncertainties familiarly quoted: it may imply, if taken seriously, that ν is a turbulent viscosity. In essense, (51) states that the Earth's core will not be a Braginskiĭ dynamo of the present type if its viscosity is "too large".

To determine Γ, u_ϕ, and $\boldsymbol{u}_p$, we imagine an iteration process in which g is assumed at the beginning of each step and is improved at its end. From g and B_ϕ, we can compute u_ϕ from eq. (31). Since $\boldsymbol{F}'$ is given, we now know all the coefficients appearing in eq. (42) for D'. This equation is hyperbolic near S and parabolic on S. Its solution may be considered, however, to consist of a particular integral and a complementary function. Since eq. (42) is of second order, there are two independent complementary functions, but one is singular on $\varpi = 0$ and has to be discarded. When the appropriate "multiple" of the other is added to the particular integral, we can satisfy the condition that $v_n' = 0$ on S, as we must. Having in this way determined $\boldsymbol{v}'$, the source fields Γ and Λ given by eqs. (19) and (46) can be evaluated. But, by eqs. (20) and (47), the boundary condition (55) can be written in integrated form as

$$\frac{2}{\lambda R}\frac{\sin^2\theta g(\sin\theta)}{\sqrt{(2\cos\theta)}} = \int_{z=-\cos\theta}^{z=\cos\theta}\left[\frac{1}{\varpi}\{\nabla(\varpi B_\phi)\times\nabla(\varpi A_e)\}_\phi + \varpi\Lambda\right]\mathrm{d}z, \quad (56)$$

the right-hand side of which is now known. We have, therefore, obtained a new value of $g(\varpi)$, from which we can commence the next step of the iteration. If, or when, this iteration has converged, we can compute $\boldsymbol{u}_{ep}$ directly from eqs. (20) and (47). [Indeed, by using our knowledge of $\boldsymbol{v}'$ we could compute w from eq. (16), and hence derive $\boldsymbol{u}_p$ and A, if we required them.] In other words, when $g(\varpi)$ has converged, we can obtain Γ and all components of the (effective) velocity at the time-step concerned, and can proceed to the next.

5. Appendix: detailed reductions

Derivation of equations (39) to (42). For the mth component of eq. (38), eqs. (36) and (37) become

$$\frac{\partial}{\partial z}\{\varpi g[\varpi(v_\varpi' + \varpi L\nabla\cdot\boldsymbol{v}_p')]\} - \varpi g L m^2 v_z' = -\frac{\varpi}{R^{\frac{1}{2}}}(\nabla\times\boldsymbol{F}')_\varpi,$$

i.e. defining D' by

$$D' = R^{\frac{1}{2}}\varpi(v_\varpi' + \varpi L\nabla\cdot\boldsymbol{v}_p'), \quad (A.1)$$

we obtain result (40). By eq. (A.1), eq. (35) may be written

$$\frac{1}{\varpi}\frac{\partial}{\partial\varpi}\left[\varpi^3 g\left(\frac{D'}{\varpi R^{\frac{1}{2}}} - v_\varpi'\right)\right] + \varpi g L(1-m^2)v_\varpi' - \varpi^2 g\left[\frac{1}{\varpi L}\left(\frac{D'}{\varpi R^{\frac{1}{2}}} - v_\varpi'\right) - \frac{1}{\varpi}\frac{\partial}{\partial\varpi}(\varpi v_\varpi')\right] = \frac{\varpi}{R^{\frac{1}{2}}}(\nabla\times\boldsymbol{F}')_z,$$

which reduces to

$$\frac{\varpi g}{R^{\frac{1}{2}}}\frac{\partial D'}{\partial\varpi} + (M-(m^2-1)L^2)\frac{g}{LR^{\frac{1}{2}}}D' - \frac{\varpi g M}{L}v_\varpi' = \frac{\varpi}{R^{\frac{1}{2}}}(\nabla\times\boldsymbol{F}')_z,$$

where M is defined by eq. (41). Regrouping terms leads to eq. (39).

On using eqs. (39) and (40) to eliminate v_ϖ' and v_z' from definition (A.1), we obtain

$$\frac{D'}{\varpi} = \varpi L\left[\frac{\partial}{\partial\varpi} + \frac{1+L}{\varpi L}\right]\times \times\left[\frac{L}{M}\left\{\frac{\partial D'}{\partial\varpi} + \left(\frac{M}{\varpi L} - \frac{(m^2-1)}{\varpi}L\right)D' - \frac{1}{g}(\nabla\times\boldsymbol{F}')_z\right\}\right] + \varpi L\frac{\partial}{\partial z}\left[\frac{1}{m^2 L}\left\{\frac{\partial D'}{\partial z} + \frac{1}{g}(\nabla\times\boldsymbol{F}')_\varpi\right\}\right],$$

which, after some rearrangement and cancellation, results in eq. (42).

Derivation of equation (45). We wish to transform the last term,

$$T = \frac{1}{\varpi^2}\left[\hat{\boldsymbol{v}}_p'\cdot\nabla\left\{\varpi^4 B_\phi^2\nabla\cdot\left(\frac{\boldsymbol{v}_p'}{\varpi}\right)\right\}\right]^A, \quad (A.2)$$

of eq. (44) to a more convenient expression. By eq.) we have

$$B_\phi^2 = \varpi(u_\phi - \varpi g) = \varpi^2 g L,$$

and, substituting into expression (A.2) and expanding, we obtain

$$T = \frac{1}{\varpi^2}[\hat{\boldsymbol{v}}_p'\cdot\nabla\{\varpi^5 g L\nabla\cdot\boldsymbol{v}_p' - \varpi^3(u_\phi-\varpi g)v_\varpi'\}]^A$$

$$= \frac{1}{\varpi^2}\left[\hat{v}_\varpi'\left\{\varpi\frac{\partial Q'}{\partial\varpi} + Q' - \varpi v_\varpi'\frac{\partial}{\partial\varpi}[\varpi^2(u_\phi-\varpi g)] - \right.\right.$$

$$\left.\left. - \varpi^3(u_\phi-\varpi g)\left(\nabla\cdot\boldsymbol{v}_p' - \frac{\partial v_z'}{\partial z}\right)\right\} +\right.$$

$$+\hat{v}_z'\left(\varpi\frac{\partial Q'}{\partial z}-\varpi^3(u_\phi-\varpi g)\frac{\partial v_\varpi'}{\partial z}-\varpi^3 v_\varpi'\frac{\partial}{\partial z}(u_\phi-\varpi g)\right)\Bigg]^A. \tag{A.3}$$

Now, for any asymmetric scalar fields F_1' and F_2', we have

$$[F_1'\hat{F}_1']^A=0,\quad [F_1'\hat{F}_2']^A=-[\hat{F}_1'F_2']^A. \tag{A.4}$$

On applying the first of these to v_ϖ' and v_z', we find that eq. (A.3) may be written in the form

$$T=\left[\hat{v}_\varpi'\left(\frac{1}{\varpi}\frac{\partial Q'}{\partial\varpi}+\varpi gL\left(1+\frac{\partial^2}{\partial\phi^2}\right)v_\varpi'-\right.\right.$$
$$-\varpi^2 g\frac{\partial v_z'}{\partial z}+\varpi u_\phi\frac{\partial v_z'}{\partial z}\Big)+$$
$$+\hat{v}_z'\left(\frac{1}{\varpi}\frac{\partial Q'}{\partial z}+\varpi gL\frac{\partial^2 v_z'}{\partial\phi^2}+\varpi^2 g\frac{\partial v_\varpi'}{\partial z}-\right.$$
$$\left.\left.-\varpi u_\phi\frac{\partial v_\varpi'}{\partial z}-\varpi v_\varpi'\frac{\partial u_\phi}{\partial z}\right)\right]^A. \tag{A.5}$$

By using eqs. (35) and (36), expression (A.5) becomes

$$T=\left[\hat{v}_\varpi'\frac{\varpi}{R^{\frac{1}{2}}}(\boldsymbol{\nabla}\times\boldsymbol{F}')_z+\varpi u_\phi\hat{v}_\varpi'\frac{\partial v_z'}{\partial z}-\right.$$
$$\left.-\hat{v}_z'\frac{\varpi}{R^{\frac{1}{2}}}(\boldsymbol{\nabla}\times\boldsymbol{F}')_\varpi-\varpi\hat{v}_z'\frac{\partial}{\partial z}(u_\phi v_\varpi')\right]^A. \tag{A.6}$$

Employing now the second of (A.4), we find that eq. (A.6) may be transformed to give

$$T=\left[\frac{\varpi}{R^{\frac{1}{2}}}\{\boldsymbol{v}'\times(\boldsymbol{\nabla}\times\hat{\boldsymbol{F}}')\}_\phi+\frac{\partial}{\partial z}\{\varpi u_\phi\hat{v}_\varpi' v_z'\}\right]^A$$
$$=\Lambda+\frac{\partial}{\partial z}(u_\phi w), \tag{A.7}$$

where we have used definitions (16) and (46) for w and Λ. On substituting expression (A.7) for the last term in eq. (44), we obtain eq. (45), as required.

Acknowledgements

We would like to thank Drs. S. Childress and R. Hide for comments made to one of us (P.H.R.) when he attended the 1967 Summer Programme in Geophysical Fluid Dynamics at Woods Hole Oceanographic Institution. This research was supported in part by Grant GR/3/425 from the Natural Environment Research Council.

References

COWLING, T. G. (1933), Monthly Notices Roy. Astron. Soc. **94**, 39.

BACKUS, G. (1957), Astrophys. J. **125**, 500.

BRAGINSKIĬ, S. I. (1964,a), J.E.T.P. (U.S.S.R.), **47**, 1084. Trans. in Soviet Phys. JETP, **20**, (1965) 726.

BRAGINSKIĬ, S. I. (1964,b), Geomagnetizm i Aéronomiya (U.S.S.R.) Vol. IV, **4**, 732. Trans. in Geomagnetism and Aeronomy Vol. IV, **4**, 572.

ROBERTS, P. H. (1967) *An Introduction to Magnetohydrodynamics* (Longmans, Green & Co.).

TOUGH, J. G. (1967) Geophys. J. **13**, 393. (Corrigendum to appear.)

Physics of the Earth and Planetary Interiors, 8 (1974) 389–390

RESEARCH NOTE

MAGNETIC CORE–MANTLE COUPLING BY INERTIAL WAVES

P.H. ROBERTS

School of Mathematics, University of Newcastle upon Tyne, Newcastle upon Tyne (Great Britain)

Accepted for publication January 17, 1974

It is generally believed that the geomagnetic field in the core is predominantly toroidal and axisymmetric, and that many of the features of the secular variation can be attributed to long wave-length asymmetric waves that propagate on these field lines (Braginsky, 1964a; 1967). Indeed, it has been suggested (Braginsky, 1964b) that these asymmetries, which break the force of Cowling's theorem, arise from instabilities that provide the driving motor of the dynamo mechanism. Because of the dominance of Coriolis forces, expressed by the large size of the parameter $2\Omega a/V$ (where Ω is the angular velocity of the earth, a is the core radius, and V is the Alfvén velocity based on the toroidal field strength), the waves can be divided into two main types: inertial waves, little influenced by magnetic or Archimedean (buoyancy) forces, of typical frequency $\omega_F = 2\Omega$, and MAC waves, in which Magnetic, Archimedean, and Coriolis forces are of comparable magnitude, which have a typical frequency of $\omega_S = V^2/2\Omega a^2$ (Braginsky, 1964b, 1967). Since $\omega_F \gg \omega_S$, Hide (1966) has called these waves the fast and the slow, respectively. He has suggested that the westward drift of the secular variation pattern is a manifestation of the slow-wave propagation, an opinion strengthened by the analysis of Acheson (1972). He has pointed out that the conductivity of the mantle is sufficient to screen the fields associated with the fast waves from observation at the earth's surface. He has suggested that, nevertheless, these fields may affect the electromagnetic coupling of core and mantle in a significant way. It is the object of this note to exhibit an inherent difficulty in this novel idea. The argument is an elaboration of a point made by Rochester (1970, p. 143 et seq.); see also MacDonald and Ness (1961, p. 1875 et seq.), and Toomre (1966, p. 40 et seq.).

Let $\boldsymbol{B}_F$ be the rapidly varying part of $\boldsymbol{B}$, the geomagnetic field, and let $\boldsymbol{B}_S = \boldsymbol{B} - \boldsymbol{B}_F$ be the remainder, which contains all steady and slowly varying parts. Let the current density, $\boldsymbol{j}$ be similarly divided into $\boldsymbol{j}_F$ and $\boldsymbol{j}_S$. Of the four contributions, $\boldsymbol{j}_F \times \boldsymbol{B}_F$, $\boldsymbol{j}_F \times \boldsymbol{B}_S$, $\boldsymbol{j}_S \times \boldsymbol{B}_F$ and $\boldsymbol{j}_S \times \boldsymbol{B}_S$ to the Lorentz force $\boldsymbol{j} \times \boldsymbol{B}$, the first three owe their existence to the fast variations, but the second and third will average to zero over the fast time scale. We will, then, only be concerned with the couple:

$$\Gamma_F = \int \boldsymbol{r} \times (\boldsymbol{j}_F \times \boldsymbol{B}_F) \mathrm{d}V \qquad (1)$$

exerted by the fast fields on the mantle. Here $\boldsymbol{r}$ is the radius vector from the geocentre, and the integration is taken over the mantle. Because of the skin effect, $\boldsymbol{B}_F$ and $\boldsymbol{j}_F$ are significant only in the lowermost mantle, in a layer of thickness $\delta = (2/\mu\sigma\omega_F)^{1/2}$, where μ is the permeability and σ is the mantle conductivity which, if we adopt the currently popular estimate of 300 mho/m, gives a δ of only about 10 km. Taking $j_F \simeq B_F/\mu\delta$, we see from eq. 1 that Γ_F is apparently of order $a(B_F{}^2/\mu\delta)4\pi a^2\delta = 4\pi B_F{}^2 a^3/\mu$. It is not easy to estimate the magnitude of $\boldsymbol{B}_F$, because of the screening to which it is subject. Recently Acheson (in preparation) has solved a problem of the dynamical excitation of slow and fast waves, and has shown that the amplitude of the slow component

provides an extreme upper limit for that of the fast. He concludes that B_F is certainly less than 1 gauss and this yields $4\pi B_F{}^2 a^3/\mu$ of only about $4 \cdot 10^{18}$ kg m^2/sec^2. Now the observations of the variations in the length of the day suggest that the unbalanced couples required between core and mantle are of order 10^{19}–10^{20} kg m^2/sec^2 (e.g., Roberts, 1972). Given the uncertainties of the estimate of B_F, it at first sight appears that Γ_F may, indeed, as Hide suggests, contribute significantly to the couples on core and mantle.

Consider, however, the direction of the Lorentz force associated with the fast wave. Since $\delta \ll a$, skin effect theory shows that B_{Fr} in the mantle is small compared with the horizontal component, B_{FH}, of B_F; in fact, $B_{Fr} = 0(\delta/a)B_{FH}$ and similarly $j_{Fr} = 0(\delta/a)j_{FH}$. Thus the dominant component, $j_{FH} \times B_{FH}$, of $j_F \times B_F$ is radial, and does not contribute to the integral of eq. 1. The estimate for Γ_F must be decreased by the factor δ/a to give:

$$\Gamma_F \simeq 4\pi B_F{}^2 a^2 \delta/\mu \qquad (2)$$

which is now only of order 10^{16} kg m^2/sec^2, which may fail by 3 or 4 orders of magnitude.

Acknowledgements

I am grateful to Dr. D. Acheson for communicating his unpublished results quoted above. I wish to thank him, and Dr. Hide, for his comments on this paper.

References

Acheson, D., 1972. J. Fluid Mech., 52: 529.
Acheson, D., in preparation. On fast hydromagnetic waves within the earth, and core–mantle coupling.
Braginsky, S.I., 1964a. Geomagn. Aeron., 4: 698.
Braginsky, S.I., 1964b. Sov. Phys. JETP, 20: 1462.
Braginsky, S.I., 1967. Geomagn. Aeron., 6: 851.
Hide, R., 1966. Philos. Trans. R. Soc. London, Ser. A, 259: 615.
MacDonald, G.J.F. and Ness, N.F., 1961. J. Geophys. Res., 66: 1865.
Roberts, P.H., 1972. J. Geomagn. Geoelectr., 24: 231.
Rochester, M.G., 1970. In: L. Mansinha, D.E. Smylie and A.E. Beck (Editors), Earthquake Displacement Fields and the Rotation of the Earth. Reidel, Dordrecht, p. 136.
Toomre, A., 1966. In: B.G. Marsden and A.G.W. Cameron (Editors), The Earth–Moon System. Plenum Press, New York, p. 33.

Proc. R. Soc. Lond. A. **344**, 235–258 (1975)
Printed in Great Britain

A three-dimensional kinematic dynamo

BY S. KUMAR† AND P. H. ROBERTS
School of Mathematics, University of Newcastle upon Tyne

(*Communicated by R. Hide, F.R.S. – Received* 3 *December* 1974)

A number of steady (marginal) solutions of the induction equation governing the magnetic field created by a particular class of three-dimensional flows in a sphere of conducting fluid surrounded by an insulator are derived numerically. These motions possess a high degree of symmetry which can be varied to confirm numerically that the corresponding asymptotic limit of Braginsky is attained. The effect of altering the spatial scale of the motions without varying their vigour can also be examined, and it is found that dynamo action is at first eased by decreasing their characteristic size. There are, however, suggestions that the regenerative efficiency does not persistently increase to very small length scales, but ultimately decreases. It is further shown that time-varying motions, in which the asymmetric components of flow travel as a wave round lines of latitude, can sustain fields having co-rotating asymmetric parts. It is demonstrated that, depending on their common angular velocity, these may exist at slightly smaller magnetic Reynolds numbers than the corresponding models having steady flows and fields.

The possible bearing of the integrations on the production of the magnetic field of the Earth is considered, and the implied ohmic dissipation of heat in the core of the Earth is estimated for different values of the parameters defining the model.

1. INTRODUCTION

It is widely believed that the magnetism of cosmic bodies such as the Earth and Sun arises from a self-excited dynamo process, that is the internal fluid motions in the core of the Earth and the solar convection zone cross the ambient magnetic fields, $\boldsymbol{B}$, to create the electromotive forces necessary to sustain the currents producing those fields. The simplest dynamo models are kinematic, that is the fluid motions, $\boldsymbol{u}$, in the volume V of interest are chosen in some qualitatively plausible way and no effort is made to provide a dynamical justification for them. The mathematical problem is then linear and reduces to solving the induction equation which, in dimensionless units, is

$$\partial \boldsymbol{B}/\partial t = R \operatorname{curl}(\boldsymbol{u} \times \boldsymbol{B}) + \nabla^2 \boldsymbol{B}, \tag{1}$$

where R, the magnetic Reynolds number, is a dimensionless measure of the fluid velocity. At the boundary, S, of V it is required that $\boldsymbol{B}$ shall be continuous with the potential field residing in the surrounding insulator $\hat{\mathrm{V}}$. The vital condition of self-excitation demands that this potential field be free of singularities and approaches

† Now at Computer Centre, Hatfield Polytechnic, Hatfield, Herts.

zero at least as rapidly as r^{-3} with increasing distance, r, from some origin, O, in V. We also apply

$$\operatorname{div} \boldsymbol{u} = 0, \quad \operatorname{div} \boldsymbol{B} = 0, \tag{2,3}$$

the latter of which is demanded by the electrodynamics, and the former by simplicity; to assume otherwise would be to add complications devoid of essential electromagnetic content.

The current state of dynamo theory has been recently reviewed by Gubbins (1974) and by Soward & Roberts (1975). We summarize here only a few of the relevant facts. Axisymmetric fields, $\boldsymbol{B}$, cannot be maintained by any motion $\boldsymbol{u}$ (Cowling 1933). Toroidal† motions, $\boldsymbol{u}$, cannot sustain any field $\boldsymbol{B}$ (Elsasser 1946). Regeneration is impossible if the magnetic Reynolds number, R, is 'too small' (Backus 1958; Childress 1969). The combined effects of these theorems is to exclude from consideration nearly all simple models, and not surprisingly progress has been slow. Several of the clearest examples of dynamo action have been transparent only because cosmical realism has been sacrificed to mathematical expediency.

In the search for simplifications that make the theory tractable without undue remoteness from cosmical applications, there have been two major successes both stemming from a fundamental paper by Parker (1955). Stated in modern terminology (Moffatt 1970), Parker showed that dynamo action is to be expected when the motions possess sufficient helicity, i.e. when their velocities and vorticities are sufficiently correlated. The turbulent dynamo theory initiated by Steenbeck, Krause & Rädler (1966)‡ divided $\boldsymbol{u}$ and $\boldsymbol{B}$ into ensemble means, $\bar{\boldsymbol{u}}$ and $\overline{\boldsymbol{B}}$, and fluctuating turbulent remainders, $\boldsymbol{u}'$ and $\boldsymbol{B}'$. By solving for $\boldsymbol{B}'$ and averaging $\boldsymbol{u}' \times \boldsymbol{B}'$ they could show, in the simplest case of pseudo-isotropic turbulence (in which $\boldsymbol{u}'$ has isotropic statistical properties with non-zero helicity $h = \overline{\boldsymbol{u}' . \operatorname{curl} \boldsymbol{u}'}$ and zero $\bar{\boldsymbol{u}}$), that $\overline{\boldsymbol{B}}$ obeyed

$$\partial \overline{\boldsymbol{B}}/\partial t = \operatorname{curl} (\alpha \overline{\boldsymbol{B}} + \bar{\boldsymbol{u}} \times \overline{\boldsymbol{B}}) + \nabla^2 \overline{\boldsymbol{B}}, \tag{4}$$

where α is linearly related to h. Comparing (4) with (1) it is seen that the induction by microscale fields and motion adds a novel electromotive force to the induction of mean fields. Steenbeck, Krause & Rädler developed this idea in subsequent papers to form a new subject which they christened 'mean field electrodynamics'. From the symbol they happened to choose to represent it, they called the new electromotive force, $\alpha \overline{\boldsymbol{B}}$, the '$\alpha$-effect'. It is relevant to our later considerations to remark that, although the theory of mean-field electrodynamics is most transparent when it is assumed that the scale of the microscale motions is small compared with macroscale lengths such as the dimensions of V, this does not necessarily mean that dynamo action is most efficient when the scale of motion is small.

Turbulent dynamo theory is often felt to be of particular relevance to the Sun, since its photosphere is observed to be in highly disorganized motion. The other strikingly successful theory referred to above is probably of greater relevance to the

† A motion is said to be toroidal when V is spherical and the flow has no component in the direction of the radius vector, $\boldsymbol{r}$, from the centre O of V; see the first term of (5) below.

‡ This paper, and many of its successors, have appeared in English translation (Roberts & Stix 1971).

geodynamo, which it was, in fact, designed to model. This theory defines and studies the nearly-symmetric dynamo (Braginsky 1964*a*). The flow is supposed to be dominantly zonal, that is parallel to circles of latitude in the core. This axisymmetric toroidal motion, $\bar{\boldsymbol{u}}_0$, cannot regenerate field, but it is imagined that an asymmetric motion, $\boldsymbol{u}'$, of order $R^{-\frac{1}{2}}\bar{u}_0$ is present which is at least in part poloidal;† for completeness, an axisymmetric motion, $\bar{\boldsymbol{u}}_p$, in meridian planes is also added of order $R^{-\frac{1}{2}}\bar{u}_0$. When the field $\boldsymbol{B}$ is similarly divided, it is found that the axisymmetric component $\bar{\boldsymbol{B}} = \bar{\boldsymbol{B}}_0 + \bar{\boldsymbol{B}}_p$ obeys (4), with $\bar{\boldsymbol{u}} = \bar{\boldsymbol{u}}_0 + \bar{\boldsymbol{u}}_p$. In order for these simple results to hold, it is necessary to give $\boldsymbol{u}$ and $\boldsymbol{B}$ a Lagrangian interpretation (Soward 1972). As a result, the Eulerian $\bar{\boldsymbol{u}}_p$ and $\bar{\boldsymbol{B}}_p$ must be given new 'effective' forms, $\bar{\boldsymbol{u}}_{pe}$ and $\bar{\boldsymbol{B}}_{pe}$, before (4) is obeyed. Braginsky gave the prescription for these vectors and also for α in terms of $\boldsymbol{u}'$. Soward (1972) related this α directly to the helicity of the flow.

Several authors have sought to solve the kinematic dynamo problem directly by numerical means. Ideally this approach would provide a 'black box' that would automatically disgorge a magnetic field when fed by a motion. Such a box would allow dynamical questions to be examined without too great a distraction from the induction problem. Unfortunately, this ideal situation is still some distance from realization. Convincing results have been obtained by G. O. Roberts (see P. H. Roberts, 1971) and Gubbins (1973), who took advantage of the fact that, while axisymmetric fields are ruled out by Cowling's theorem, axisymmetric motions are not. They integrated (1)–(3) directly for such flows and obtained convincingly converged working dynamos, of completely asymetric magnetism.

Further progress has recently been made by Pekeris, Accad & Shkoller (1973), who obtained a dynamo by using a particularly simple, highly helical, but three-dimensional motion. The geophysical motivation for this flow (Pekeris 1972) while of some independent interest seems somewhat remote from many current views of the geodynamo (Braginsky 1964*b*; Bullard 1971; Bullard & Gellman 1954; Parker 1955).

The present paper takes a step nearer to the full black box. There is only one novelty (see § 2) added to the work described above, the success of the results depending largely on generous allocations of time on large computers, the IBM 360/67 at Newcastle University and the IBM 370/165 at Cambridge University. Even with these machines, however, it is difficult to obtain convincingly converged solutions unless the motion $\boldsymbol{u}$ is well chosen, for which we had no better criterion than trial and error.

After explaining (§ 2) the methods used and the motion chosen, we try to convince the reader (§ 3) that our results have reliably converged. We then use our model to study a number of questions of interest. First (§ 4), by varying a parameter that essentially describes the degree of asymmetry of our model, we ask whether it passes continuously into the corresponding Braginsky limit. Next (§ 5), by varying a parameter that essentially describes the degree of spatial complexity of our model,

† A poloidal motion has radial components but has toroidal vorticity; see the second term of (5) below.

we ask whether the regenerative efficiency of the dynamo is improved as we move towards the small-scale motions of Steenbeck, Krause & Rädler. Next (§ 6), we study the inductive effect of time-varying motions in the form of a travelling wave round the fluid sphere in the particular case in which the magnetic field co-rotates with it: the significance of such structures has already been noted by Soward (1974) in his magnetohydrodynamic dynamo model. Finally (§ 7), we set out some integrated properties of our solutions which may be of possible usefulness to geomagnetic theorists.

2. Method and motion

One of the difficulties of satisfying (1)–(3) arises from the matching of their solutions to the potential field in $\hat{\mathrm{V}}$. This essentially poses a non-local boundary condition on S, i.e. one that relates the field at any point of S to that at every other point of S. One way out of this difficulty in the case in which V is the sphere $r = 1$ is to adopt the procedure of Bullard & Gellman (1954). The motion and field are divided into toroidal and poloidal parts (Elsasser 1946)

$$\boldsymbol{u} = \operatorname{curl} t\hat{\boldsymbol{r}} + \operatorname{curl}^2 s\hat{\boldsymbol{r}}, \quad \boldsymbol{B} = \operatorname{curl} T\hat{\boldsymbol{r}} + \operatorname{curl}^2 S\hat{\boldsymbol{r}}, \tag{5, 6}$$

where $\hat{\boldsymbol{r}}$ is the unit vector from the centre O of V, and t, T, s and S are the defining toroidal and poloidal scalar fields;† clearly, (2) and (3) are now obeyed. Next, the defining scalars are divided into their harmonic components:

$$t = \sum_\alpha t_\alpha(r,t) Y_\alpha(\theta,\phi), \quad T = \sum_\beta T_\beta(r,t) Y_\beta(\theta,\phi), \tag{7}$$

etc., where Y_α denotes a surface harmonic, i.e. a linear combination for one n_α of the $2n_\alpha + 1$ combinations, $P_{n_\alpha}^{m_\alpha}(\theta)\sin m_\alpha\phi$ and $P_{n_\alpha}^{m_\alpha}(\theta)\cos m_\alpha\phi$, for all m_α from 0 to n_α; here (r,θ,ϕ) are spherical polar coordinates and Schmidt normalized associated Legendre functions were employed. Regularity of $\boldsymbol{u}$ and $\boldsymbol{B}$ at $r = 0$ requires

$$s_\alpha = O(r^{n_\alpha+1}), \quad T_\beta = O(r^{n_\beta+1}), \quad \text{etc.}, \tag{8}$$

and the vital condition of self-excitation is

$$T_\beta = 0, \quad \partial S_\beta/\partial r + n_\beta S_\beta = 0 \quad \text{on} \quad r = 1, \tag{9, 10}$$

essentially local conditions that evade the difficulty mentioned at the outset of this section.

On substituting (5)–(7) into (1) and seeking time independent solutions ($\partial\boldsymbol{B}/\partial t = 0$), Bullard & Gellman obtained

$$r^2 \mathrm{d}^2 S_\gamma/\mathrm{d}r^2 - n_\gamma(n_\gamma+1) S_\gamma = R \sum_{\alpha,\beta} [(s_\alpha S_\beta S_\gamma) + (t_\alpha S_\beta S_\gamma) + (s_\alpha T_\beta S_\gamma)], \tag{11}$$

$$r^2 \mathrm{d}^2 T_\gamma/\mathrm{d}r^2 - n_\gamma(n_\gamma+1) T_\gamma = R \sum_{\alpha,\beta} [(s_\alpha S_\beta T_\gamma) + (t_\alpha S_\beta T_\gamma) + (s_\alpha T_\beta T_\gamma) + (t_\alpha T_\beta T_\gamma)], \tag{12}$$

where the coupling coefficients have complicated forms which they give in full. The condition of time independence is intended to locate the marginal value R_c of R

† The use of lower case and capital letters to distinguish velocity fields from magnetic fields suffers from the objection that t is also used for time. It should, however, be clear from the context how t should be interpreted.

dividing the regime $R < R_c$ in which all self-excited solutions of (1)–(3) ultimately evanesce from the regime $R > R_c$ in which dynamo action occurs. Unfortunately, partly because of the state of computer technology at the time Bullard and Gellman wrote and partly because the motions they assumed were (as it appears now) insufficiently helical, they were unable to use the apparatus described above in a convincing way (Gibson & Roberts 1969; Lilley 1970; Gubbins 1973). Also, it is now known that circumstances exist in which the field oscillates at criticality (e.g. Soward 1974). Such a.c. dynamos are excluded from the formulation (11) and (12), which can only deal with d.c. regeneration.

To solve the eigenvalue problem (9)–(12) for R, we may further expand t_α, T_β, etc., into, for example, sums of spherical Bessel functions (Elsasser 1946; Backus 1958; Kropachev 1971), or we may use collocation methods (Gibson & Roberts, 1969). We shall, however, follow Bullard & Gellman in dividing the range $0 \leqslant r \leqslant 1$ into M equal intervals, and use (10)–(12) to relate defining scalars at neighbouring grid points. In an effort to eliminate coding errors we constructed two partially independent programmes, in one of which R was found by a matrix method and in the other of which parallel shooting was employed. In each case, the series (7) for T and S were truncated at $n_\alpha = N$, implying for the motion (17) below that $K = \frac{1}{2}N(N+2)$ harmonics of the form $P_n^m(\theta)\overset{\cos}{\sin} m\phi$ had to be considered for even N, and

$$K = \tfrac{1}{2}(N^2+2N-1)$$

for odd N (K = 17, 24, 31, 40, 49 and 60 for N = 5, 6, 7, 8, 9 and 10 respectively).

In the matrix method, the finite difference equivalents of (9)–(12) are used to construct a sparse determinant of $(KM)^2$ elements. This consists of M^2 blocks of K^2 elements, in each of which the interaction of the K harmonics for that particular grid-point are represented. The eigenvalues in R of this determinant were located by inverse iteration (Isaacson & Keller 1966).

In the parallel shooting method (Keller 1968; Davey 1973) first order equations replace (11) and (12): they are

$$\mathrm{d}T_\gamma/\mathrm{d}r = r^{n_\gamma}\mathscr{T}_\gamma, \quad \mathrm{d}S_\gamma/\mathrm{d}r = r^{n_\gamma}\mathscr{S}_\gamma, \tag{13, 14}$$

$$\mathrm{d}\mathscr{T}_\gamma/\mathrm{d}r = -n_\gamma\mathscr{T}_\gamma/r + n_\gamma(n_\gamma+1)\,T_\gamma/r^{n_\gamma+2} + (R/r^{n_\gamma+2}) \sum_{\alpha,\beta} [(s_\alpha S_\beta T_\gamma) + (t_\alpha S_\beta T_\gamma) + (s_\alpha T_\beta T_\gamma) + (t_\alpha T_\beta T_\gamma)], \tag{15}$$

$$\mathrm{d}\mathscr{S}_\gamma/\mathrm{d}r = -n_\gamma\mathscr{S}_\gamma/r + n_\gamma(n_\gamma+1)\,S_\gamma/r^{n_\gamma+2} + (R/r^{n_\gamma+2}) \sum_{\alpha,\beta} [(s_\alpha S_\beta S_\gamma) + (t_\alpha S_\beta S_\gamma) + (s_\alpha T_\beta S_\gamma)], \tag{16}$$

for the $2K$ independent variables T_γ, $\mathscr{T}_\gamma$, S_γ and $\mathscr{S}_\gamma$. The choice (13) and (14) for $\mathscr{T}_\gamma$ and $\mathscr{S}_\gamma$ is dictated by (8). Then K independent solutions of (13)–(16) satisfying (8) are simultaneously integrated from $r = 0$ to $r = 1$ by Runge–Kutta methods. Gram-Schmidt orthonormalization is undertaken every few steps at an interval decided by experience. In this way the linear independence of the solutions is preserved. A $2K \times K$ matrix, M_{ij}, is formed, where M_{ij} is the value of the ith variable

of the set $(T_\gamma, S_\gamma, \mathscr{T}_\gamma, \mathscr{S}_\gamma)$ for the jth independent solution. From this matrix a second $K \times K$ matrix, m_{ij}, is formed where m_{ij} is the value of the ith variable of the set $(T_\gamma, \mathscr{S}_\gamma, +n_\gamma, S_\gamma)$ for the jth solution. The condition that a linear combination of the K independent solutions should exist which meets (9) and (10) is then the condition that the determinant of m_{ij} is zero. Once R has been obtained in this way, the corresponding eigenfunction can be derived using the method of Conte (1966).

Although parallel shooting was found to be less efficient than inverse iteration, it has the advantage of using less computer storage, since grid values that have outlived their usefulness can be discarded as the integration proceeds. The $(s_\alpha S_\beta T_\gamma)$ interaction in (15) requires a knowledge of $\mathrm{d}\mathscr{S}_\beta/\mathrm{d}r$, and for this reason each forward step of the solution is taken first for S_γ and $\mathscr{S}_\gamma$ by means of (14) and (16), and then for T_γ and $\mathscr{T}_\gamma$ by means of (13) and (15).

Our choice of motion was governed by several considerations, computational and geophysical. First, from the numerical point of view, it was desirable to minimize the band-width for the matrix method. Thus no harmonics of motion were included for $n_\alpha > 2$. Secondly, we desired flows that were comparatively easy to compare with the Braginsky limit (§4). We focused our attention on the following motion:

$$\boldsymbol{u} = t_1 + \epsilon_1 s_2 + \epsilon_2 s_2^{2s} + \epsilon_3 s_2^{2c}, \tag{17}$$

where

$$t_1 = r^2(1-r^2) + \lambda r^2, \tag{18}$$

$$s_2 = r^6(1-r^2)^3, \tag{19}$$

$$s_2^{2s} = r^4(1-r^2)^2 \cos pr, \tag{20}$$

$$s_2^{2c} = r^4(1-r^2)^2 \sin pr, \tag{21}$$

and λ, ϵ_i and p are constants. The axisymmetric part of the motion is illustrated in figure 1.

A few comments about (17)–(21) are in order. First, the flow admits fields $\boldsymbol{B}$ of different parity with respect to the equatorial plane $z = 0$. For quadrupole type solutions B_x and B_y are symmetric and B_z is antisymmetric in z; for the dipole type solutions that we will consider the reverse is true and the sums (7) involve odd n_β for S and even n_β for T. Second, when $\lambda = 0$ the motion obeys the non-slip boundary conditions on $r = 1$. Even when $\lambda \neq 0$, there is no slip at the boundary when viewed in a frame rotating with angular velocity λ. Since the induction equation (1) is invariant to such a change of frame, we can by making λ finite examine solutions in which field and motion are time-varying and co-rotate about Oz. Third, while t_α and s_α obey (8), s_2^{2s}/r^3 and s_2/r^3 are not even in r, as one might have expected from considerations of the analyticity of $\boldsymbol{u}$ at O (Bullard 1972, private communication). In fact, the assumption of even $t_\alpha/r^{n_\alpha+1}$ and $s_\alpha/r^{n_\alpha+1}$ has seldom been observed in kinematic dynamo computations, even though it would be a necessary ingredient of a dynamical calculation. Our choice for s_α was made to simplify the analysis of the Braginsky limit (§4). Since s_2^{2s}/r^3 vanishes linearly at O according to (20) the rates of strain upon which the advection of field depend will not be singular at O, and it seemed to us that the non-analyticity arising from s_2^{2s} would not lead to pathologies in the solution, a belief strengthened by the results reported below.

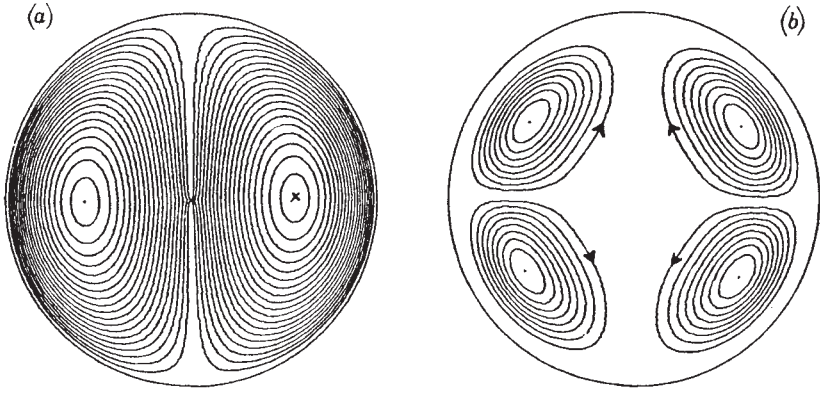


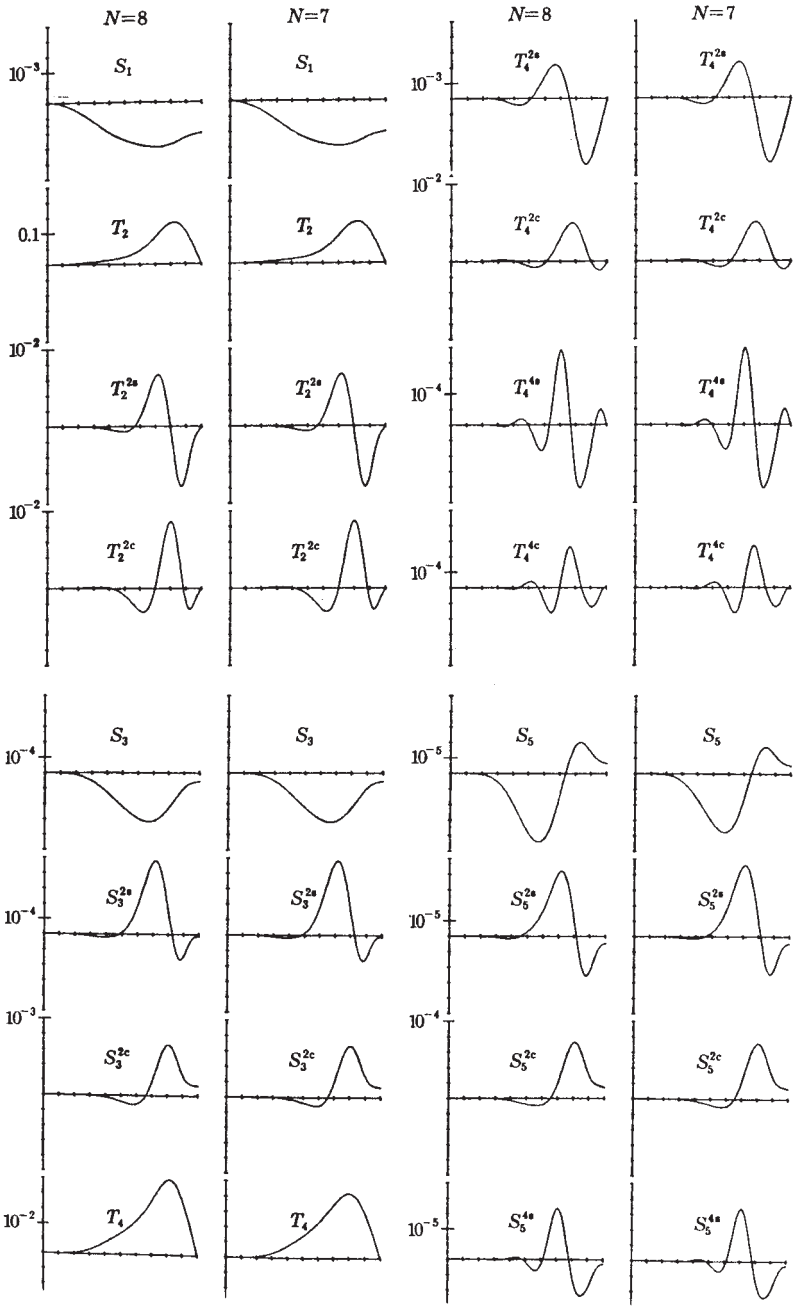
FIGURE 1. (*a*) Contours of constant axisymmetric toroidal flow in a meridian plane. (Contours surrounding a cross signify motion into the paper; those containing a dot, motion out of the paper.) (*b*) Streamlines of axisymmetric poloidal flow in a meridian plane. If $\epsilon_1 > 0$ in (17) the sense of circulation is given by the arrows.

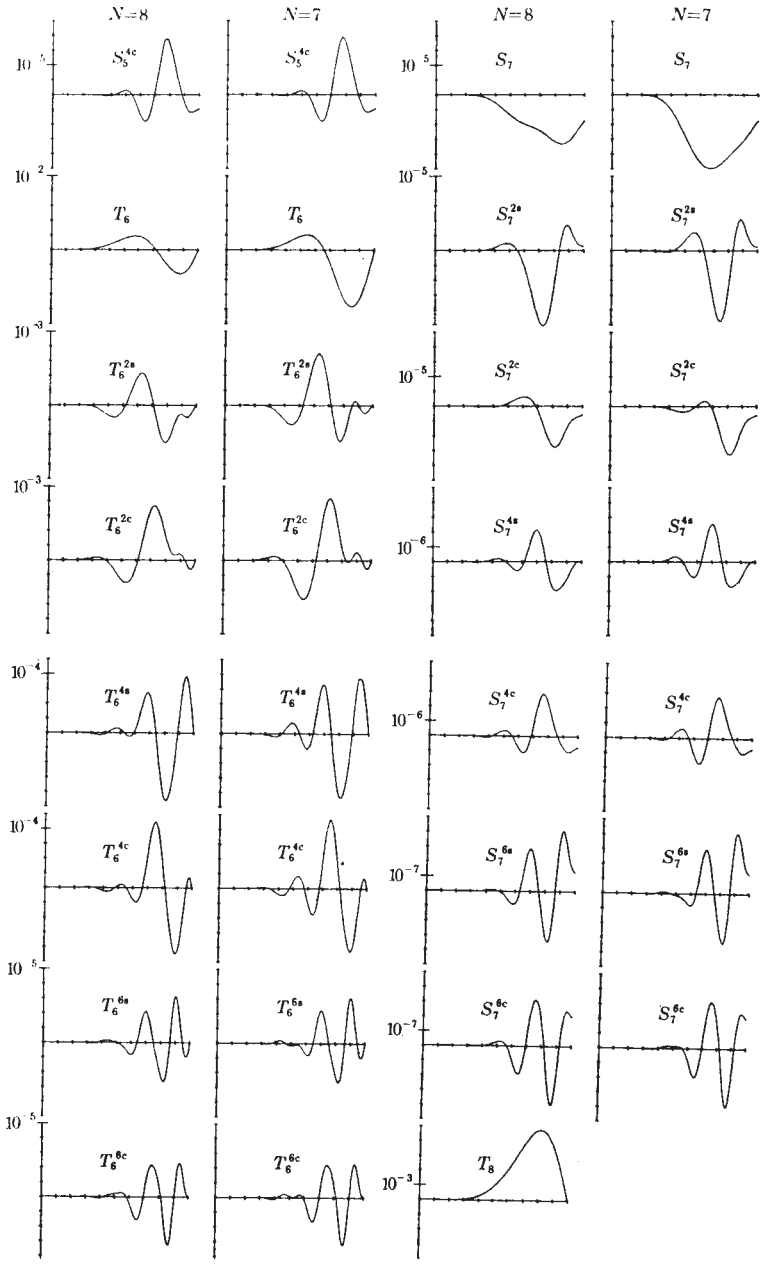
3. CONVERGENCE

Even when, as for instance for purely toroidal flows, it is known that regenerative solutions of (1)–(3) do not exist, it may still happen that the finite difference form of the truncated equations (9)–(12) possesses real finite eigenvalues, R. And, even when the convergence of the eigenvalue at low truncation levels can appear convincing, it may nevertheless be found to be illusory at higher truncation levels. It is not surprising that some writers, amongst whom the present authors hope not to be numbered, have been beguiled by a verisimiltude of dynamo action which has not been substantiated by subsequent studies on larger and more capable machines. Our experience has led us to value the advice of Gubbins (1973) that no reliance should be placed on an apparent convergence of the eigenvalue: although such a convergence is necessary, it is not sufficient. The eigenfunctions should be compared at different truncation levels before deciding whether the solution is to be believed. Pressure of space will prevent us from displaying the eigenfunctions in most cases, but we have in fact generally followed Gubbins's counsel.

In addition to the application of two different integration methods (§ 2), we also performed several calculations in single and double precision, and compared the results. During the course of our work, the model by Pekeris *et al.* (1973) was published, and we digressed from our studies (which had been reported briefly in a NATO Conference proceedings†) to test our methods against their results. We were

† Moffatt (1973). The model reported there was somewhat different to that defined by (17)–(21) above, whose superior convergence properties were not appreciated at the time of the meeting.


N=8
N=7
N=8
N=7
S_1
T_2
T_2^{2s}
T_2^{2c}
T_4^{2s}
T_4^{2c}
T_4^{4s}
T_4^{4c}
S_3
S_3^{2s}
S_3^{2c}
T_4
S_5
S_5^{2s}
S_5^{2c}
S_5^{4s}


N=8
N=7
N=8
N=7
S_5^{4c}
S_5^{4c}
S_7
S_7
T_6
T_6
S_7^{2s}
S_7^{2s}
T_6^{2s}
T_6^{2s}
S_7^{2c}
S_7^{2c}
T_6^{2c}
T_6^{2c}
S_7^{4s}
S_7^{4s}
T_6^{4s}
T_6^{4s}
S_7^{4c}
S_7^{4c}
T_6^{4c}
T_6^{4c}
S_7^{6s}
S_7^{6s}
T_6^{6s}
T_6^{6s}
S_7^{6c}
S_7^{6c}
T_6^{6c}
T_6^{6c}
T_8

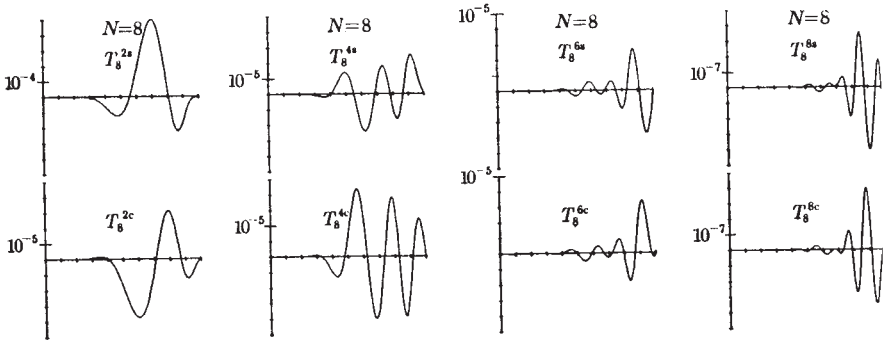


FIGURE 2. A comparison of the eigenfunctions of the induction equation at $N = 7$ and $N = 8$ levels of truncation. Both are normalized by $S_1(1) = 10^{-3}$. (The modes T_8 to T_8^{8c} vanish, by definition, at the $N = 7$ truncation level.)

TABLE 1. CONVERGENCE OF EIGENVALUE IN THE CASE
$\lambda = 0$, $p = 3\pi$, $\epsilon_1 = 0.03$, $\epsilon_2 = \epsilon_3 = 0.04$

	M				parallel shooting (double precision)
N	60	80	100	120	
5	3839	3842	3847 3859	3851	3868
6	3837	3861	3851 3871†	3856	3889
7	3865	3891 3891†	3875	3883 3889†	3918
8	3844	3841	3852	3856	3908
9	3846	3840	—	—	—
10	3843	—	—	—	—

† Double precision result from inverse iteration. (Absence of a dagger signifies single precision.)

ultimately† able to confirm in the cases we tested both their eigenvalues and eigenfunctions. (We used the shooting method and obtained the same eigenvalues as they did to three significant figures.)

Table 1 shows the convergence of the eigenvalues of our model for the case $p = 3\pi$, $\lambda = 0$, $\epsilon_1 = 0.03$ and $\epsilon_2 = \epsilon_3 = 0.04$. In figure 2 the functions $T_\beta(r)$ and $S_\beta(r)$

† The appearance of their model led us on an unexpectedly lengthy detour, since we were at first unable to reproduce their eigenvalues. Recently, however, we detected an error in our coding of their motion (though not in our main programme) which, on correction, gave us faith in both their results and our own.

are compared in the $N = 7$ and $N = 8$ cases with $M = 120$. As anticipated, the amplitude of functions of the same type (e.g. the sequence of axisymmetic toroidal or poloidal fields) decreases with ascending harmonic number n_β. At the same time the percentage differences in the harmonics at the $N = 7$ and $N = 8$ truncations progressively increases until of course at $n_\beta = 8$ the former harmonics vanish by definition while the latter do not. It will be seen that the convergence is least good for the sequence $T_2, T_4, T_6, \ldots$ of axisymmetic toroidal fields, but that there appears to be no serious cause for alarm.

The example summarized in table 1 and figure 2 is one of our more convincingly converged cases, and it was found by trial and error. The model defined by (17) to (21) appears to be sensitive to the choice of ϵ_i in (17).

4. The Braginsky limit

Consider the form (17) without at first specializing by (18)–(21). It follows from results (3.14), (3.19) and (3.21) of Braginsky (1964*a*) that the corresponding axisymmetric limit (valid for

$$\epsilon_1 \to 0, \quad \epsilon_2 \epsilon_3 \to 0, \quad R \to \infty, \tag{22}$$

with

$$\bar{\epsilon}_1 = \epsilon_1 R, \quad \bar{\epsilon}_2 = \epsilon_2 R^{\frac{1}{2}}, \quad \bar{\epsilon}_3 = \epsilon_3 R^{\frac{1}{2}} \tag{23}$$

held constant) is given by

$$\tilde{\boldsymbol{u}}_0 = \boldsymbol{t}_1, \quad \tilde{\boldsymbol{u}}_{p\epsilon} = \epsilon_{1\epsilon} \boldsymbol{s}_2 + \epsilon_{4\epsilon} \boldsymbol{s}_4, \tag{24, 25}$$

$$\alpha = \epsilon_2 \epsilon_3 (A_1 P_1 + A_3 P_3), \tag{26}$$

where

$$\epsilon_{4\epsilon} = \tfrac{6}{70} \epsilon_2 \epsilon_3 t_1 \left(\sigma_2^{2c} \frac{\partial \sigma_2^{2s}}{\partial r} - \sigma_2^{2s} \frac{\partial \sigma_2^{2c}}{\partial r} \right), \tag{27}$$

$$\epsilon_{1\epsilon} = \epsilon_1 - \tfrac{10}{21} \epsilon_{4\epsilon}, \tag{28}$$

$$A_1 = \frac{9}{5} \left[\frac{\partial \sigma_2^{2c}}{\partial r} \frac{\partial^2 \sigma_2^{2s}}{\partial r^2} - \frac{\partial \sigma_2^{2s}}{\partial r} \frac{\partial^2 \sigma_2^{2c}}{\partial r^2} \right) - \left(\frac{t_1'}{t_1} + \frac{5}{r} \right)' \left(\sigma_2^{2c} \frac{\partial \sigma_2^{2s}}{\partial r} - \sigma_2^{2s} \frac{\partial \sigma_2^{2c}}{\partial r} \right) \Bigg], \tag{29}$$

$$A_3 = -\frac{9}{5} \left[\left(\frac{\partial \sigma_2^{2c}}{\partial r} \frac{\partial^2 \sigma_2^{2s}}{\partial r^2} - \frac{\partial \sigma_2^{2s}}{\partial r} \frac{\partial^2 \sigma_2^{2c}}{\partial r^2} \right) - \left(\frac{t_1'}{t_1} - \frac{5}{r} \right)' \left(\sigma_2^{2c} \frac{\partial \sigma_2^{2s}}{\partial r} - \sigma_2^{2s} \frac{\partial \sigma_2^{2c}}{\partial r} \right) \right], \tag{30}$$

and

$$\sigma_2^{2s} = s_2^{2s}/t_1, \quad \sigma_2^{2c} = s_2^{2c}/t_1, \tag{31}$$

the primes in (29) and (30) denoting differentiation with respect to r.

These expressions assume particularly simple forms on taking

$$\sigma_2^{2s} = g(r) \cos pr, \quad \sigma_2^{2c} = g(r) \sin pr. \tag{32}$$

Noting that then

$$\sigma_2^{2c} \frac{\partial \sigma_2^{2s}}{\partial r} - \sigma_2^{2s} \frac{\partial \sigma_2^{2c}}{\partial r} = -pg^2, \tag{33}$$

$$\frac{\partial \sigma_2^{2c}}{\partial r} \frac{\partial^2 \sigma_2^{2s}}{\partial r^2} - \frac{\partial \sigma_2^{2s}}{\partial r} \frac{\partial^2 \sigma_2^{2c}}{\partial r^2} = p(gg'' - 2g'^2 - p^2 g^2), \tag{34}$$

we see that A_1 and A_3 will become singular at a zero of t_1 unless g is chosen to have a zero of the same order. Although the role of such 'resonant surfaces' and 'layers

of concentrated regeneration' has been elucidated by Braginsky (1964*a*), their presence hampers numerical comparison between the three-dimensional integrations of § 3 and two-dimensional computations based on (24)–(26). For this reason, we took

$$t_1 = r^2(1-r^2), \quad g = r^2(1-r^2), \tag{35}$$

so recovering the model (18)–(21) with $\lambda = 0$.

In applying (24)–(26) to (4), we find it convenient to scale $\overline{\boldsymbol{B}}_0$ by R^{-1}. We then obtain in place of (4)

$$\partial\overline{\boldsymbol{B}}/\partial t = \operatorname{curl}(\alpha\overline{\boldsymbol{B}}_0 + \boldsymbol{u}_e \times \overline{\boldsymbol{B}}) + \nabla^2\overline{\boldsymbol{B}}, \tag{36}$$

with $\bar{\epsilon}_i$ replacing ϵ_i in (24)–(28). For given $\bar{\epsilon}_1$, (36) and the relevant boundary conditions pose an eigenvalue problem for $\bar{\epsilon}_2\bar{\epsilon}_3$. Results are given in table 2. They were obtained by using a specially constructed programme for axisymmetric fields (see Roberts 1972).

TABLE 2. THE EIGENVALUE $\bar{\epsilon}_2\bar{\epsilon}_3$ IN THE BRAGINSKY LIMIT

($M = 20$, 10 harmonics.)

$\bar{\epsilon}_1$	60	80	100	120*	140	160	180	200
$\bar{\epsilon}_2\bar{\epsilon}_3$	5.30	4.78	4.56	4.49	4.47	4.49	4.55	4.57
$\bar{\epsilon}_1/\bar{\epsilon}_2\bar{\epsilon}_3$	11.3	16.7	21.9	26.7	31.3	35.6	39.6	43.8

* An additional calculation with $M = 24$ and 12 harmonics gave $\bar{\epsilon}_2\bar{\epsilon}_3 = 4.50$ and $\bar{\epsilon}_1/\bar{\epsilon}_2\bar{\epsilon}_3 = 26.6$.

TABLE 3. SOME THREE-DIMENSIONAL CASES RELEVANT TO THE BRAGINSKY LIMIT

($M = 120$, $\epsilon_1/\epsilon_2\epsilon_3 = 25$.)

	$N = 5$		$N = 6$		$N = 7$		$N = 8$	
$\epsilon_2 = \epsilon_3$	R	$\epsilon_2\epsilon_3 R$	R	$\epsilon_2\epsilon_3 R$	R	$\epsilon_2\epsilon_3 R$	R	$\epsilon_2\epsilon_3 R$
0	—	—	—	—	—	—	(∞	4.51)
10^{-2}	—	—	53051	5.30	46631	4.66	47250	4.72
2×10^{-2}	—	—	13055	5.22	12928	5.17	12615	5.05
3×10^{-2}	5903	5.31	6120	5.51	6251	5.63	6180	5.56
5×10^{-2}	2696	6.74	2734	6.84	2830	7.07	2808	7.02
10^{-1}	1216	12.2	1104	11.0	1251	12.5	1239	12.4

Consider the case $\bar{\epsilon}_1/\bar{\epsilon}_2\bar{\epsilon}_3 = 25$, in which the convergence appears to be good. According to table 2, this implies that $\bar{\epsilon}_1 \approx 113$ and $\bar{\epsilon}_2\bar{\epsilon}_3 \approx 4.51$. We should therefore find that, if in the three-dimensional integrations of § 2 we write $\epsilon_1 = \bar{\epsilon}_1/R$ and $\epsilon_2 = \epsilon_3 = \sqrt{(\bar{\epsilon}_2\bar{\epsilon}_3/R)}$ as (23) dictates, we should obtain the eigenvalue

$$R = \bar{\epsilon}_1/\epsilon_1 = \bar{\epsilon}_2\bar{\epsilon}_3/\epsilon_2\epsilon_3.$$

Results of the three-dimensional integrations are shown in table 3 and figure 3; the Braginsky limiting value is also displayed in figure 3.

Further comparisons of the two- and three-dimensional integrations are given in figures 4 and 5. In figure 4 the toroidal field of the Braginsky limit for the case

$N = 6$, $M = 24$, $\bar{\epsilon}_1 = 120$ and $\bar{\epsilon}_2\bar{\epsilon}_3 = 4.50$, is compared with the axisymmetric toroidal field of the full three-dimensional integration for the case $N = 7$, $M = 120$, $\epsilon_1 = 0.0225$, $\epsilon_2 = \epsilon_3 = 0.03$ and $R = 6251$. The axisymmetric poloidal fields are contrasted (through the ϕ component of their vector potential) in figure 5 for the same cases.†

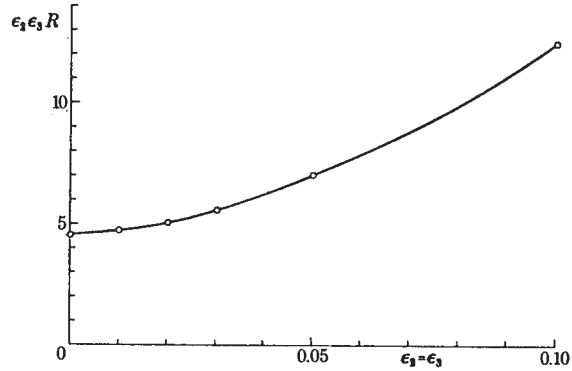


FIGURE 3. $\epsilon_2\epsilon_3 R$ for different value of $\epsilon_2 (= \epsilon_3)$ and $\epsilon_1/\epsilon_2\epsilon_3 = 25$. The corresponding Braginsky value is shown on the $\epsilon_2\epsilon_3 R$ axis.

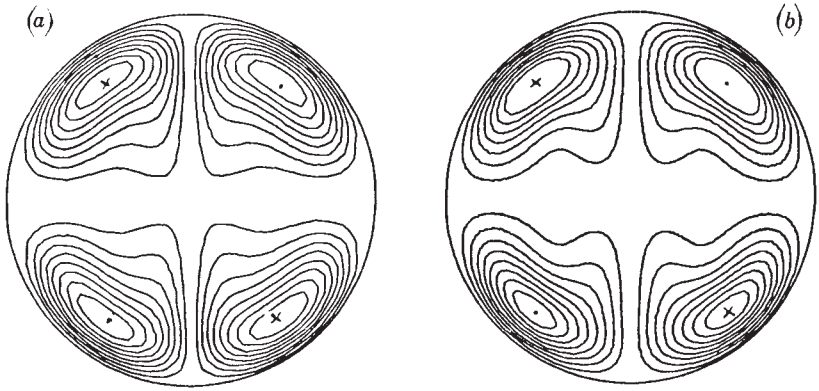


FIGURE 4. (*a*) Contours of equal axisymmetric toroidal field in a meridian plane for the Braginsky limit. ($\epsilon_1/\epsilon_2\epsilon_3 = 25$; contour interval 8×10^{-5}.) (*b*) Contours of equal axisymmetric toroidal field in a meridian plane for the case $\epsilon_1 = 0.0225$, $\epsilon_2 = \epsilon_3 = 0.03$, $p = 3\pi$, $R = 6251$, $N = 7$, $M = 120$. (Contour interval 0.05.) Contours surrounding a cross signify fields into the paper; those containing a dot, fields out of the paper.

† The two-dimensional field of figure 5(*a*) is the effective meridional field, $\mathbf{B}_{pe}$, of Braginsky (1964*a*), and before comparing figure 5(*a*) and (*b*) one or the other should be transformed according to Braginsky's prescription $\mathbf{B}_{pe} = \mathbf{B}_p + \text{curl}\,(w\mathbf{B}_0)$. The differences between $\mathbf{B}_p$ and $\mathbf{B}_{pe}$ are, however, in the present model too small to display graphically.

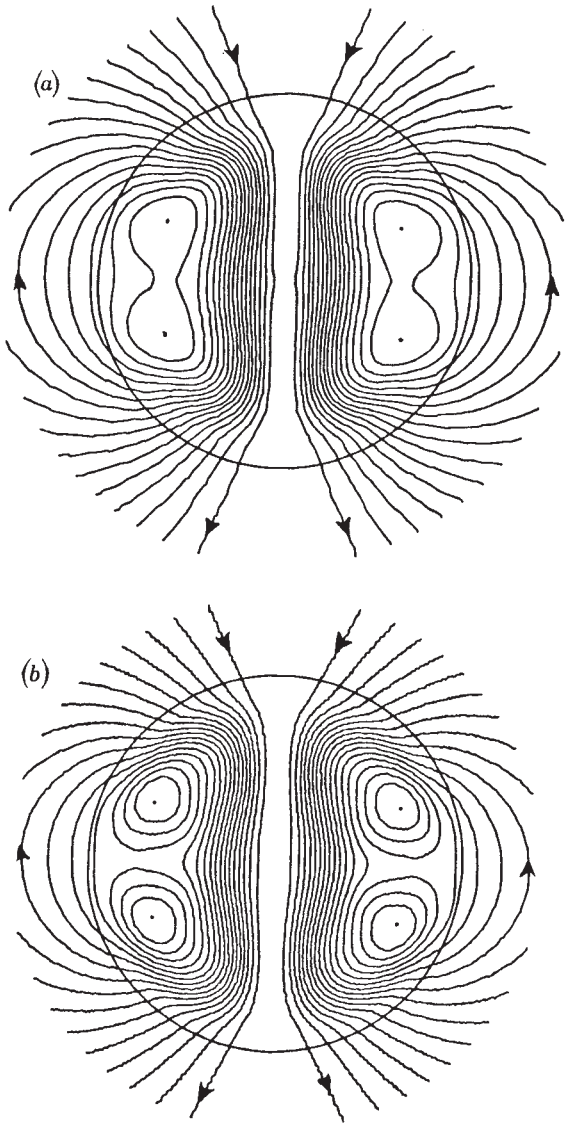


FIGURE 5. (*a*) Contours of equal poloidal axisymmetric vector potential in a meridian plane for the Braginsky limit. ($\epsilon_1/\epsilon_2\epsilon_3 = 25$; Contour interval 6×10^{-5}). (*b*) Contours of equal poloidal vector potential in a meridian plane for the case $\epsilon_1 = 0.0225$, $\epsilon_2 = \epsilon_3 = 0.03$, $p = 3\pi$, $R = 6251$, $N = 7$, $M = 120$. (Contour interval 7×10^{-5}.) The direction of the field is given by the arrows.

Our Braginsky integrations did not lend themselves to comparison of asymmetric fields. In Figure 6 we show (for the three-dimensional case considered in figures 4 and 5 above) contours of equal radial field B_r in projection on the northern hemisphere of the spheres $r = r_V$ and $r = 1$, where r_V is the grid point nearest to the sphere dividing V into two equal volumes. The displacements of two flux centres from the north pole is clear. They are the analogues, for the $m = 2$ asymmetries of (17), of the displacement of geomagnetic from geographic axes that presumably would have arisen had we considered† asymmetric component of flow with $m = 1$.

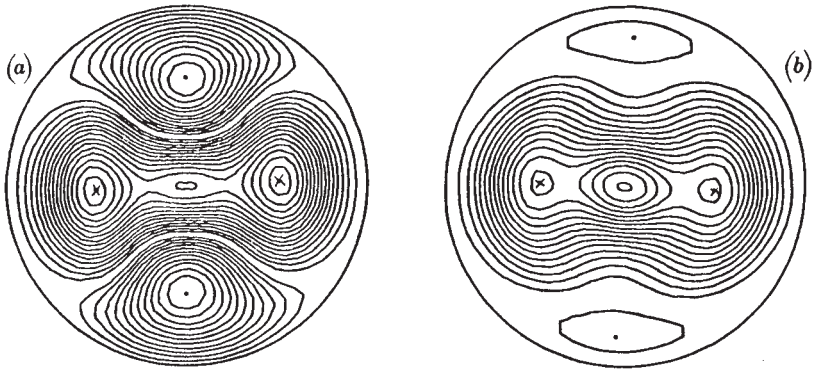


FIGURE 6. (*a*) Contours of equal radial magnetic field on the northern hemisphere of the sphere $r = r_V$ of half volume for the case $\epsilon_1 = 0.0225, \epsilon_2 = \epsilon_3 = 0.03, R = 6251, p = 3\pi, N = 7, M = 120$. (Contours from -0.015 to 0.011 in intervals of 0.001.) (*b*) Contours of equal radial magnetic field on the northern hemisphere of the sphere $r = 1$ for the case $\epsilon_1 = 0.0225$, $\epsilon_2 = \epsilon_3 = 0.03, p = 3\pi, R = 6251, N = 7, M = 120$. (Contours from -0.0032 to 0.0002 in intervals of 0.0002.) Contours surrounding a cross signify fields directed inwards; those containing a dot, fields directed outwards.

According to the Braginsky theory, the critical value of R for given ϵ_1 should be independent of ϵ_2 and ϵ_3 provided their product $\epsilon_2\epsilon_3$ is preserved; see (26) and (27) above. We found, however, that except for very small asymmetries, R is quite sensitive to ϵ_2 and ϵ_3, with regeneration most efficient for $\epsilon_2 = \epsilon_3$, approximately. For this reason, the results we report here are restricted to this case.

It may be noted that, according to (26), the regeneration coefficient α, and the helicity are antisymmetric with respect to the equatorial plane. This stands in contrast to the motion of Pekeris *et al.* (1973) in which the helicity is of the same sign in both hemispheres.

† Following most earlier writers (for example, Bullard & Gellman 1954; Lilley 1970; Pekeris *et al.* 1973) we have restricted attention to motions of $m = 2$ asymmetry rather than the geophysically more realistic case of $m = 1$.

5. Varying scales of motion

The literature of dynamo theory contains several suggestions that regeneration is hampered if the moving parts (e.g. the convection cells) of the dynamo are too few in number. We may for instance recall that the success of the two rotor model of Herzenberg (1958) depended on a somewhat delicate argument, whereas the three rotor model of Gibson (1968) functioned in a straightforward manner. Again, we may remember that the inductive interaction of N thin rectilinear streams of conducting fluid, an idealized dynamo model devised by Gailitis (1969), will not regenerate if $N = 2$ and only under somewhat special circumstances for $N = 3$. For larger values of N there appears to be no obstacle in sustaining the field if the flows are fast enough.

Table 4. Increase in the complexity of motion

($\epsilon_1 = 0.03$, $\epsilon_2 = \epsilon_3 = 0.04$.)

	N			
p/π	5	6	7	8
2.5	4733	5049	5191	5122
3	3851	3856	3883	3856
4	2830	2691	2662	2647
5	2460	2190	2145	2131
6	—	1948	1879	1886
7	—	1720	1657	1760
10	—	1238	1210	1629
15	—	1090	1169	1606
20	—	1215	1313	—
30	—	1530	1656	—
		1544†	—	—
50	—	2151	2190	—
		2322†	—	—

† $M = 200$ (otherwise $M = 120$).

An inspection of the expressions (26), (29), (30), (33) and (34) for α suggests that, for a given asymmetric flow, α will be the greater the larger the value of p, i.e. the more numerous the cells between the centre and surface of the sphere. This idea cannot be pursued too far, however, for the Braginsky theory is inapplicable when the Reynolds number of the asymmetric scales of motion becomes too small.

Consider instead, therefore, the turbulent dynamo models of Steenbeck *et al.* (see Roberts & Stix 1971) or the spatially periodic flows of Childress (1967, 1970) and G. O. Roberts (1970, 1972) in the case of small microscale Reynolds number $R_m \equiv u\lambda/\eta$, where λ is the correlation length or spatial periodicity, u is the maximum microscale velocity and η is the magnetic diffusivity. If L is the wavelength of the mean or macroscale field generated, dynamo action occurs when the macroscale Reynolds number $R_M \equiv uL/\eta$ is of order $(L/\lambda)^{\frac{1}{2}}$, implying that dynamo action on the scale L is most efficient when λ/L is made as large as possible. This idea cannot

be pursued too far, however, for the theory becomes inapplicable when the Reynolds number of the microscale becomes too large. If we identify the radius of our sphere with L, it may not be unreasonable to believe that when p is too large regeneration will be inefficient.

These arguments suggest that there may exist a most efficient scale of motion, from which a Reynolds number of order 1 may be defined (Childress 1969). It is of some interest to note that Gailitis & Freibergs (1972) have been able to locate such an optimal state for the idealized twin vortex dynamo of Gailitis (1970).

Our results obtained by varying p are shown in table 4. In agreement with G. O. Roberts (reported by P. H. Roberts 1971), Gubbins (1973) and Pekeris *et al.* (1973), we found that the dynamo would not function at all for small values of p. At large values of p the convergence became too bad for the Reynolds numbers to be trusted. It may be noted however that for each sufficiently large p the R obtained increases with increasing N and that, if this trend persists until convergence is obtained, the converged value of R will at first decrease with increasing p and then (at large values of p) increase once more, the optimal value of p possibly lying between $p = 20$ and $p = 30$. We found that the axisymmetic fields were surprisingly insensitive to the number of cells postulated between the centre of the sphere and its surface.

6. Co-rotating field and motion

In his study of a convection driven dynamo in a highly-rotating fluid layer, Soward (1974) was able to find a particularly simple solution in which field and motion were steady in a frame that rotated with an angular velocity slightly different from that of the bounding walls, the relative angular velocity being determined by the diffusivities. He could show that the field was excited in this mode at a smaller Reynolds number than for the case of steady convective overturning. As we observed in § 2, we can explore the kinematic aspects of such co-rotating patterns in spherical geometry by choosing, in definition (18) of our model, a non-vanishing λ which makes the angular velocity, $\omega(r, \theta)$, of the axisymmetric toroidal flow take the constant value λ on $r = 1$. Thus, in a frame moving with this angular velocity, the motion and field appear to drift with angular velocity $-\lambda$ about the axis $\theta = 0$.

The results of our numerical integrations are displayed in table 5. It will be seen that there is some evidence for an increase in dynamo efficiency for a small negative value, λ_c, of λ corresponding (in the frame in which $r = 1$ is at rest) to a motion of the asymmetry in pattern round the z-axis in the same sense as the *mean* zonal motion, u_0. Presumably a similar small value of λ_c would also be obtained in the appropriate Braginsky limit. Unfortunately, however, the fact that λ_c is negative implies that the Braginsky model would possess a surface of concentrated regeneration near $r = 1$, with attendant numerical complications that dissuaded us from constructing the new programme necessary.

The sense of drift in Soward's model was decided by the angular velocity, $\boldsymbol{\Omega}$, of his

layer. Internal convective motions created a gradient of α parallel to Ω. This, generalized to the context of the Earth's core, would have produced positive α in the northern hemisphere and negative in the southern. It may be seen, however, from our definitions (26), (29), (30), (33) and (34) that (for $\epsilon_2\epsilon_3 > 0$) our α is predominantly negative in the north and positive in the south. This indeed appears to be the choice that, with the assumed negative $\omega'(r)$ implied by (35), leads to modes of preferred dipole, rather than quadrupole, parity: see Roberts (1972). (A perfunctory search for modes of quadrupole symmetry for positive $\epsilon_2\epsilon_3$ and similar values of R was, in fact, unsuccessful.)

To make comparisons with Soward's model, we should reverse the roles of the hemispheres by reversing z. The drift of the asymmetries of flow and field is then in both cases westward. It is, however, disturbing that the concommitant reversal of $\omega'(r)$ leads to an average *mean* motion of the core to the west and to a positive $\omega'(r)$.

TABLE 5. TIME VARYING FIELDS AND MOTIONS

($\epsilon_1 = 0.04$, $\epsilon_2 = \epsilon_3 = 0.05$, $p = 3\pi$, $M = 120$.)

	N		
λ	5	6	7
0.05	3120	3144	3182
0.02	2888	2886	2903
0.00	2780	2765	2773
−0.02	2731	2705	2704
−0.05	2872	2845	2818

7. SOME QUANTITIES OF GEOPHYSICAL INTEREST

In view of the diversity of kinematic dynamo models possible, it would be unwise to try to impress any one of them too firmly onto the theory of the main geomagnetic field. It is natural to wonder, however, what inferences could be drawn from our integrations if our model happened to resemble the geodynamo. For example, we might enquire how large the toroidal field in the core would be. We might also ask how rapidly ohmic dissipation would heat the core, and so forth.

As a preliminary, it is necessary to estimate both σ, the conductivity of the core in S/m ($= 10^{11}$ e.m.u.) and v, the typical velocity in m/s used to define

$$R = \mu\sigma v a = 4.36\sigma v.$$

Here $\mu = 4\pi \times 10^{-7}$ H/m is the permeability and $a = 3.47 \times 10^6$ is the core radius in m.

There have been a number of estimates of σ. Bullard & Gellman (1954) took $\sigma = 3 \times 10^5$, which they regarded as uncertain to within a factor of 3 but more probably too small than too large. Braginsky (1964*c*) indeed, arguing by extrapolation of experiments at high pressures in the laboratory, gave $\sigma = 6.7 \times 10^5$. Later, however, after using geophysical observations to estimate the fundamental diffusive time-scale of the geodynamo at 9000 and relating this to a theory of core induction

that allowed for the inner body, he revised his estimate downwards to 3.3×10^5 (Braginsky 1971).

Several authors have suggested smaller values of σ. These seem attractive from two points of view. First, for given v, the toroidal field created by the zonal flow in the core is proportional to σ and the ohmic dissipation, which is dominated by the decay of this predominantly T_2 mode, will be proportional to σ also. Order of magnitude arguments with $\sigma = 3 \times 10^5$ indicate an embarrasingly large Joule heating compared with the observed heat flux from the Earth. This has led Braginsky (1964*b*) for example to argue on thermodynamic grounds that the core motions are produced by non-thermal convection, rather than therma lconvection which may be limited by Carnot efficiency. Other authors, particularly Busse (1975) in a recent paper, have cogently urged that the toroidal field is not as large as supposed in much current theory. As a second reason for smaller conductivities one can point to the geophysical evidence (Bullard 1968; Hertzler *et al.* 1968) that, when a reversal occurs, the axial dipole changes its sign in a time of the order of 1000 a. If this reversal of global field is a diffusive process (as apparently is the case for the solar cycle; see, for example, Steenbeck & Krause 1969*a*) its duration should be of the same order as the fundamental free decay time $0.03\mu\sigma a^2$ of the dominant T_2 magnetic mode, implying that $\sigma = 7 \times 10^4$. This analogy with the solar dynamo raises, however, another consideration, namely the role of turbulence in the core.

The highly irregular motions seen in the solar photosphere have motivated the development of a theory of turbulent dynamo action that has led to significant advances in understanding of cosmical magnetism in the last twenty years. Greatest emphasis has been placed on the generation of large-scale 'mean fields', $\langle \boldsymbol{B} \rangle$, by induction on the microscale, a subject now known as 'mean field electrodynamics'. As we mentioned in § 1, it was recognized early by Parker (1955) and more recently by Steenbeck & Krause (1966) that, if the microscale motions have sufficient helicity a felicious word first applied by Moffatt, 1970), that is if their velocity and vorticity vectors are systematically and sufficiently correlated, they would give rise to *mean* electromotive forces that were related to the *mean* fields in ways completely unknown to classical electromagnetic theory; they showed that such forces could regenerate the mean field without difficulty. Lacking helicity, it appeared that microscale motions would only enhance the rate at which macroscale fields, $\langle \boldsymbol{B} \rangle$, decay (see, for example, Krause & Roberts 1973*a*, *b*).

The possibility of regeneration by helicity, though interesting, remains academic while an origin for helicity is unclear. Steenbeck, Krause & Rädler (1966) showed that the action of Coriolis forces could generate helicity in non-homogeneous turbulence. Other effects being equal, Coriolis forces will be the more effective the greater the scale of motion. Thus, for example, it is sometimes held that only scales of motion comparable in spatial extent to the sector structure and the super-granulation have sufficient helicity to contribute significantly to regeneration of the solar field. The more isotropic motions on the scale of the granulation merely enhance the rate at which the mean field diffuses.

Although there is some danger of oversimplification, the attitude is often taken, when considering mean field generation, that the smallest (non-helical) motions can be ignored if a new 'turbulent electrical conductivity', σ_T, is introduced. The ratio σ_T/σ is some measure of the vigour of the non-helical microscale motions that σ_T replaces. Steenbeck & Krause (1969*a*) have estimated that $\sigma/\sigma_T = 3 \times 10^4$ in the solar convection zone. If, as suggested earlier, we take $\sigma = 3 \times 10^5$ and $\sigma_T = 7 \times 10^4$ in the Earth's core, we obtain $\sigma/\sigma_T = 4$, a value also given by Steenbeck & Krause (1969*b*) by very different considerations. The inference would seem to be that, despite the probable size of dynamical parameters such as Reynolds number and Rayleigh number, the smallest scales of turbulence in the core of the Earth, unlike those of the solar convection zone, do not give rise to large electrodynamic effects. This is consistent with the dynamical arguments of Braginsky (1964*b*, §5). An observational test is precluded by the remoteness of the core surface and the intervening electromagnetic screening.

The uncertainty in v is probably greater than that in σ. Both direct and indirect estimates have been made. The direct approach notes that the time scales associated with secular variation foci are short compared with the ohmic diffusion time of structures of their spatial extent. The conductivity of the core is then effectively infinite over these times, and the field lines emanating from the core are frozen to the fluid elements in their motion along the core surface (Roberts & Scott 1965). To deduce these motions from magnetic observations introduces problems of non-existence and non-uniqueness that are still an obstacle to progress (Roberts & Scott 1965; Backus 1969). Attempts by Vestine and his collaborators (for example, Kahle, Ball & Vestine 1967) gave surface velocities typically of order 2×10^{-4} m/s. This may be compared with the velocity 4×10^{-4} m/s at the core equator implied by a westward drift of 0.2°/a. These velocities may be typical of flow in some boundary layer near the core surface. They are likely to underestimate the zonal flow at depth.

To summarize the direct methods we list a few values of R that it implies, not forgetting that in our model the maximum fluid velocity, v_{max}, is approximately $0.385v$. Thus an estimate of 2×10^{-4} m/s for the maximum core velocity implies a value of v of 5×10^{-4}. For $\sigma = 3 \times 10^5$, we then have $R = 650$, but for $\sigma = \sigma_T = 7 \times 10^4$ we have $R = 150$ only.

The estimates of v just made imply that Coriolis forces, associated with the toroidal flow, and Lorentz forces derived from the toroidal field are of the same order. It seems likely that the same would be true of the non-geostrophic part of the toroidal flow, i.e. the part that cannot be balanced by pressure forces. The indirect method of estimating v supposes that this equality is not fortuitous but is a consequence of the core magnetohydrodynamics. This, as we will now see, gives a value of v dependent on $\sigma (v \propto \sigma^{-2})$.

The ratio $\bar{B}_\phi/|\overline{\boldsymbol{B}}_p|$ is of order R in the Braginsky limit. Indeed, if we take B_{max} to be the largest value of $\bar{B}_\phi$ and B_0 to be the polar value of $|\overline{\boldsymbol{B}}_p|$ on $r = 1$, we find that (for the model $\epsilon_1/\epsilon_2\epsilon_3 = 25$ on which we concentrated in §4) $B_{max}/B_0 = 0.03R$. In

dimensional form, taking B_0 from the axial geomagnetic dipole ($= 3.9G$), we obtain $B_{\max} = (0.5\sigma v)\,G$. ($1G \equiv 10^{-4}\,T$.) The maximum Lorentz force per unit volume is of order $B^2_{\max}/\mu a$. Making more precise comparisons with the model of § 4 we find it is $(4.6 \times 10^{-8}\sigma^2 v^2)$ N/m^3. If we equate this with the Coriolis force $2\Omega\rho v_{\max}$, where $\rho = 10^4$ k/m^3 is the core density and Ω is the angular velocity of the Earth, we obtain $\sigma^2 v = 1.2 \times 10^8$. If we take $\sigma = 3 \times 10^5$ we obtain $R = 1770$, $B_{\max} = 200$G and $v_{\max} = 5 \times 10^{-4}$ m/s. Alternatively, with $\sigma = \sigma_T = 7 \times 10^4$ we have $R = 7600$, $B_{\max} = 860$G and $v_{\max} = 10^{-2}$ m/s, a somewhat large value according to the direct method, even allowing for the likely increase of zonal flow with depth in the core.

In most of the cases described above R is not as large as our numerical integrations suggested. It was found that, with p held fixed, convergence ultimately deteriorated with increasing departures from symmetry (i.e. with increasing ϵ_2 and ϵ_3), and the suggestion was strong that the models ultimately failed if made too asymmetrical. As we observed in § 5, solutions are more readily obtained and have smaller Reynolds numbers if p is increased beyond the value of 3π on which we concentrated. If this experience can be extrapolated to other models, it may perhaps be that the small values of R inferred above imply that the number of inductive centres in the core is large. It is interesting to note that $\bar{B}_{\max}$ is not as great as one might have expected for such large values of R. This is a consequence of the comparatively small value 0.03, of $\bar{B}_{\max}/R\bar{B}_0$ in our model (see § 4). It has the consequence that the ohmic generation of heat is significantly less than might be anticipated from the large value of R and order of magnitude arguments.

Table 6 gives the magnetic energy and ohmic power dissipated by our dynamo model in a number of different cases. The individual contributions from the toroidal and poloidal fields is shown; in the case of the poloidal magnetic field, the contribution from the exterior of the conductor is included. The convergence of the results is indicated by two footnotes to the table. The departures shown by the first and last rows from the general trend appear to be due to deteriorating convergence at these extremes. The last column of the table gives the ratio of the total energy to the total dissipation. It may be compared with the value 1.49×10^{-2} for the longest lived $n = 2$ toroidal decay mode, which forms the largest component of our solutions, in cases of high symmetry.

The eigenfunctions, on which table 6 draws, are normalized to a dimensionless dipole moment m of $S_1(1) = -0.001$. To translate to geophysical magnitudes, it is necessary to identify this with the observed moment 8.07×10^{25} G cm^3 of the Earth's field. The dimensionless units of field, energy, and ohmic dissipation are then 1930 G, 1.24×10^{24} J and $(8.20 \times 10^{16}/\sigma)$ W respectively; taking $\sigma = \sigma_T = 7 \times 10^4$ the last of these is 1.2×10^{12} W. This may be compared with the total heat flux from the Earth, which has been assessed at $1\text{–}3 \times 10^{13}$ W, and is thought to be principally due to the radioactivity of the crust. Looking at the penultimate column of table 6 we see that the dynamos of lower symmetry and larger p have significantly less difficulty in making ohmic dissipation smaller than the surface heat flux. If instead of

TABLE 6. ENERGY AND POWER INTEGRALS

($\lambda = 0$, $M = 120$, $N = 8$.)

definition of case				energy			dissipation			$10^3 \frac{\text{energy}}{\text{dissipation}}$
p/π	ϵ_1	$\epsilon_2 = \epsilon_3$	R	toroidal	poloidal	total	toroidal	poloidal	total	
3	0.0225	0.03	6180	7.76×10^{-2}	3.02×10^{-4}	7.79×10^{-2}	9.14	0.155	9.29	8.39
	0.03	0.04	3856	3.19×10^{-2}	2.12×10^{-4}	3.21×10^{-2}	3.70	0.0960	3.79	8.47
	0.0625	0.05	2808†	1.49×10^{-2}	1.97×10^{-4}	1.51×10^{-2}	1.94	0.0746	2.01	7.49
	0.25	0.10	1239	2.08×10^{-4}	1.88×10^{-4}	2.26×10^{-3}	0.347	0.0411	0.389	5.82
2.5	0.03	0.04	5122	5.15×10^{-2}	3.07×10^{-4}	5.18×10^{-2}	6.25	0.139	6.39	8.10
3			3856	3.19×10^{-2}	2.12×10^{-4}	3.21×10^{-2}	3.70	0.0960	3.79	8.47
4			2647	1.48×10^{-2}	1.21×10^{-4}	1.49×10^{-2}	1.68	0.0525	1.74	8.58
5			2131	8.47×10^{-3}	8.46×10^{-5}	8.55×10^{-3}	1.009	0.0331	1.042	8.21
6			1886	5.79×10^{-3}	6.76×10^{-5}	5.85×10^{-3}	0.736	0.0229	0.759	7.71
7			1760‡	4.47×10^{-3}	5.87×10^{-5}	4.53×10^{-3}	0.607	0.0168	0.624	7.26
15			1606	3.12×10^{-3}	4.14×10^{-5}	3.16×10^{-3}	0.412	0.0130	0.425	7.44

Notes on convergence:

† For $N = 6$, $R = 2734$ and the successive entries are 1.52×10^{-2}, 1.97×10^{-4}, 1.54×10^{-2}, 1.95, 0.0726, 2.02, 7.60.

‡ For $N = 6$, $R = 1720$ and the successive entries are 4.31×10^{-3}, 5.89×10^{-5}, 4.37×10^{-3}, 0.578, 0.0160, 0.594, 7.36.

the turbulent conductivity we take $\sigma = 3 \times 10^5$, i.e. the molecular conductivity, then the unit of power dissipation is 2.7×10^{11}; the energy problem is then evidently far less acute.

Our thanks are due to the Research Committee of the University of Newcastle upon Tyne for a grant in aid of this research.

References

Backus, G. E. 1958 *Ann. Phys.* **4**, 372.
Backus, G. E. 1969 *Phil. Trans. R. Soc. Lond.* A **263**, 239.
Braginsky, S. I. 1964*a* *Soviet Phys. JETP* **20**, 726.
Braginsky, S. I. 1964*b* *Geomag. Aeron.* **4**, 698.
Braginsky, S. I. 1964*c* *Geomag. Aeron.* **4**, 572.
Braginsky, S. I. 1971 In *Transactions of the XV General Assembly, Moscow, U.S.S.R.*, 1971 (ed. L. R. Alldredge), p. 41. Paris: IUGG Publication Office.
Bullard, E. C. 1968 *Phil. Trans. R. Soc. Lond.* A **263**, 481.
Bullard, E. C. 1971 In *World Magnetic Survey* 1957–1969 (ed. A. J. Zmuda), p. 112. Paris: IUGG Publication Office.
Bullard, E. C. & Gellman, H. 1954 *Phil. Trans. R. Soc. Lond.* A **247**, 213.
Busse, F. H. 1975 To appear.
Childress, S. 1967 *Courant Institute Reports* AFOSR-67-0124 and 0976.
Childress, S. 1969 *Théorie magnetohydrodynamique de l'effet dynamo.* Report from Département Méchanique de la Faculté des Sciences, Paris.
Childress, S. 1970 *J. Math. Phys.* **11**, 3063.
Conte, S. D. 1966 *S.I.A.M. Rev.* **8**, 309.
Cowling, R. G. 1933 *Mon. Not. R. Astr. Soc.* **94**, 39.
Davey, A. 1973 *Q. Jl Mech. & Appl. Math.* **26**, 401.
Elsasser, W. M. 1946 *Phys. Rev.* **69**, 106.
Gailitis, A. 1969 *Magnitnaya Gidrodinamika* **5**, 31.
Gailitis, A. 1970 *Magnitnaya Gidrodinamika* **6**, 19.
Gailitis, A. & Freibergs, Ya. J. 1972 *Proc. 6th Riga Conf. on Magnetohydrodynamics*, vol. 1, p. 193.
Gibson, R. D. 1968 *Q. Jl Mech. Appl. Math.* **21**, 257.
Gibson, R. D. & Roberts, P. H. 1969 In *The application of modern physics to the Earth and planetary interiors* (ed. S. K. Runcorn), p. 577. London: Wiley.
Gubbins, D. 1973 *Phil. Trans. R. Soc. Lond.* A **274**, 493.
Gubbins, D. 1974 *Rev. Geophys. Space Phys.* **12**, 137.
Herzenberg, A. 1958 *Phil. Trans. R. Soc. London.* A **250**, 543.
Heirtzler, J. R., Dickson, G. O., Herron, E. M., Pitman III, W. C., LePichon, X. 1968 *J. Geophys. Res.* **73**, 2119.
Isaacson, E. & Keller, H. B. 1966 *Analysis of numerical methods.* New York: Wiley.
Kahle, A. B., Ball, R. H. & Vestine, E. H. 1967 *J. Geophys. Res.* **72**, 4917.
Keller, H. B. 1968 *Numerical methods for two-point boundary-value problems.* New York: Blaisdell.
Kropachev, E. P. 1971 *Geomag. Aeron.* **11**, 690.
Krause, F. & Roberts, P. H. 1973*a* *Astrophys. J.* **181**, 977.
Krause, F. & Roberts, P. H. 1973*b* *Mathematika* **20**, 24.
Lilley, F. E. M. 1970 *Proc. R. Soc. Lond.* A **316**, 153.
Moffatt, H. K. 1970 *J. Fluid Mech.* **41**, 435.
Moffatt, H. K. (editor) 1973 *J. Fluid Mech.* **57**, 625.
Parker, E. N. 1955 *Astrophys. J.* **122**, 293.
Pekeris, C. L. 1972 *Proc. natn. Acad. Sci. U.S.A.* **69**, 2460.

Pekeris, C. L., Accad, Y. & Shkoller, B. 1973 *Phil. Trans. R. Soc. Lond.* A **275**, 425.
Roberts, G. O. 1970 *Phil. Trans. R. Soc. Lond.* A **266**, 535.
Roberts, G. O. 1972 *Phil. Trans. R. Soc. Lond.* A **271**, 411.
Roberts, P. H. 1971 In *World Magnetic Survey* 1957–1969 (ed. A. J. Zmuda), p. 123. IUGG Publication Office.
Roberts, P. H. 1972 *Phil. Trans. R. Soc. Lond.* A **272**, 663.
Roberts, P. H. & Scott, S. 1965 *J. Geomag. & Geoelec.* **17**, 137.
Roberts, P. H. & Stix, M. 1971 *The turbulent dynamo.* Tech. Note 60, National Center for Atmospheric Research.
Soward, A. M. 1972 *Phil. Trans. R. Soc. Lond.* A **272**, 431.
Soward, A. M. 1974 *Phil. Trans. R. Soc. Lond.* A **275**, 611.
Soward, A. M. & Roberts, P. H. 1975 *Magnitnaya Gidrodinamika* (to appear).
Steenbeck, M. & Krause, F. 1966 *Z. Naturforsch.* **21***a*, 1285.
Steenbeck, M. & Krause, F. 1969*a* *Astronom. Nachr.* **291**, 49.
Steenbeck, M. & Krause, F. 1969*b* *Astronom. Nachr.* **291**, 271.
Steenbeck, M., Krause, F. & Rädler, K.-H. 1966 *Z. Naturforsch.* **21***a*, 369.

Physics of the Earth and Planetary Interiors, 20 (1979) 192–193

ARE PLANETARY DYNAMOS DRIVEN BY GRAVITATIONAL SETTLING?

DAVID E. LOPER [1] and PAUL H. ROBERTS [2]

[1] *Geophysical Fluid Dynamics Institute, Florida State University, Tallahassee, FL (U.S.A.)*
[2] *School of Mathematics, University of Newcastle upon Tyne, Newcastle upon Tyne (Great Britain)*

(Accepted for publication in revised form April 5, 1979)

Loper, D.E. and Roberts, P.H., 1979. Are planetary dynamos driven by gravitational settling?, Phys. Earth Planet. Inter., 20: 192–193.

The basic features of gravitationally driven convection are summarized, and an expression for the power available to drive a hydromagnetic dynamo is given. Some thermal consequences of gravitationally driven convection are presented.

Several of the planets are known to possess magnetic fields. It is generally believed that these fields are sustained by dynamo action within the planetary interiors. One of the most challenging problems of planetary physics concerns the energy source for these dynamos. The two energy sources mentioned most often in the literature are thermal convection and precessionally induced motions, but each of these mechanisms encounters difficulties in efficiently converting the available energy into mechanical motion: in a metallic fluid such as the Earth's core, thermal convection may be short-circuited by thermal conduction, while the energy of precession appears to be dissipated in boundary layers. There is a third possible energy source which is not subject to these problems of efficiency: the energy released by gravitational settling. This idea was first proposed by Braginsky (1963) and has recently been discussed by Gubbins (1977) and Loper (1978a, b). A complete theory of dynamo operation driven by gravitational settling is currently being developed (Loper and Roberts, 1978).

The basic features of gravitationally driven convection may be summarized as follows. Consider an initially homogeneous self-gravitating sphere of molten binary alloy representing a planetary core, the alloy being composed of a heavy metal (principally iron) and a light non-metal in the case of a terrestrial planet, or of helium and metallic hydrogen in the case of a Jovian planet. As the planet cools over geological time, a phase change is likely to occur, either to a new liquid phase immiscible with the first phase, or to a solid phase. Assuming the adiabatic gradient to be smaller than the liquidus gradient, the phase change would first occur at the center of the sphere. Metallurgical studies have shown that the new phase which forms from a binary melt does not in general have the same composition as the liquid. To simplify the discussion we shall assume the melt to contain a larger fraction of the heavy component than does the eutectic. In this case, the new phase has a larger fraction of the heavy component and hence is more dense than the initial liquid, even in the absence of a density change associated with the phase change. The denser phase will accumulate at the center, forming an inner core. Associated with the growth of this inner core, there is a net motion of heavy material downward and light material upward, releasing gravitational potential energy. More specifically as the inner core grows by accretion of the denser phase, a residue of light material is left in the liquid near the inner core. Since diffusion of matter is ineffective over planetary length-scales, this excess of light material must distribute

itself throughout the outer core by means of convective motions driven by compositional buoyancy. The fact that the material diffusion coefficient is typically much smaller than the thermal diffusion coefficient accounts for the inherent efficiency of the gravitationally driven dynamo, first pointed out by Bragin Braginsky (1963) and qualified by Gubbins (1977).

The power available from gravitational settling depends upon the density contrast $\Delta\rho$ between the two phases and may be crudely estimated by the formula:

$$P = \Delta\rho\ VgL/t$$

where V is the inner core volume, L is the outer core radius, g is a typical value of the local acceleration of gravity and t is the time which has elapsed since the inner core began to form. Typical estimates for the Earth's core give $P = 4 \times 10^{11}$ W, which is sufficient to sustain a large field (see Loper, 1978a). In the light of the uncertainty in composition and structure of the planetary cores, it is difficult at present to estimate the viability of gravitational settling as a power source for the other planetary dynamos.

If gravitationally-driven convection occurs, the outer core is in effect mechanically mixed. Loper (1978b) has shown that several interesting thermal regimes are possible. For example, if the liquidus gradient is less than the conduction gradient (i.e., the gradient which would occur in the absence of motion), a slurry must occur in the fluid at the bottom of the outer core. We have recently been able to show that the Earth's core must contain a slurry. A general thermodynamic theory for the motion of a binary-alloy fluid containing a slurry has been developed by Loper and Roberts (1978). Also, compositionally driven convection can occur even if the actual temperature gradient is less than the adiabatic gradient, making the fluid thermally stably stratified. In this case the fluid motions convect heat radially inward – up the temperature gradient. Hence the heat conducted down the adiabat cannor be taken as a lower limit on the heat transferred out of the core.

Acknowledgements

This work was supported by the National Science Foundation under grant No. EAR 78–00818 and is contribution No. 147 of the Geophysical Fluid Dynamics Institute, Florida State University, Tallahassee, Florida.

References

Braginsky, S.I., 1963. Structure of the F layer and reasons for convection in the Earth's core. Dokl. Akad. Nauk. SSSR, 149: 8–10.

Gubbins, D., 1977. Energetics of the Earth's core. J. Geophys., 43: 453–464.

Loper, D.E., 1978a. The gravitational powered dynamo Geophys. J. R. Astron. Soc., 54: 389–404.

Loper, D.E., 1978b. Some thermal consequences of a gravitationally powered dynamo. J. Geophys. Res., 83: 5961–5970.

Loper, D.E. and Roberts, P.H., 1978. On the motion of an iron-alloy core containing a slurry. I. General theory. Geophys. Astrophys. Fluid. Dyn., 9: 289–321.

Physics of the Earth and Planetary Interiors, 24 (1981) 302–307
Elsevier Scientific Publishing Company, Amsterdam – Printed in The Netherlands

A STUDY OF CONDITIONS AT THE INNER CORE BOUNDARY OF THE EARTH *

DAVID E. LOPER [1] and PAUL H. ROBERTS [2]

[1] *Department of Mathematics and Geophysical Fluid Dynamics Institute, Florida State University, Tallahassee, FL 32306 (U.S.A.)*
[2] *School of Mathematics, The University, Newcastle upon Tyne NE1 7RU (Great Britain)*

(Received February 26, 1980; accepted for publication June 27, 1980)

Loper, D.E. and Roberts, P.H., 1981. A study of conditions at the inner core boundary of the Earth. Phys. Earth Planet. Inter., 24: 302–307.

A study is made of the thermal and compositional conditions which the liquid outer core must satisfy at the inner core boundary, assuming the inner core to be growing by continual solidification of the heavy component of the liquid alloy in the outer core. It is found that the outer core is strongly destabilized by the compositional gradients driven by the separation process associated with the freezing. Further, it is argued that all the freezing necessary for the growth of the solid inner core cannot occur on a flat interface; most of it must occur above the solid boundary in a region labeled the slurry layer.

1. Introduction

The F-layer in the Earth's core was postulated by Jeffreys in 1939 to explain the short-period precursors to the phase PKIKP at epicentral distances between 125 and 143° which had been discovered by Gutenberg (Jeffreys, 1976). The original explanation was in terms of a region lying between the liquid outer core and the solid inner core in which the compressional wave speed was anomalously low. Subsequently Bolt (1962, 1964) suggested that the speed in the F-layer was anomalously high with a sharp jump at the top of the layer. These explanations were not completely satisfactory because the data were insufficient to determine the depth and structure of the layer and no theory was available. When better data became available with the advent of large seismic arrays such as NORSAR, the evidence resolved the question of the precursors in favor of scatterers at or near the mantle–core boundary, and the need for an F-layer to explain these seismic observations was removed (Cleary and Haddon, 1972; King et al., 1973; Haddon and Cleary, 1974; Husebye et al., 1976). Consequently, most Earth models constructed in the last few years have omitted the F-layer (Dziewonski et al., 1975; Gilbert and Dziewonski, 1975; Anderson and Hart, 1976). However, the seismic data do not rule out the possibility of an F-layer (Haddon, private communication, 1979) and recently Stevenson (1980) has presented evidence which appears to support the view that the lower portion of the outer core is of a different composition than the upper portion. This evidence is a plot of dK/dp against p for several Earth models which shows a sharp drop near the inner core boundary. Unfortunately, the strength of this evidence is considerably weakened if the data from model 1066A of Gilbert and Dziewonski (1975) are included, since this model gives a sharp rise near the inner core boundary. We are forced to conclude that the composition and thermal regime of the lower portion of the outer core are unknown at present.

In this paper we shall study the portion of the outer core immediately above the solid inner core

* Contribution no. 161, Geophysical Fluid Dynamics Institute.

from the thermodynamic point of view and present a new theoretical argument which strongly suggests that this region may indeed be anomalous. The core will be modeled as a binary alloy composed of heavy constituents (Fe and Ni) and a light constituent (FeS, FeSi or FeO), with the precise nature of the constituents being irrelevant to the analysis; and it will be assumed that the solid inner core has formed by freezing of material from the liquid. This model has been developed in part to explain the power source for the geodynamo (Braginsky, 1963; Loper, 1978a, b). The growth of the inner core is a continual process as the Earth gradually cools at a rate of 50–100°C/10^9 y (Davies, 1979; Schubert and Stevenson, 1979; Stacey, 1980). This growth implies certain boundary conditions on the liquid outer core. We shall develop these boundary conditions and investigate their consequences for the thermal regime of the outer core.

2. Boundary conditions for the outer core

We begin by supposing the core to be cooled sufficiently rapidly that convection occurs and the temperature gradient of the liquid outer core to be nearly adiabatic away from rigid boundaries

$$\mathrm{d}T_A/\mathrm{d}p = \alpha T/\rho C_P \tag{1}$$

where p is the pressure, T the temperature, ρ the density, α the coefficient of thermal expansion and C_P the specific heat at constant pressure. The mixing associated with the convective motion maintains the liquid alloy in a state of nearly constant composition and the gradient of the liquidus temperature is given by the Clausius–Clapeyron equation

$$\mathrm{d}T_L/\mathrm{d}p = T\,\delta/L \tag{2}$$

where L is the latent heat and δ the change of specific volume upon melting. We shall assume that

$$\alpha L < C_P \rho\,\delta \tag{3}$$

so that the adiabat intersects the melting curve at the bottom of the liquid and the "core paradox" (Kennedy and Higgins, 1973) is avoided. Assuming Lindemann's law to be valid, (3) is satisfied provided that $\gamma > 2/3$ where $\gamma = \alpha K/\rho C_P$, K being the adiabatic incompressibility. Current opinion (Irvine and Stacey, 1975; Stevenson, 1980) is that (3) is well satisfied within the core.

Immediately above the rigid inner core boundary convective motions are suppressed and the latent heat released by the freezing process must be removed from the interface by conduction into the fluid, giving

$$k\,\mathrm{d}T_c/\mathrm{d}r = -L\dot{m} \tag{4}$$

where k is the thermal conductivity, r the radius and $\dot{m}$ the mass per unit area per unit time of material frozen.

In writing (4) we have neglected the heat which must be conducted from the solid to the liquid as the inner core gradually cools. This is much smaller than the latent heat and has been ignored for simplicity. (Its inclusion would only strengthen the conclusion we shall obtain.) Also in writing (4) we have assumed all freezing to occur on a flat interface separating the solid and liquid. We shall show that this assumption leads to a contradiction and discuss the resolution of the apparent paradox. To a good approximation the Earth is in hydrostatic balance

$$\mathrm{d}p/\mathrm{d}r = -\rho g \tag{5}$$

allowing (4) to be written as

$$\mathrm{d}T_c/\mathrm{d}p = L\dot{m}/\rho g k \tag{6}$$

Let us now consider the compositional effects of the freezing process, assuming that the mass fraction ξ of the lighter constituent is less than the eutectic value ξ_e. [In a similar study of the F-layer, Braginsky (1963) assumed $\xi > \xi_e$ but his model encounters serious difficulties in removing the latent heat from the freezing interface (Loper, 1978b).] It is known from metallurgical studies that the solid which crystallizes from such a liquid has a mass fraction ξ_s of light component less than that of the liquid. For simplicity we shall assume that $\xi_s = 0$. As a liquid layer of mass $\mathrm{d}m$ per unit area freezes on to the solid inner core in time $\mathrm{d}t$, the light mass $\xi \mathrm{d}m$ must be expelled into the liquid above and replaced by heavy material. This generates a vertical flux i of light material relative to the mass center given by

$$i = \xi\dot{m} \tag{7}$$

To express this boundary condition in terms of the compositional gradient, we note first that quite generally (Loper and Roberts, 1980, eq. 2.1) the vector flux $\mathbf{i}$ is proportional to $\nabla\mu$ and ∇T where the relative chemical potential μ is defined as

$$\mu = \mu_1/M_1 - \mu_2/M_2 \tag{8}$$

where μ_1 is the chemical potential of the light constituent and M_1 its molecular weight, with a subscript 2 denoting the heavy constituent. The diffusion i driven by the temperature gradient, known as the Soret effect is assumed small and will be ignored. Further we shall assume * that $\partial\mu/\partial T = 0$, giving

$$i = -\rho D\left(\frac{\mathrm{d}\xi}{\mathrm{d}r} + \frac{\bar{\delta}}{\bar{\mu}}\frac{\mathrm{d}p}{\mathrm{d}r}\right) \tag{9}$$

where $\bar{\delta} = \partial\mu/\partial p$, $\bar{\mu} = \partial\mu/\partial\xi$ and D is the coefficient of diffusion of the light component. This is identical to the model used by Braginsky (1964); his $\bar{\alpha}$ is the same as $\bar{\delta}/\rho$. We may combine (5), (7) and (9) to obtain an expression for the compositional gradient at the bottom of the liquid

$$\frac{\mathrm{d}\xi}{\mathrm{d}p} = -\frac{\bar{\delta}}{\bar{\mu}} + \frac{\xi\dot{m}}{\rho^2 g D} \tag{10}$$

If $\dot{m}$ were zero, (10) would yield a negative gradient of ξ with respect to p, indicating a stable chemical layering near the bottom of the outer core. As $\dot{m}$ increases from zero, this gradient is diminished and eventually reversed, suggesting the possibility of a convective instability.

To study this instability, we introduce the equation of state for density

$$\mathrm{d}\rho = \beta\rho\ \mathrm{d}p - \alpha\rho\ \mathrm{d}T - \bar{\delta}\rho^2\ \mathrm{d}\xi \tag{11}$$

where β is the isothermal compressibility and, through a Maxwell relation, the $\bar{\delta}$ in (11) is the same as that introduced in (9). The actual density gradient immediately above the inner core boundary may be determined from (11), using (6) and (10), to be

$$\frac{\mathrm{d}\rho}{\mathrm{d}p} = \beta\rho + \frac{(\rho\bar{\delta})^2}{\bar{\mu}} - \left(\frac{\alpha L}{k} + \frac{\xi\bar{\delta}}{D}\right)\frac{\dot{m}}{g} \tag{12}$$

On the other hand, if a parcel of fluid is moved adiabatically, holding its composition constant, its density varies as

$$\mathrm{d}\rho_A/\mathrm{d}p = \beta\rho - \alpha^2 T/C_P \tag{13}$$

If the parcel moves upward and becomes less dense than its surroundings, convective overturning is possible. This occurs if $\mathrm{d}\rho/\mathrm{d}p < \mathrm{d}\rho_A/\mathrm{d}p$ or if

$$\frac{\alpha^2 T}{C_P} + \frac{(\rho\bar{\delta})^2}{\bar{\mu}} < \left(\frac{\alpha L}{k} + \frac{\xi\bar{\delta}}{D}\right)\frac{\dot{m}}{g} \tag{14}$$

Strictly speaking (14) is a necessary condition for instability but not sufficient; the freezing rate $\dot{m}$ must exceed (14) sufficiently to overcome the dissipation associated with the convective motion. However, for the length scales relevant to the core, this latter effect is small and it has been shown by Loper and Roberts (1980) that convective instability occurs when (14) is only marginally satisfied.

The terms on the left-hand side of (14) represent the stabilizing effects of the conduction of heat down the adiabat and the diffusion of material induced by the pressure gradient, respectively, while the terms on the right-hand side represent the destabilizing effects of latent heat release and the release of light material associated with the freezing process. We now wish to determine which of these terms are important and whether the inequality is satisfied in the core. This requires estimation of the parameters contained in (14).

3. Parameter estimation

The estimation of parameter values under core conditions is an uncertain process and only the fact that the inequalities which we shall obtain hold so strongly allows us to place some confidence in the conclusions. We begin with several quantities from standard Earth models (e.g. Jacobs, 1975, table 1.3): $\rho = 12 \times 10^3$ kg m^{-3} and $g = 4.3$ m s^{-2}. From Stacey (1977a) we adopt $T = 4000$ K, $C_P = 660$ J kg^{-1} K^{-1}, $k = 35$ W m^{-1} K^{-1}, and $\alpha = 8 \times 10^{-6}$ K^{-1}, while from Stacey (1977b) we have $L = 8 \times 10^5$ J kg^{-1} and $\rho\delta = 0.012$. The composition and properties of the light constituent in the core are the subjects of speculation. Loper (1978a) has estimated that $\rho\bar{\delta} \approx 1.1$ for silicon and 2.4 for sulfur, although oxygen is also a possibility (Ringwood, 1977). We shall take the former value since it yields a more conservative estimate in (14). Usselman (1975) has estimated that $\xi = 0.1$, but allowing for the possibility of alloying in the solid, we shall take a smaller value: $\xi = 0.05$. From ideal solution

* The magnitudes of these thermal effects are poorly known within the core, but it is likely that they will be small compared with the terms retained in (9).

theory (Guggenheim, 1952), we have * $\bar{\mu} \approx 3RT/M\xi(1-3\xi)$ where $R = 8314$ J K^{-1} kmol^{-1} and $M \approx 50$ kg kmol^{-1}, giving $\bar{\mu} \approx 4.4 \times 10^7$ J kg^{-1}. From both theory (Frenkel, 1946) and experiment (Majdic et al., 1969; Calderon et al., 1971) we have $D \approx 3 \times 10^{-9}$ m^2 s^{-1}; this is considerably larger than Braginsky's (1964) estimate and accordingly yields a more conservative stability criterion in (15). To determine $\dot{m}$, we may estimate the mass of the solid inner core to be 10^{23} kg from a crude integration of Jacobs' (1975) Table 1.3. Davies (1979) has estimated that the solid inner core began to form roughly 3×10^9 y ago. Assuming the solid to have formed at a constant rate since then, we have, using a current inner core radius of 1215 km, $\dot{m} = 5 \times 10^{-8}$ kg m^{-2} s^{-1}. For ease of reference, these parameter estimates have been compiled in Table I.

With the parameter estimates listed, we find that the thermal terms in (14) are negligibly small

$$\frac{\alpha^2 T}{C_P}\left(\frac{(\rho\bar{\delta})^2}{\bar{\mu}}\right)^{-1} \approx 1.4 \times 10^{-2},$$

$$\frac{\alpha L}{k}\left(\frac{\xi\bar{\delta}}{D}\right)^{-1} \approx 1.2 \times 10^{-4}$$

The stability criterion (14) now simplifies to

$$g\rho^2\bar{\delta}D/\bar{\mu}\xi < \dot{m} \tag{15}$$

The left-hand side of (15) is approximately 8×10^{-11} kg m^{-2} s^{-1}. This is the critical mass rate of freezing for instability to occur. Our estimate of $\dot{m}$ is more than 600 times larger than critical. Therefore, the core is likely to be in vigorous convective motion and any analysis which is based upon a small departure from the critical is not geophysically relevant. The fact that thermal buoyancy is negligible leads us to conclude that the dynamo is driven primarily by the gravitational energy released by the separation process (Loper, 1978a).

In writing (2), it was assumed that the liquid was well mixed with small compositional gradients. This is true away from rigid boundaries, but near the freezing interface motions are suppressed and the compositional gradient may become large. This requires that we account for the variation of the liquidus temperature T_L with composition as well as pressure near the freezing interface. From Loper and Roberts (1980)

$$\frac{dT_L}{dp} = \frac{T\bar{\delta}}{L} - \frac{T\bar{\mu}\xi}{L}\frac{d\xi}{dp} \tag{16}$$

Using (10) this becomes

$$\frac{dT_L}{dp} = \frac{T}{L}(\delta + \xi\bar{\delta}) - \frac{\bar{\mu}\xi^2 T\dot{m}}{L\rho^2 gD} \tag{17}$$

Now let us compare (6) and (17). If the conduction gradient is smaller than the liquidus gradient, the temperature of the liquid at the bottom of the outer core lies above the liquidus and nothing untoward occurs; the core is in regime A of Loper (1978b). On the other hand, if the core is in regime B, i.e. if

$$dT_L/dp < dT_c/dp \tag{18}$$

it would appear that the liquid above the freezing interface is frozen, clearly a contradictory situation.

Before discussing the resolution of this paradox, let us determine whether (18) is satisfied within the core. Combining (6), (17) and (18) we have

$$g\rho(\delta + \xi\bar{\delta}) < (\bar{\mu}\xi^2/\rho D + L^2/kT)\,\dot{m} \tag{19}$$

With the estimates from Table I we have

$$\frac{L^2}{kT}\left(\frac{\bar{\mu}\xi^2}{\rho D}\right)^{-1} \approx 1.5 \times 10^{-3}$$

That is, $dT_c/dp \ll |dT_L/dp|$ close to the freezing interface. Once again thermal effects are small and (19) may be simplified to

$$\frac{g\rho^2\bar{\delta}D}{\bar{\mu}\xi}\left(1 + \frac{\delta}{\xi\bar{\delta}}\right) < \dot{m} \tag{20}$$

This expression is similar to (15). With the given estimates $\delta/\xi\bar{\delta}$ is small, roughly 0.2. It follows that the critical freezing rate for the core to lie in regime B is approximately 20% higher than that for the onset of convection. However, it has been estimated that $\dot{m}$ is 600 times larger than critical for convection, so it is very likely that the bottom of the outer core satisfies condition (18).

* From Loper and Roberts (1978), $\bar{\mu} = \partial^2\Phi/\partial\xi^2$ where Φ is the Gibbs free energy. If we define ξ as the mass fraction of light constituent and x as the mass fraction of a compound such as FeS or FeSi then $x \approx 3\xi$. Noting that $\Phi = \Delta_m F/M$, we may use (3.02.3) of Guggenheim (1952) to obtain $\bar{\mu} = 3RT/M\xi(1-3\xi)$.

TABLE I

Parameter estimates at the inner core boundary

Parameter	Symbol	Estimate	Source
Density	ρ	12×10^3 kg m^{-3}	Jacobs (1975)
Gravity	g	4.3 m s^{-2}	Jacobs (1975)
Temperature	T	4000 K	Stacey (1977a)
Heat capacity	C_P	660 J kg^{-1} K^{-1}	Stacey (1977a)
Thermal conductivity	k	35 W m^{-1} K^{-1}	Stacey (1977a)
Thermal expansion	α	8×10^{-6} K^{-1}	Stacey (1977a)
Latent heat	L	8×10^5 J kg^{-1}	Stacey (1977b)
Change of volume upon melting	$\rho\delta$	0.012	Stacey (1977b)
Change of volume with composition	$\rho\bar{\delta}$	1.1	Loper (1978a)
Mass fraction of light constituent	ξ	0.05	Usselman (1975)
Energy of mixing	$\bar{\mu}$	4.4×10^7 J kg^{-1}	Guggenheim (1952)
Material diffusion coefficient	D	3×10^{-9} m^2 s^{-1}	Frenkel (1946)
Mass rate of freezing	$\dot{m}$	5×10^{-8} kg m^{-2} s^{-1}	

When (18) is satisfied the liquid alloy near the freezing interface is said to be constitutionally supercooled (Chalmers, 1964; Christian, 1965). The resolution of the paradox it implies lies in reducing the amount of freezing directly at the interface and instituting freezing above it. In laboratory experiments this is accomplished by the formation of dendrites extending from the freezing interface into the liquid forming a so-called "mushy zone" (Copley et al., 1970; Ockendon and Hodgkins, 1975). These dendrites are tree-like structures which, under core conditions, are likely to be very tenuous and easily broken. (In fact, Copley et al. (1970) found that the dendrites can be broken by natural-convection flow in the laboratory.) Once broken, the dendrites will tend to move with the fluid, freezing more solid or remelting as conditions dictate. The region occupied by these free-floating dendrites (snowflakes) has been labeled the "slurry layer" (Loper, 1978b; Loper and Roberts, 1978).

Once solidification occurs away from the flat boundary, the parameter appearing in (6) and (17) must be taken to mean the mass rate of freezing directly at this interface rather than the total. Since the fluid above the interface must lie on the liquidus, the latent heat released at the boundary must be removed by conduction down this gradient. This statement has the effect of making (18) and (20) equalities and the mass rate of freezing directly at the boundary must be given by

$$\dot{m} = g\rho^2 D(\delta + \xi\bar{\delta})/\bar{\mu}\xi^2 \qquad (21)$$

Since we have estimated the total mass rate of freezing to be much larger than this, it follows that virtually all of the solidification occurs above the inner core boundary.

The nature of the slurry layer and its consequences for seismology and dynamo theory are currently being investigated.

Acknowledgments

We wish to thank David Stevenson for providing a copy of his paper (Stevenson, 1980) prior to its publication. The assistance of David Fearn in the preparation of Table I and in sharpening the arguments in the paper is gratefully acknowledged. This work was supported by the National Science Foundation under Grant no. EAR 78-00818.

References

Anderson, D.L. and Hart, R.S., 1976. An Earth model based on free oscillations and body waves. J. Geophys. Res., 81: 1461–1475.

Bolt, B.A., 1962. Gutenberg's early PKP observations. Nature (London), 196: 122–124.

Bolt, B.A., 1964. The velocity of seismic waves near the Earth's center. Bull. Seismol. Soc. Am., 54: 191–208.

Braginsky, S.I., 1963. Structure of the F layer and reasons for convection in the Earth's core. Dokl. Akad. Nauk SSSR, 149: 8–10.

Braginsky, S.I., 1964. Magnetohydrodynamics of the Earth's core. Geomagn. Aeron. (USSR), 4: 698–712.

Calderon, F.P., Sano, N. and Matsushita, Y., 1971. Diffusion of Mn and Si in liquid Fe over the whole range of composition. Metall. Trans., 2: 3325–3332; see also Diffusion Data, 6: 72–73 (1972).

Chalmers, B., 1964. Principles of Solidification. Wiley, New York, NY, pp. 154–183.

Christian, J.W., 1965. The Theory of Transformations in Metals and Alloys. Pergamon, Oxford, pp. 572–576.

Cleary, J.R. and Haddon, R.A.W., 1972. Seismic wave scattering near the core–mantle boundary: a new interpretation of precursors to PKP. Nature (London), 240: 549–551.

Copley, S.M., Giamei, A.F., Johnson, S.M. and Hornbecker, M.F., 1970. The origin of freckles in unidirectionally solidified castings. Metall. Trans., 1: 2193–2204.

Davies, G.F., 1979. Timing of core formation and onset of the geomagnetic dynamo in convecting Earth models. Paper 10/30, XVII General Assembly, IUGG, Canberra.

Dziewonski, A.M., Hales, A.L. and Lapwood, E.R., 1975. Parametrically simple Earth models consistent with geophysical data. Phys. Earth Planet. Inter., 10: 12–48.

Frenkel, J., 1946. Kinetic Theory of Liquids. Oxford University Press, London.

Gilbert, F. and Dziewonski, A.M., 1975. An application of normal mode theory to the retrieval of structural parameters and source mechanisms from seismic spectra. Philos. Trans. R. Soc. London, Ser. A, 278: 187–269.

Guggenheim, E.A., 1952. Mixtures. Oxford University Press, London.

Haddon, R.A.W. and Cleary, J.R., 1974. Evidence for scattering of seismic PKP waves near the mantle–core boundary. Phys. Earth Planet. Inter., 8: 211–234.

Husebye, E.S., King, D.W. and Haddon, R.A.W., 1976. Precursors to PKIKP and seismic wave scattering near the mantle–core boundary. J. Geophys. Res., 81: 1870–1882.

Irvine, R.D. and Stacey, F.D., 1975. Pressure dependence of the thermal Grüneisen parameter, with application to the Earth's lower mantle and outer core. Phys. Earth Planet. Inter., 11: 157–165.

Jacobs, J.A., 1975. The Earth's Core. Academic Press, New York, NY.

Jeffreys, H., 1976. The Earth. 6th edn. Cambridge University Press.

Kennedy, G.C. and Higgins, G.H., 1973. The core paradox. J. Geophys. Res., 78: 900–904.

King, D.W., Haddon, R.A.W. and Cleary, J.R., 1973. Evidence for seismic wave scattering in the D″ layer. Earth Planet. Sci. Lett., 20: 353–356.

Loper, D.E., 1978a. The gravitationally powered dynamo. Geophys. J.R. Astron. Soc., 54: 389–404.

Loper, D.E., 1978b. Some thermal consequences of a gravitationally powered dynamo. J. Geophys. Res., 83: 5961–5970.

Loper, D.E. and Roberts, P.H., 1978. On the motion of an iron-alloy core containing a slurry. I. General theory. Geophys. Astrophys. Fluid Dyn., 9: 289–321.

Loper, D.E. and Roberts, P.H., 1980. On the motion of an iron-alloy core containing a slurry. II. A simple model. Geophys. Astrophys. Fluid Dyn., 16: 83–127.

Majdic, A., Graf, D. and Schenck, H., 1969. Diffusion of Si, P, S and Mn in molten Fe. Arch. Eisenhüttenwes., 40: 627–630; see also Diffusion Data, 4: 338–339 (1970).

Ockendon, J.R. and Hodgkins, W.R., 1975. Moving-Boundary Problems in Heat Flow and Diffusion. Clarendon Press, Oxford.

Ringwood, A.E., 1977. Composition of the core and implications for origin of the Earth. Geochem. J., 11: 111–135.

Schubert, G. and Stevenson, D., 1979. Radiogenic heat source contents of the Earth and Moon. Paper 10/14, XVII General Assembly, IUGG, Canberra.

Stacey, F.D., 1977a. A thermal model of the Earth. Phys. Earth Planet. Inter., 15: 341–348.

Stacey, F.D., 1977b. Applications of thermodynamics to fundamental Earth physics. Geophys. Surveys, 3: 175–204.

Stacey, F.D., 1980. The cooling Earth: a reappraisal. Phys. Earth Planet. Inter., 22: 89–96.

Stevenson, D.J., 1980. Applications of liquid-state physics to the Earth's core. Phys. Earth Planet. Inter., 22: 42–52.

Usselman, T.M., 1975. Experimental approach to the state of the core: Part II. Composition and thermal regime. Am. J. Sci., 275: 291–303.

Reprinted from Nature, Vol. 292, No. 5820, pp. 232-233, July 16 1981

Structure of the Earth's inner core

David R. Fearn* & David E. Loper

Geophysical Fluid Dynamics Institute, Florida State University, Tallahassee, Florida 32306, USA

Paul H. Roberts

School of Mathematics, The University, Newcastle upon Tyne NE1 7RU, UK

Jacobs[1] proposed that the Earth's inner core is growing through the freezing of outer-core material as the Earth gradually cools[2]. Recent studies have shown that compositional effects associated with this freezing process can release energy at a rate sufficient to power the geodynamo[3,4] and may be crucial in determining the dynamic state of the outer core[5]. Here we investigate the effects of composition on the freezing process itself, drawing on metallurgical experience, and speculate on the structure and state of the inner core. We propose that the interface separating the inner and outer core is dendritic and argue that the region in which freezing takes place may extend throughout the entire inner core. Consequently the compositionally driven convective motions which stir the outer core and sustain the geodynamo also occur within the interdendritic spaces of the inner core. The seismic evidence which corroborates this proposal is briefly reviewed.

The material composing the Earth's core has often been treated as being pure iron but it is in fact an "uncertain mixture of all the elements"[6]. The freezing of a mixture can be very different in nature from the freezing of a pure substance[7] so it is most important that the effects of composition be included in any study of the evolution of the Earth's core. In particular there are two properties which must be modelled. First the temperature (the liquidus temperature T_L) below which material freezes out of a fluid with a certain composition is a function of that composition. Second, in general, the composition of the solid which freezes is different from that of the fluid from which it has frozen. As a first step towards investigating the effects of composition we shall study the simplest possible system which incorporates these two properties; a binary alloy. We shall assume that as freezing proceeds, one component of the binary mixture goes to form the solid while the other is expelled into the remaining fluid. We shall use ξ to denote the mass fraction of the latter (so by definition $\xi = 0$ in the solid). As already mentioned, the liquidus temperature is a function of composition and we shall use ξ_e to denote that composition ξ for which T_L takes its minimum T_e. In applying the results of studies of binary alloys to the Earth's core, it is very likely[8] that ξ represents the less dense constituents and the solid which freezes is richer in the denser constituents.

A binary mixture in a single liquid phase has, by Gibbs' phase rule, three independent thermodynamic variables. When a solid phase is present in equilibrium with the adjacent liquid, a constraint is placed on the variables reducing the number of independent variables to two. Usually the constraint is expressed as the liquidus condition $T = T_L(p, \xi)$ where T is the temperature of the liquid and p is the pressure. If

$$T < T_L(p, \xi) \tag{1}$$

the liquid is said to be supercooled and is unstable to the growth of solid, usually in the form of dendrites extending from a boundary. In particular if this condition is caused by a spatial gradient of composition normal to a freezing interface, the liquid is said to be constitutionally supercooled. To determine whether equation (1) is satisfied near the inner-core boundary of the Earth, let us assume the boundary to be spherical and $T = T_L(p, \xi)$ on the boundary. Expanding equation (1) as a Taylor series in powers of $p - p_0$, where p_0 is the pressure at the boundary, and keeping only the linear terms allows equation (1) to be rewritten as

$$(T/L)(\delta - \bar{\mu}\xi\, \mathrm{d}\xi/\mathrm{d}p) < \mathrm{d}T/\mathrm{d}p \tag{2}$$

evaluated at the boundary, where L is the latent heat, δ is the specific-volume change on melting, $\bar{\mu}$ is the energy of mixing[9], and

$$\partial T_L/\partial p \equiv T\delta/L, \qquad \partial T_L/\partial \xi \equiv -T\bar{\mu}\xi/L$$

Because fluid motions are inhibited at the rigid freezing boundary, the fluxes $\xi\dot{m}$ of light material and $L\dot{m}$ of heat, generated by the freezing of material at a mass rate $\dot{m}$ per unit area, must be removed from the interface by diffusion. In addition to the usual flux down the gradient, there is an upward barodiffusive flux of light material, giving[9,10]

$$\xi\dot{m} = \rho^2 D g[(\bar{\delta}/\bar{\mu}) + \mathrm{d}\xi/\mathrm{d}p] \tag{3}$$

$$L\dot{m} = k\rho g\, \mathrm{d}T/\mathrm{d}p \tag{4}$$

where ρ is the fluid density, D the material diffusion coefficient, g the acceleration due to gravity, $\bar{\delta}$ the specific-volume change with composition and k the thermal conductivity. In writing equations (3) and (4) we have assumed a hydrostatic balance to replace radial derivatives by derivatives with respect to pressure. Using equations (3) and (4) to eliminate $\mathrm{d}\xi/\mathrm{d}p$ and $\mathrm{d}T/\mathrm{d}p$ from equation (2) yields[11]

$$\dot{m} > \dot{m}_c = (\delta + \xi\bar{\delta})\rho^2 g D/\bar{\mu}\xi^2 \tag{5}$$

In writing equation (5) a thermal term, which can be shown to be negligible, has been neglected. This is the condition for constitutional supercooling to occur in the liquid outer core immediately above a spherical inner-core boundary.

Using parameter estimates from ref. 11 ($\rho = 1.2 \times 10^4\ \mathrm{kg\,m^{-3}}$, $g = 4.3\ \mathrm{m\,s^{-2}}$, $D = 3 \times 10^{-9}\ \mathrm{m^2\,s^{-1}}$, $\bar{\mu} = 4.4 \times 10^7\ \mathrm{J\,kg^{-1}}$, $\rho\bar{\delta} = 1.1$, $\rho\delta = 0.016$, $\xi = 0.05$), the critical growth rate is $\dot{m}_c = 10^{-10}\ \mathrm{kg\,m^{-2}\,s^{-1}}$. On the other hand, assuming the solid inner core formed 3,000 Myr ago and grew at a constant rate since then, we have $\dot{m} = 5 \times 10^{-8}\ \mathrm{kg\,m^{-2}\,s^{-1}}$. By these estimates, the inner core is growing at a rate roughly 500 times faster than the critical value for the occurrence of constitutional supercooling. Even allowing for a considerable variation in the parameter estimates, it appears that if the inner-core boundary were spherical, the fluid immediately above it would be constitutionally supercooled. With the above parameter estimates, the supercooling increases with distance from the boundary at a rate of $0.8\ \mathrm{K\,m^{-1}}$. It is extremely doubtful that such a strongly supercooled fluid could persist for long in the vicinity of a solid crystalline interface. Studies by metallurgists[7] of the freezing of binary alloys have shown that for growth rates satisfying equation (5), the system reduces the local rate of freezing $\dot{m}$ by increasing the surface area over which freezing occurs, most commonly by forming dendrites extending into the fluid.

The process of freezing and dendritic growth has been described by Copley, *et al.*[12] who studied the evolution of a solution of ammonium chloride in water as it was cooled from below. This solution was studied because it is transparent, allowing the freezing process to be observed directly, and because it has a low entropy of fusion, thus solidifying, as would a metallic solution[13-15]. The cooling causes crystals of ammonium chloride to freeze onto a dendritic interface at the bottom of the solution. The interdendritic liquid is enriched in water and hence is lighter than the overlying solution. This unstable layering drives convective motions despite the stabilizing temperature gradient. The buoyant plumes of lighter fluid erode the dendrites and create isolated chimneys free of solid.

In a quasi-steady state, the cooling system studied by Copley *et al.*[12] may be divided into three regions: a lower region completely solid, an upper region completely liquid and a middle region, the 'mushy zone', in which both solid (dendrites)

* Present address: Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Silver Street, Cambridge CB3 9EW, UK.

and liquid are present. The compositionally driven convective motions circulate throughout the upper two regions. The top of the mushy zone is the level at which $T = T_L(p, \xi_o)$ where ξ_o is the mass fraction of the less dense constituent (water) in the well-mixed upper region, while the bottom is the level at which $T = T_L(p, \xi_c)$.

The mushy zone may be thought of as a layer which matches the composition ξ_o of the upper region to the lowest-melting-point composition ξ_c. The thickness of the zone thus depends on the difference $\xi_c - \xi_o$. In the laboratory the dependence of the liquidus temperature on pressure is unimportant and the imposed temperature gradient is sufficiently large that the mushy zone is only a few centimetres deep. In the Earth, pressure effects are important and the temperature gradient is much smaller, causing the mushy zone to be much thicker. To estimate its thickness, the liquidus relation may be rewritten as

$$g\rho\delta \, dr = -\tilde{\mu}\xi \, d\xi - (L/T) \, dT \tag{6}$$

Within the mushy zone, $dT/dr < 0$ and a lower bound on its thickness can be found by setting $dT = 0$ in equation (6). Integration yields

$$g_o\rho\delta(r_o^2 - r^2)/r_o > \tilde{\mu}(\xi^2 - \xi_o^2) \tag{7}$$

where $g = g_o r/r_o$, subscript o denotes conditions at the inner-core boundary (at the top of the mushy zone) and we have assumed $\tilde{\mu}$, ρ, δ, r_o and g_o are constant. We may use this inequality to calculate the greatest compositional difference which can be accommodated by a mushy inner core. Setting $r = 0$, $\xi = \xi_c$, $r_o = 1.2 \times 10^6$ m and using the parameter estimates given previously, equation (7) predicts that $\xi_c - \xi_o < 0.016$. This figure is, of course, very uncertain. If we had chosen a larger value of ξ_o, say 0.15, then the allowable difference would have been smaller: $\xi_c - \xi_o < 0.006$. An independent bound on the compositional difference sustained by a mushy inner core can be found using Fig. 5 of ref. 16. With the pressure difference across the outer core known[17] to be 1.93×10^{11} N m^{-2}, the liquidus gradient may be estimated as $(d\xi/dp)_T \leq 10^{-13}$ m^2 N^{-1}. With the pressure difference across the inner core being[17] 3.3×10^{10} N m^{-2}, it follows that $\xi_c - \xi_o \leq 0.003$.

Whether the composition of the core satisfies such an inequality is not known. The point of this calculation is to show that the thickness of the mushy zone is strongly dependent on the difference in composition between the outer core and the corresponding liquid with the lowest melting point. In fact, unless the two compositions are very nearly equal, the mushy zone will be very thick, possibly extending to the centre of the Earth.

When the idea of compositionally driven convection in the core was first introduced, it was assumed that the resulting flow patterns would be quite similar to those found in thermally driven convection. With the model of a dendritic inner-core boundary, that assumption is not valid. As the core gradually cools, compositionally buoyant fluid is produced within the inner core. This fluid rises upwards into the outer core as thin plumes and is replaced by a general downward motion of heavier fluid. An important feature of this flow pattern, resulting from the fact that diffusion of material is very slow, is that all streamlines must pass through the inner core. Models of fluid flow within the outer core, such as those constructed for kinematic dynamo problems, should have this property.

There may be some seismic evidence which corroborates the view that the inner core is in a partially molten state. Specifically, several authors[18-20] have suggested that the attentuation of seismic waves in the inner core may be due to the presence of fluid inclusions. It has also been demonstrated experimentally[21] that the anelasticity of a simple binary system ($NaCl$–H_2O) is much larger if partially molten than if completely solid. Several attempts[19,20] have been made to demonstrate that thermal diffusive effects in a partially molten inner core can account for the observed anelasticity, but preliminary calculations show that compositional effects are likely to dominate thermal effects; whether these can account for the observed anelasticity remains an open question.

This work was supported by NSF under grant EAR78-00818 and is contribution no. 173 of the Geophysical Fluid Dynamics Institute, Florida State University, Tallahassee, Florida.

Received 9 April; accepted 1 June 1981.

1. Jacobs, J. A. *Nature* **172,** 297 (1953).
2. Stacey, F. D. *Phys. Earth planet. Inter.* **22,** 89–96 (1980).
3. Gubbins, D. *J. Geophys.* **43,** 453–464 (1977).
4. Loper, D. E. *Geophys. J. R. astr. Soc.*, **54,** 389–404 (1978).
5. Fearn, D. R. & Loper, D. E. *Nature* **289,** 393–394 (1981).
6. Birch, F. *J. geophys. Res.* **57,** 227–286 (1952).
7. Chalmers, B. *Principles of Solidification* (Wiley, New York, 1964).
8. Fearn, D. R. & Loper, D. E. in *The Fluid Mechanics of Astrophysics and Geophysics* Vol. 2 (ed. Soward, A. M.) (Gordon and Breach, London, in the press).
9. Loper, D. E. & Roberts, P. H. *Geophys. Astrophys. Fluid Dyn.* **9,** 289–321 (1978).
10. Loper, D. E. & Roberts, P. H. *Geophys. Astrophys. Fluid Dyn.* **16,** 83–127 (1980).
11. Loper, D. E. & Roberts, P. H. *Phys. Earth planet. Inter.* **24,** 302–307 (1981).
12. Copley, S. M., Giamei, A. F., Johnson, S. M. & Hornbecker, M. F. *Met. Trans.* **1,** 2193–2204 (1970).
13. Jackson, K. A. & Hunt, J. D. *Acta. met.* **13,** 1212–1215 (1965).
14. Jackson, K. A., Hunt, J. D., Uhlmann, D. R. & Seward, T. P. *Trans. met. Soc. AIME* **236,** 149–158 (1966).
15. Streat, N. & Weinberg, F. *Met. Trans.* **3,** 3181–3184 (1972).
16. Usselman, T. M. *Am. J. Sci.* **275,** 291–303 (1975).
17. Jacobs, J. A. *The Earth's Core* (Academic, London, 1975).
18. Dourbos, D. J. *Geophys. J. R. astr. Soc.* **38,** 397–415 (1974).
19. Stiller, H., Franck, S. & Schmit, U. *Phys. Earth planet. Inter.* **22,** 221–225 (1980).
20. Cormier, V. F. *Phys. Earth planet Inter.* **24,** 291–301.
21. Spetzler, H. & Anderson, D. L. *J. geophys. Res.* **73,** 6051–6060 (1968).

Printed in Great Britain by Macmillan Production Ltd., Basingstoke, Hampshire

Physics of the Earth and Planetary Interiors, 36 (1984) 49–60
Elsevier Science Publishers B.V., Amsterdam – Printed in The Netherlands

Time-dependent electromagnetic core–mantle coupling *

M. Stix

Kiepenheuer-Institut für Sonnenphysik, D-7800 Freiburg (W. Germany)

P.H. Roberts

University of Newcastle, Newcastle upon Tyne, NE1 7RU (Gt. Britain)

(Received and accepted November 11, 1983)

Stix, M. and Roberts, P.H., 1984. Time-dependent electromagnetic core–mantle coupling. Phys. Earth Planet. Inter., 36: 49–60.

The magnetic field in the Earth's mantle is computed using a depth-dependent electrical conductivity, of form $\sigma = \sigma_a (r/a)^{-\alpha}$, and an approximation scheme in which the electromagnetic time constant of the mantle is assumed small compared with the time scales of the secular variation, and in which the induced currents and fields are obtained iteratively. We first associate the toroidal fields in the mantle with motions at the core surface ($r = a$) which create the observed geomagnetic field by flux rearrangement, and compute the resulting couple, Γ, parallel to the geographical axis. Using only zonal core motions, and values $\sigma_a = 3 \times 10^3\ \Omega^{-1}\ m^{-1}$, $\alpha = 30$ for the conductivity profile, we find that the toroidal induced fields create a couple, Γ_T, that over most of this century has been roughly ten times greater than the poloidal part, Γ_S, of Γ, and has the same sign. The total couple, Γ, has fluctuations of order 10^{18} Nm as required for the observed decade fluctuations in the length of the day. Its average is $\sim -1.5 \times 10^{18}$ Nm, i.e., it is too large to remain unbalanced. We suppose that an equally important couple in the opposite sense is created by flux leakage from the core, and we estimate the necessary gradient of toroidal field in the core to be of order -0.5 Gs km^{-1} at the core surface. During the course of the data analysis needed for the present work, we found some evidence for a torsional wave in the Earth's core with a period of ~ 60 y.

1. Introduction: conduction in the mantle

Fluctuations in the Earth's rate of rotation on the time scale of decades require unbalanced torques on the Earth's mantle of order 10^{18} Nm (see, for example, Morrison, 1979). In this paper we examine the hypothesis (Munk and Revelle, 1952) that these torques are created electromagnetically by the Lorentz force, $\mathbf{j} \times \mathbf{B}$, produced by the electric currents, density $\mathbf{j}$, created in the mantle by the core, and the geomagnetic field $\mathbf{B}$, both poloidal and toroidal. We therefore seek to evaluate the z-component of the total electromagnetic couple

$$\Gamma = \int_M \mathbf{r} \times (\mathbf{j} \times \mathbf{B}) \mathrm{d}V \qquad (1)$$

i.e., the component parallel to the geographic axis, $0z$. Here, M denotes the mantle, which is modelled by a spherical shell, $a \le r \le R$, with centre 0 and $\mathbf{r}$ is the radius vector from 0. The Γ_x and Γ_y components of Γ will alter the direction of the mantle's angular velocity vector $\boldsymbol{\Omega}$ in space. Here we shall set these zero, therefore writing Γ_z as Γ.

To evaluate Γ we need $\mathbf{j}$ and $\mathbf{B}$ in M. These are related by Ampère's law in pre-Maxwell form

$$\mu \mathbf{j} = \nabla \times \mathbf{B} \qquad (2)$$

(where μ, the permeability, will be taken to be that of free space), and by Ohms law

$$\mathbf{j} = \sigma \mathbf{E} \qquad (3)$$

where the electric field, $\mathbf{E}$, is measured in a frame

* Mitteilung aus dem Kiepenheuer-Institut No. 235.

0031-9201/84/$03.00

of reference fixed to the mantle, and σ is the electrical conductivity of the mantle, which is spatially varying and which we shall model by a function of r alone. When combined with Faraday's law

$$\nabla \times \mathbf{E} = -\dot{\mathbf{B}} \tag{4}$$

(2) and (3) lead to the familiar induction equation

$$\nabla \times (\eta \nabla \times \mathbf{B}) = -\dot{\mathbf{B}} \tag{5}$$

where $\eta(r) = 1/\mu\sigma$ is the magnetic diffusivity of the mantle.

In principle, we solve (5) subject to the appropriate boundary conditions, compute $\mathbf{j}$ from (2), and substitute into (1) to obtain Γ. In practice, this ambitious programme encounters two main difficulties. First, although we know $\mathbf{B}$ and $\dot{\mathbf{B}}$ in $r \geq R$ through the "observed" coefficients g_{nm}, h_{nm}, $\dot{g}_{nm}$, $\dot{h}_{nm}$ in the harmonic expansion

$$\mathbf{B} = -\nabla \Phi^{(0)} \tag{6}$$

$$\Phi^{(0)} = R \sum_{n=1}^{\infty} \sum_{m=0}^{n} (g_{nm} \cos m\phi + h_{nm} \sin m\phi) \times (R/r)^{n+1} \lambda_{nm} P_n^m(\theta) \tag{7}$$

these boundary conditions on solutions of (5) suffice at most to determine the poloidal part of $\mathbf{B}$ in M. And there is no reason to believe that the toroidal field in M will not be equally decisive in determining Γ. [We note here that λ_{nm} in (7) is a normalization constant for the associated Legendre function P_n^m. In the geophysically common Schmidt normalization, $\lambda_{nm} = 1$. We shall later use the Ferrer normalization, however, for which λ_{nm} is given by (40) below. In (7), θ is colatitude and ϕ is East longitude.]

Second, the conductivity law for the mantle is determined poorly although a few points seem to be generally agreed. For example (except for the oceanic layers) σ increases with depth from the Earth's surface, to such an extent that we may ignore all but the lowermost range, $a \leq r \leq b$, in the integral (1). The thickness, $b-a$, of this "conducting layer" we will take to be 2×10^3 km. It is also thought that, even within this layer, the conductivity decreases rapidly with increasing r. In what follows, we shall adopt the simple power law

$$\sigma(r) = \sigma_a (r/a)^{-\alpha}, \quad a < r < b \tag{8}$$

where α is a constant, which has been given values between 5.1 and 30 (McDonald, 1957; Kolomiytseva, 1972; Alldredge, 1977). We shall take below

$$\sigma_a = 3 \times 10^3\ \Omega^{-1}\ \mathrm{m}^{-1}, \alpha = 30 \tag{9}$$

The electromagnetic properties of a uniform solid body of size L are often summarized by its electromagnetic time-constant, $\tau = \mu\sigma L^2$. There are well known dangers in applying this concept too superficially. For example, $\mu\sigma L^2$ is derived supposing the scale l of the field to be of the same order as the scale L of the body, but the body will generally contain fields on many smaller scales that will evolve on correspondingly shorter times $\mu\sigma l^2$. In the spherical case, τ is a function of the harmonic n, and when a time-constant of 10^4 y is quoted for the Earth's core, what is meant is $\tau_{c1} = \mu\sigma_c a^2/\pi^2$, the time-constant for $n = 1$. In a similar way, we must expect the mantle to have an infinity of time-constants, τ_n (cf. Backus, 1983). It should also be recalled that, as usually defined, τ_n is associated with the natural decay of the field harmonic, n, and further complications may arise: there should be such a time-constant for both the poloidal and toroidal fields; the answers should depend on the conditions assumed on $\mathscr{S}$, the core–mantle boundary ($r = a$); and so forth. Moreover, we are not interested really in the natural decay of fields in M, but with the response of M to conditions imposed by the core on $\mathscr{S}$. That response will depend clearly on the frequency, ω, of the Fourier component, $\mathbf{B}_a \exp(i\omega t)$, of the field $\mathbf{B}$ on $r = a$. We may define a high frequency time-constant, τ_∞, and a low frequency time-constant, τ_0, corresponding to the $\omega \to \infty$ and $\omega \to 0$ limits of solutions to (5). The former is easy to obtain by solving (5) by WKBJ methods. These show that $\mathbf{B}_R \exp(i\omega t)$, the Fourier component of the field at $r = R$ generated by $\mathbf{B}_a$, is proportional to $\mathbf{B}_a \exp[(i\omega\tau_\infty)^{1/2}]$ where

$$\tau_\infty \equiv \mu \left[\int_a^b \sigma^{1/2} \mathrm{d}r \right]^2 \tag{10}$$

or, using (8) and $\alpha \gg 1$

$$\tau_\infty = 4\mu\sigma_a a^2/\alpha^2 \tag{11}$$

Evidently τ_∞ is independent of the harmonic number, n, of $\mathbf{B}_a$.

Evaluation of the low frequency time-constant, $\tau_{0,n}$, which does depend on n, affords us also the opportunity of introducing the principal solution technique that we will employ later. We solve (5) by iteration (see Roberts, 1972; Backus, 1982; Benton and Whaler, 1983)

$$\mathbf{B} = \mathbf{B}^{(0)} + \mathbf{B}^{(1)} + \ldots \tag{12}$$

where

$$\nabla \times (\eta \nabla \times \mathbf{B}^{(0)}) = 0 \tag{13}$$

$$\nabla \times (\eta \nabla \times \mathbf{B}^{(1)}) = -\dot{\mathbf{B}}^{(0)} \tag{14}$$

. . . . . .

Low conductivity expansions of this type are used in a number of laboratory contexts in magnetohydrodynamics (cf. e.g., Roberts, 1967, section 5.4). They are not uniformly valid in t since an arbitrary initial state will not obey (13). There will therefore be an initial period of time, of order τ (and corresponding to the more familiar "inner domain" of boundary layer theory), in which the full equation (5) is required to follow the evolution of $\mathbf{B}$. As t increases, this inner solution merges continuously with a solution of (12)–(14). This outer expansion is then valid for all $t \gtrsim \tau$. By (2), (3) and (13)

$$\mathbf{E}^{(0)} = -\nabla \psi^{(0)} \tag{15}$$

i.e., to leading order, the only electric field in the mantle is electrostatic and is set up by potential differences over $\mathscr{S}$ that create poloidal currents in M and therefore toroidal fields also. The poloidal field thus has no sources in M to this order, and (6) and (7) can be extrapolated through the entire mantle $r \geq a$. This is not true at the next order, at which (14) gives

$$\nabla^2 (r B_r^{(1)}) = i\omega\mu\sigma (r B_r^{(0)}) \tag{16}$$

Solving, subject to the boundary condition $B_{ra}^{(1)} = 0$ and the requirement that $B_{rb}^{(1)}$ matches to a potential field in $r > b$, gives in a straightforward way

$$B_{rR}^{(1)} = -i\omega\tau_{0,n} B_{ra}^{(0)} \tag{17}$$

where

$$\tau_{0,n} = \frac{\mu}{2n+1} \int_a^R \left[1 - \left(\frac{a}{r} \right)^{2n+1} \right] r\sigma(r)\,\mathrm{d}r \tag{18}$$

or, using (8)

$$\tau_{0,n} = \frac{\mu\sigma_a a^2}{2n+1} \left[\frac{1}{(\alpha-2)} - \frac{1}{(2n+\alpha-1)} - \frac{1}{(\alpha-2)} \left(\frac{a}{b} \right)^{\alpha-2} + \frac{1}{(2n+\alpha-1)} \left(\frac{a}{b} \right)^{2n+\alpha-1} \right] \tag{19}$$

The result in (18) appears first in Backus (1983) and was independently obtained by Benton and Whaler (1983). Backus (1983) also gave (19) for the special case of a power law conductivity. For large α, $\tau_{0,n}$ again becomes independent of n

$$\tau_0 = \mu\sigma_a a^2/\alpha^2 = \tau_\infty/4 \tag{20}$$

We will use this limit as a measure of the time-constant of the mantle. The time-constant of the mantle is often assessed, on the basis of secular variation studies, to be a few years. The numbers chosen in (9) above give 1.6 y.

Although, clearly, we require no knowledge of the dynamo processes at work deep within the core, in order to solve (5) and evaluate (1), it is true that Γ depends totally on conditions on $\mathscr{S}$ set up by the geodynamo (and possibly by potential differences of thermoelectric origin on $\mathscr{S}$), and it is useful to relate the conditions on $\mathscr{S}$ to the magnetohydrodynamic state of the top of the core (or more precisely below the thin Ekman-type boundary layer at the top of the core in which the fluid velocity, $\mathbf{u}$, of the core is brought to zero to obey the non-slip conditions on $\mathscr{S}$). From a typical size, U, of $\mathbf{u}$ we may construct a magnetic Reynolds number for the mantle, e.g.

$$R_m = Ua/\alpha\eta_a \tag{21}$$

and this is small. For example, a common estimate $U = 10^{-4}$ m s^{-1} with (9) gives $R_m \doteq 0.04$. It is true that U may be as much as 5 times larger, leading to $R_m \doteq 0.2$ which appears to make the success of the method less certain. But assuming, as we shall, that $R_m \ll 1$, $\mathbf{E}^{(0)}$ on $\mathscr{S}$ is small compared with $\eta_a \mathbf{B}/a$ and so, when we adopt the expansion (12), we may set $\psi^{(0)} = 0$, i.e., $\mathbf{E}^{(0)} \doteq 0$. The first term of (12) is therefore the potential field (6) and (7). To

leading order, then, $\Gamma = \Gamma^{(1)}$ where

$$\Gamma^{(1)} = \int_M \mathbf{r} \times (\mathbf{j}^{(1)} \times \mathbf{B}^{(0)}) dV \tag{22}$$

In evaluating $\Gamma^{(1)}$ it is useful to divide $\mathbf{B}$ into its toroidal and poloidal parts, writing in the usual way

$$\mathbf{B} = \mathbf{B}_T + \mathbf{B}_S \tag{23}$$

where

$$\text{(a) } \mathbf{B}_T = \nabla \times (T\hat{\mathbf{r}}), \text{ (b) } \mathbf{B}_S = \nabla \times \nabla \times (S\hat{\mathbf{r}}) \tag{24}$$

and $\hat{\mathbf{r}}$ is the radial unit vector. The corresponding electric current densities will be written

$$\mathbf{j}_T = \mu^{-1} \nabla \times \nabla \times (T\hat{\mathbf{r}}),$$
$$\mathbf{j}_S = -\mu^{-1} \nabla \times \left(\left[r \nabla^2 (S/r) \right] \hat{\mathbf{r}} \right) \tag{25}$$

and $\Gamma^{(1)}$ will be divided similarly with

$$\text{(a) } \Gamma_T = \int_M \mathbf{r} \times (\mathbf{j}_T^{(1)} \times \mathbf{B}^{(0)}) dV,$$

$$\text{(b) } \Gamma_S = \int_M \mathbf{r} \times (\mathbf{j}_S^{(1)} \times \mathbf{B}^{(0)}) dV \tag{26}$$

Since $\mathbf{B}_S^{(1)}$ is set up by toroidal-type currents $\mathbf{j}_S^{(1)}$ that can only come from $\mathbf{B}_S^{(0)}$ by electromagnetic induction, we can evaluate Γ_S directly from (6), and known values of g_{nm}, $\dot{g}_{nm}$, etc., without reference to conditions on $\mathscr{S}$. This is carried out in section 2 below. To calculate Γ_T we recall that the magnetic Reynolds number

$$R_{mc} = Ua/\eta_c \tag{27}$$

of the core is of order 50–500, and that, especially when we look at events on the decade time-scale, we should be able to assume, at least in a first approximation, that it is a perfect electrical conductor. Then, the electric field within it is

$$\mathbf{E}_c = -\mathbf{u} \times \mathbf{B}_c \tag{28}$$

Since $\hat{\mathbf{r}} \cdot \mathbf{B}$ and $\hat{\mathbf{r}} \times \mathbf{E}$ are continuous on $\mathscr{S}$, we then know that the electric field in the mantle must obey

$$\hat{\mathbf{r}} \times \mathbf{E} = -B_r \mathbf{u} \quad \text{on} \quad r = a \tag{29}$$

Equation 29 in conjunction with (4) is, of course, the basis of the well-known "frozen-field" representation of the secular variation, in which $\dot{\mathbf{B}}$ is *totally* produced by a rearrangement of flux from $\mathscr{S}$ through fluid motions below $\mathscr{S}$. These motions are also *totally* responsible for Γ_T: in approximation (28) (and with thermoelectric effects ignored), Γ_T arises from flux rearrangement and for no other reason.

By (2) and (3), we now see that solutions to (14) must obey

$$\hat{\mathbf{r}} \times (\nabla \times \mathbf{B}^{(1)}) = -B_r^{(0)} \mathbf{u}/\eta_a \quad \text{on } r = a \tag{30}$$

and that therefore both toroidal and poloidal $\mathbf{B}^{(1)}$ are created, the latter being responsible for Γ_S which we have already dealt with more directly above. In section 3 below we therefore use only the extract

$$L^2 \frac{\partial T^{(1)}}{\partial r} = -\frac{a^2}{\eta_a} \hat{\mathbf{r}} \cdot \nabla \times (B_r^{(0)} \mathbf{u}) \quad \text{on } r = a \tag{31}$$

of (30), where

$$L^2 F = -\left[\frac{1}{\sin\theta} \frac{\partial}{\partial \theta} \left(\sin\theta \frac{\partial F}{\partial \theta} \right) + \frac{1}{\sin^2\theta} \frac{\partial^2 F}{\partial \phi^2} \right] \tag{32}$$

Solving (5), subject to (31) and to $T_b^{(1)} = 0$, we obtain $\mathbf{B}_T$ and hence Γ_T.

Equation 28 is really only a first approximation to $\mathbf{E}_c$, a more complete treatment of Ohms law leading to

$$\mathbf{E}_c = -\mathbf{u} \times \mathbf{B}_c + \eta_c \nabla \times \mathbf{B}_c \tag{33}$$

In a second approximation, (30) is therefore replaced by

$$\hat{\mathbf{r}} \times (\nabla \times \mathbf{B}^{(1)}) = -B_r^{(0)} \mathbf{u}/\eta_a + (\eta_c/\eta_a) \hat{\mathbf{r}} \times (\nabla \times \mathbf{B}_c^{(0)}), \; r = a \tag{34}$$

There are two geophysical arguments suggesting that the last term in (34) may be influential. First, although Voorhies and Benton (1982) concluded from MAGSAT data that the frozen field model performs well, recently both Braginsky (1983) on analytic grounds and Gubbins (1983) by analysis of observations have suggested significant departures from frozen field behaviour. Gubbins confirmed the earlier conclusion of Booker (1969) that failure could even be detected at spherical harmonic level $n = 1$. The failure is evidently not dramatic and, even if an allowance were made for the diffusion of poloidal flux from the core, Γ_S and

Γ would not be much altered. But the toroidal field in the core may be large and, if expulsion of poloidal field is detectable, one must wonder whether the diffusion of toroidal field from the core through the last term of (34) and the concomitant alteration in Γ_T may not be important. This is the second of the geophysical points referred to above. When we attempt in section 3 to evaluate Γ_T on the basis of (30) we find that Γ_T is larger but of the same sign as Γ_S, and that these unbalanced torques would create an acceleration of the mantle large enough to be observed, but in fact absent. We surmise that, despite the smallness of (η_c/η_a), the last term in (34) may be significant. That this should be so is natural for the so called "strong field dynamo models", in which U is dominated by a zonal shear u_ϕ and in which R_{mc} is large. This shear creates a zonal field $B_\phi = 0(R_{mc}\mathbf{B}_p)$, large compared with $\mathbf{B}_p$, and proportional to $1/\eta_c$. The last term in (34) is then effectively independent of η_c and is not negligible.

We return to this question in section 4 and make only two further points here. First, when (34) is adopted it is no longer true that Γ is created by flux rearrangement alone. We now create Γ by flux diffusion, and for this reason we may call the Γ arising from the first and second sources in (34) "the couple due to flux rearrangement", and "the couple due to flux leakage", respectively. Second, since the former leads to unacceptable unbalanced torques we shall take the strong field dynamo model seriously, and shall therefore adopt a zonal shear

$$\mathbf{u} = u(\theta,t)\hat{\boldsymbol{\phi}} \tag{35}$$

as the dominant part of the core surface motion when applying (31) in section 3.

As indicated in this discussion, the plan of the paper is as follows: in section 2, the poloidal couple, Γ_S, is evaluated in the $R_m \ll 1$ approximation; in section 3, we expand $u(\theta,t)$ in a series of Legendre polynomials

$$u(\theta,t) = a \sin\theta \sum_{n=0}^{\infty} \zeta_n(t) P_n(\theta) \tag{36}$$

and try to determine the best values for the coefficients $\zeta_n(t)$ by fitting as well as possible the resulting $\bar{\mathbf{B}}$ to the observed secular variation. We then compute Γ_T. The results of section 3 depend on assuming that the core is perfectly conducting, but, observing that $\Gamma_S + \Gamma_T$ is too large to be unbalanced, we add in section 4 flux diffusion and, assuming a simple quadrupolar toroidal $\mathbf{B}_c^{(0)}$ as the source in (34), we make an estimate of the toroidal field gradient at the top of the core. In section 5 we briefly discuss our findings.

2. The poloidal couple

To evaluate the poloidal couple (26b) we need

$$\left(\mathbf{j}_S^{(1)} \times \mathbf{B}^{(0)}\right)_\phi = \frac{-1}{\mu \sin\theta} \frac{\partial}{\partial\phi} \nabla^2 \left(\frac{S}{r}\right) \frac{\partial \Phi^{(0)}}{\partial r} \tag{37}$$

Introducing, in analogy to (6) and (7), the potential of the secular variation

$$\dot{\Phi}^{(0)} = R \sum_{n=1}^{\infty} \sum_{m=0}^{n} (\dot{g}_{nm} \cos m\phi + \dot{h}_{nm} \sin m\phi) \times (R/r)^{n+1} \lambda_{nm} P_n^m(\theta) \tag{38}$$

or equivalently, the defining scalar $\dot{S}^{(0)}$ (with $\dot{\Phi}^{(0)} = -\partial \dot{S}^{(0)}/\partial r$), we find from eq. (14) that $\eta r \nabla^2 (S/r) = \dot{S}_0$. Hence, the poloidal couple is

$$\Gamma_S = -\int_M \sigma \frac{\partial \dot{S}^{(0)}}{\partial\phi} \frac{\partial \Phi^{(0)}}{\partial r} \, dV \tag{39}$$

The main field coefficients g_{nm} and h_{nm} for the period 1903.5 to 1975.5 were kindly given to us by D.R. Barraclough; the secular variation coefficients $\dot{g}_{nm}$ and $\dot{h}_{nm}$ were then determined in precisely the same way as described by Hodder (1981). As we use the Ferrer-normalized associated Legendre functions, the constants λ_{nm} are

$$\lambda_{nm} = \sqrt{\frac{2(n-m)!}{(1+\delta_{m0})(n+m)!}} \tag{40}$$

To smooth the coefficients g_{nm}, $\dot{g}_{nm}$... we have computed running means, with weights $\frac{1}{4}$, $\frac{1}{2}$, $\frac{1}{4}$ given at 3 consecutive epochs (which are in steps of 2 y). In this way we have eliminated much of the short term fluctuations, consistent with our assumed time-constant of a few years for the mantle. When $\sigma(r)$ is of the form (8), the integral (39) can be performed analytically. We obtain

$$\Gamma_S = 4\pi\sigma_a a^5 \sum_{n=1}^{\infty} \sum_{m=1}^{n} \frac{m(n+1)(R/a)^{2n+4}}{(2n+1)n(2n+\alpha-1)}$$

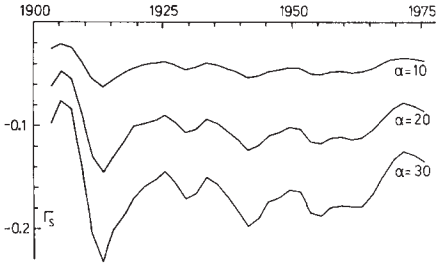


Fig. 1. The poloidal couple, Γ_S in units of 10^{18} Nm, for $\alpha = 10$, 20, 30 and $\sigma_a = 333$, 1333, 3000 $\Omega^{-1}\,m^{-1}$, respectively (so that $\sigma_a/\alpha^2 = \text{const}$). The proportionality of Γ_S to α is evident.

$$\times\left[1-\left(\frac{a}{b}\right)^{2n+\alpha-1}\right]\left(g_{nm}\dot{h}_{nm}-h_{nm}\dot{g}_{nm}\right) \quad (41)$$

It is interesting to note that, when the time-constant τ_0 is fixed, we have $\sigma_a \sim \alpha^2$; as observed by Braginsky and Fishman (1976), low order harmonics ($2n \ll \alpha$) therefore make contributions to the couple, proportional to α, while high order harmonics ($2n \gg \alpha$) make contributions proportional to α^2. Since $N = 5$ (before 1926) and $N = 6$ (after 1926) are the truncation levels for the coefficients available to us, we find that $\Gamma_S \propto \alpha$ to a good approximation. Figure 1 shows this behaviour. The toroidal couple to be calculated later depends on α in the same way. We shall therefore restrict our results below to the choice (9) of σ_a and α, knowing that we can readily deduce, if need be, values of Γ for other α.

3. The toroidal couple

In this section we shall consider the toroidal couple, Γ_T, due to flux rearrangement, i.e., the core will be treated as a perfect conductor (Roberts and Scott, 1965). As Γ_T depends on the motions of the core relative to the mantle, these motions must be obtained first.

Obviously $(\Phi,\dot{\Phi})$ determined by the analyses of the main field and secular variation will not be consistent with the frozen flux hypothesis. In this paper, we will consider that we have satisfied the frozen field constraint as well as possible within our class of chosen motions $\mathbf{u}$ if we have maximized the likelihood

$$\mathscr{L} = -\int\left[\hat{\mathbf{r}}\cdot\left(\dot{\mathbf{B}}^{(0)}-\nabla\times(\mathbf{u}\times\mathbf{B}^{(0)})\right)\right]^2 dS \quad (42)$$

where the integral is taken over the core surface, $\mathscr{S}$. Having found the best $\mathbf{u}$ we can compute a new $\dot{\Phi}$ and ask how close it is to the $\dot{\Phi}$ originally provided. We may also compute the toroidal fields generated and the resulting toroidal couple.

We restrict attention to axisymmetric zonal $\mathbf{u}$: on $\mathscr{S}$

$$u_r = 0, \quad u_\theta = 0, \quad u_\phi = \zeta(\theta)\, a \sin\theta \quad (43)$$

so that

$$\zeta = \sum_{l=0}^{A} \zeta_l P_l(\theta) \quad (44)$$

is the local angular velocity about 0z. Although Whaler (1982) has pointed out that meridional velocity components may not be negligible, we shall determine only a flow of form (43). This can be done uniquely, in contrast to the general case where there is a considerable degree of arbitrariness in $\mathbf{u}$ (Backus, 1968). Since $[\nabla\times(\mathbf{u}\times\mathbf{B}^{(0)})]_r = -\zeta\partial B_r^{(0)}/\partial\phi$, we evaluate

$$P_l\frac{\partial B_r}{\partial\phi} = \sum_{n=1+l}^{NB}\sum_{m=1}^{n}(n+1)m\left(\frac{R}{a}\right)^{n+2} \times\lambda_{nm}\left(H_{nm}^l\cos m\phi - G_{nm}^l\sin m\phi\right) \quad (45)$$

where $NB = N + l$

$$\left\{G_{nm}^l, H_{nm}^l\right\} = \frac{(2n+1)\lambda_{nm}}{4(n+1)}\sum_{j=n-l}^{JB}(j+1)\lambda_{jm} \times\left(\frac{R}{a}\right)^{j-n} G_{ljn}^{0mm}\left\{g_{jm}, h_{jm}\right\} \quad (46)$$

and

$$G_{ljn}^{0mm} = \int_0^\pi P_l P_j^m P_n^m \sin\theta\, d\theta \quad (47)$$

is a Gaunt integral; the sum (46) includes terms up to $JB = min(N, l+n)$. We are now in a position to write down $\mathscr{L}$ as a function of ζ_l. The conditions $\partial\mathscr{L}/\partial\zeta_l = 0$ then define the system of equations

$$\sum_{r=0}^{A} T_{lr}\zeta_r = U_l \quad (48)$$

where

$$T_{lr} = \sum_{n=1}^{NB} \sum_{m=1}^{n} \frac{(n+1)^2}{(2n+1)} \left(\frac{R}{a}\right)^{2n} m^2 \times \left(G^l_{nm} G^r_{nm} + H^l_{nm} H^r_{nm}\right) \quad (49)$$

and

$$U_l = \sum_{n=1}^{NB} \sum_{m=1}^{n} \frac{(n+1)^2}{(2n+1)} \left(\frac{R}{a}\right)^{2n} \times m\left(\dot{h}_{nm} G^l_{nm} - \dot{g}_{nm} H^l_{nm}\right) \quad (50)$$

In the case of a pure westward drift, i.e., only one term, ζ_0, in (44), the system (48)–(50) reduces to

$$\zeta_0 = \sum_{n=1}^{N} W_n \left[\sum_{m=1}^{n} m\left(g_{nm}\dot{h}_{nm} - \dot{g}_{nm} h_{nm}\right) \Big/ \sum_{m=1}^{n} m^2\left(g^2_{nm} + h^2_{nm}\right) \right] \quad (51)$$

with the weight

$$W_n = \frac{(n+1)^2}{2n+1} \left(\frac{R}{a}\right)^{2n} \sum_{m=1}^{n} m^2\left(g^2_{nm} + h^2_{nm}\right) \Big/ \sum_{n=1}^{N} \frac{(n+1)^2}{2n+1} \left(\frac{R}{a}\right)^{2n} \sum_{m=1}^{n} m^2\left(g^2_{nm} + h^2_{nm}\right) \quad (52)$$

Richmond (1969) listed these weighting factors in his table 1 for the epoch 1965. They are (0.014, 0.040, 0.133, 0.167, 0.165, 0.482) and do not appear to converge. Our weighting factors also do not seem to converge well, a problem indigenous to studies of the present type that use spherical harmonic coefficients obtained by direct least squares analysis, and which will later be associated with a marked dependence of Γ on the truncation index, N.

The first 3 motion harmonics of the expansion (44) are depicted as a function of time in Figs. 2–4. For each truncation index, N, of the field expansion, we have maximized $\mathscr{L}$ with $A + 1 = 1 \ldots N$ motion harmonics included in the expansion (44). The dependence on A is only weak: in particular the westward drift, $-\zeta_0$, changes little when more motion harmonics are retained. Nevertheless the higher motion harmonics do improve our fit of the secular variation. This can be seen, when the integral (42) is rewritten, using (44) and

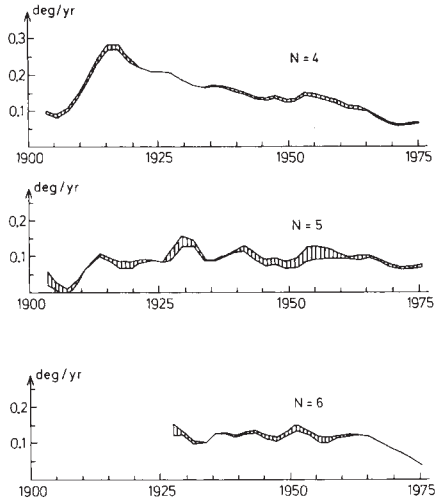


Fig. 2. The westward drift, $-\zeta_0$, as a function of the truncation index, N, of the field expansion. In each case the band shows the variation of $-\zeta_0$ as $1 \ldots N$ motion harmonics are retained in the expansion (44).

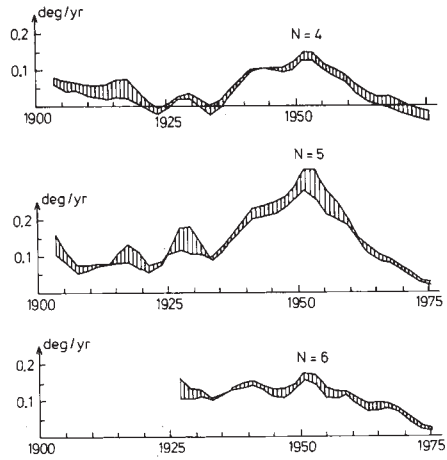


Fig. 3. The first antisymmetric velocity, ζ_1, of the expansion (44). The band shows the variation of ζ_1 as a function of the total number, $2 \ldots N$, of motion harmonics included.

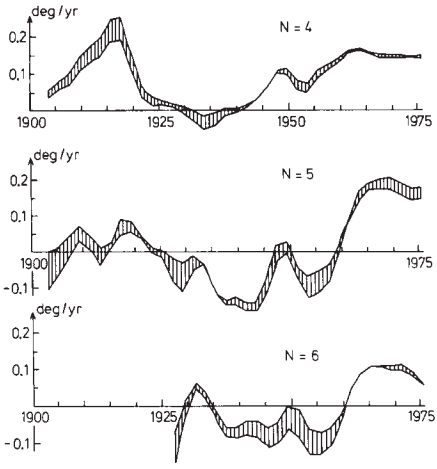


Fig. 4. The second symmetric velocity, ζ_2, of the expansion (44). Otherwise as Fig. 2. The band shows the variation of ζ_2 as a function of the total number, $3 \ldots N$, of motion harmonics included.

(48)–(50), as

$$\mathscr{L}' = -8\pi a^2 \left(\frac{R}{a}\right)^4 \left[\sum_{n=1}^{NB} \sum_{m=1}^{n} \frac{(n+1)^2}{(2n+1)} \left(\frac{R}{a}\right)^{2n} \times \left(\dot{g}_{nm}^2 + \dot{h}_{nm}^2\right) - \sum_{l=0}^{A} \zeta_l U_l \right] \tag{53}$$

and, accordingly, an "excellence of fit" parameter is defined by

$$E = \sum_{l=0}^{A} \zeta_l U_l \Bigg/ \sum_{n=1}^{NB} \sum_{m=1}^{n} \frac{(n+1)^2}{(2n+1)} \left(\frac{R}{a}\right)^{2n} \times \left(\dot{g}_{nm}^2 + \dot{h}_{nm}^2\right) \tag{54}$$

The dash in (53) indicates that we have omitted the $m = 0$ terms which are unaffected by ζ and are therefore irrelevant to, and not removable from the maximum $\mathscr{L}$. Figure 5 shows E as a function of the truncation indices N and A of field and motion, respectively. Adding the first antisymmetric motion, ζ_1, markedly improves the fit, and some further improvement is obtained by still higher harmonics. The best values of E are, however, around 0.5, indicating that a perfect fit, i.e., $E = 1$, is impossible with zonal motions alone. The fit deteriorates in particular after 1960, consistent with Whaler's (1982) result, obtained for 1965, that meridional motions should be included.

A result of particular interest (Fig. 4) is the strong dependence on time of the second symmetric zonal motion, ζ_2. We think that this is suggestive of torsional oscillations in the core, with a period of ~ 60 y, as has been advanced by Braginsky (1970), in the context of variations in the length of the day.

We can now calculate the toroidal couple. To this end we expand the scalar $T^{(1)}$ in expression (24a) into spherical harmonics, substitute into the toroidal part of eq. 14, and solve, subject to the boundary conditions $T^{(1)}(b) = 0$ and (31). The resulting toroidal mantle field is then used, in combination with the zero order field $\mathbf{B}^{(0)}$, to calculate the couple Γ_T according to (26a). This recipe has been described in more detail by Roberts (1972); we must only replace his q_n by the more general quantity

$$q_n = \frac{\beta_1 (a/b)^{\beta_1 - 1} - \beta_2 (a/b)^{\beta_2 - 1}}{(a/b)^{\beta_2 - 1} - (a/b)^{\beta_1 - 1}} \tag{55}$$

where

$$\beta_{1,2} = \tfrac{1}{2}(1-\alpha) \pm \tfrac{1}{2}\sqrt{(1-\alpha)^2 + 4n(n+1)} \tag{56}$$

Again making use of Gaunt integrals of the type (47) we finally obtain, with $M = min(n, j)$

$$\Gamma_T = \pi \sigma_a a^5 \sum_{n=1}^{N} \sum_{j=1}^{N} \sum_{m=0}^{M} \frac{j+1}{2n+1} \left(\frac{R}{a}\right)^{j+n+4} \times \sqrt{\frac{(j-m)!(n-m)!}{(j+m)!(n+m)!}} \left(g_{nm} g_{jm} + h_{nm} h_{jm}\right) \times \sum_{l=1}^{A+1} \left(\frac{\zeta_{l-1}}{2l-1} - \frac{\zeta_{l+1}}{2l+3} \right) \Bigg\{ \frac{(n+1)(n+m)}{n q_{n-1}} \times \left[j(j+1) - n(n-1) - l(l+1) \right] G^{0mm}_{lj,n-1} - \frac{(n+1-m)}{q_{n+1}} \left[j(j+1) - (n+1)(n+2) - l(l+1) \right] G^{0mm}_{lj,n+1} \Bigg\} \tag{57}$$

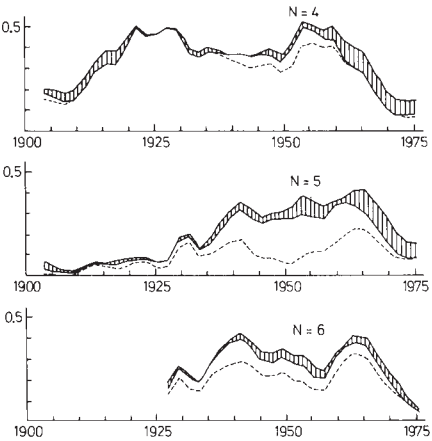


Fig. 5. The parameter E describing the quality of the fit of the secular variation by zonal motions in a perfectly conducting core. The broken curve is for a pure westward drift; the band is for $2 \ldots N$ motion harmonics.

Figure 6 shows the total couple, $\Gamma = \Gamma_S + \Gamma_T$. The main contribution to it is Γ_T, which exceeds Γ_S by about an order of magnitude. Comparison with

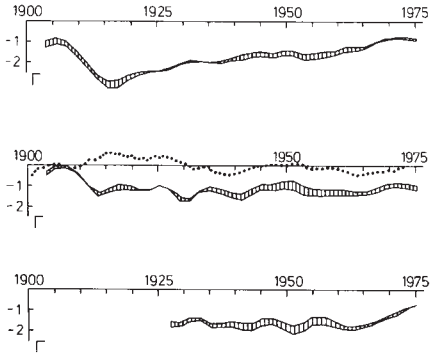


Fig. 6. The total couple, $\Gamma_S + \Gamma_T$, in units of 10^{18} Nm according to eqs. 41 and 57, as a function of the truncation index, N, of the field expansion. The main contribution is the toroidal part, Γ_T; the bands indicate its dependence on the total number, $A+1$, of motion harmonics. The dotted curve is the torque derived by Morrison (1979, Fig. 5) from fluctuations of the Earth's rate of rotation.

Fig. 2 immediately correctly suggests that the westward drift, $-\zeta_0$, i.e., the mean rotation of the core relative to the mantle, is almost exclusively responsible for the coupling. Although the higher motion harmonics are of the same order, 0.1–0.2° of longitude per year, they do not contribute much to the couple, probably because the torques obtained at different latitudes more or less cancel each other. In fact, the case $A = 0$, where only the mean core rotation is used to maximize $\mathscr{L}$, yields a couple, Γ_0, which is mostly at the lower (more negative) edge of the bands in Fig. 6; i.e., the couple arising from the inclusion of higher harmonics is (in absolute magnitude) somewhat smaller than Γ_0.

An interesting question is of course whether our total couple, Γ, can be related to the unbalanced torques on the Earth's mantle mentioned in our Introduction. Morrison's (1979) results, derived from fluctuations of the Earth's rate of rotation, are shown as a dotted curve in Fig. 6. Its variation is rather similar to our Γ-band for the case $N = 5$, somewhat less so for $N = 6$, and rather dissimilar from $N = 4$. Of course, a more detailed comparison would require a better knowledge of the electromagnetic time-constants of the mantle. Or the comparison could be used to infer information concerning these time-constants (e.g., Backus, 1983). In any case, our calculation shows that sufficiently large *variations* of Γ can arise easily when the coupling is electromagnetic in nature. This result was not obtainable from calculations for a single epoch (e.g., Rochester, 1960; Roden, 1963; Stix, 1982) alone.

4. The leakage couple

The couple which we have obtained from flux rearrangement by core motions was always negative, while Morrison's (1979) curve fluctuates around the comparatively small tidal couple of -5.10^{16} Nm. As explained in the Introduction we think that this difference is due to our neglect of diffusion in the core. Unfortunately, when (30) is replaced by (34), one can only proceed if the field in the core is known. To determine the latter a dynamo model is necessary in principle. However,

instead of specifying such a model we simply suppose here that the toroidal field in the core is of the form

$$\mathbf{B}_c = \nabla \times (T_c \hat{\mathbf{r}}) \quad T_c = T_{c2}(r) P_2(\theta) \tag{58}$$

which would naturally be the dominant harmonic in a "strong field dynamo" which produces such fields by zonal shearing motions. We substitute (58) into condition (34), repeat the calculations of the preceding section which led to the toroidal couple Γ_T, but obtain this time

$$\tilde{\Gamma}_T = \Gamma_T + \Gamma_L \tag{59}$$

where the additional term, Γ_L, is the couple due to "leakage"

$$\Gamma_L = \frac{16\pi a^3 \eta_c}{5\mu q_2 \eta_a} \left(\frac{R}{a}\right)^3 \left(g_{10} - \frac{6}{7}\left(\frac{R}{a}\right)^2 g_{30}\right) \times \left(\frac{dT_{c2}}{dr}\right)_{r=a} \tag{60}$$

We now require that the time average of the total couple, $\Gamma_S + \Gamma_T + \Gamma_L$, is zero so that there is no net acceleration of the Earth's rotation (except for the small tidal effect). This allows us to estimate $(dT_{c2}/dr)_{r=a}$, viz.

$$\left(\frac{dT_{c2}}{dr}\right)_{r=a} = \frac{-\frac{5\mu q_2 \eta_a}{16\pi a^3 \eta_c}\left(\frac{a}{R}\right)^3 \bar{\Gamma}}{\left[\overline{g_{10}} - \frac{6}{7}\left(\frac{R}{a}\right)^2 \overline{g_{30}}\right]} \tag{61}$$

where $\bar{\Gamma} = \bar{\Gamma}_T + \bar{\Gamma}_S$ is the average couple due to flux rearrangement. Since η_c/η_a is small, so is $B_\phi(a)$, and the field gradient is $\partial B_{c\phi}/\partial r = (3/a)\cos\theta \sin\theta \, dT_{c2}/dr$, or for $\theta = \pi/4$, $(3/2a)$ times the right-hand side of (61). Using the values σ_a and α specified by (9), and $\sigma_c = 5.10^5\ \Omega^{-1}\ m^{-1}$, we thus find

$$\left(\frac{\partial B_{c\phi}}{\partial r}\right)_{r=a,\theta=\pi/4} \doteq 3 \times 10^{-26}\, \bar{\Gamma}(T/m) \tag{62}$$

where $\bar{\Gamma}$ is in Nm. Figure 6 shows that $\bar{\Gamma}$ is typically -1.5×10^{18} Nm, so that we obtain a field gradient of -5×10^{-8} T m^{-1} or -0.5 G km^{-1}. This compares reasonably well with the estimate made by Benton and Muth (1979) who used a completely different method. It indicates that the toroidal field in the core is very large indeed, if only the gradient continues to be of the same order for a few hundred km! The sign of the core field is consistent with a picture proposed by Bullard (1949) and Bullard et al. (1950): positive in the Northern Hemisphere, negative in the Southern, and generated within the core by a large shear which is opposite to the shear between mantle and core. The induction equation suggests that the magnitude of the shear $\partial u_\phi/\partial r$ on $r = a$ should be of order $(\mu\sigma_c a B_P)^{-1}\partial B_T/\partial r$, i.e., $\sim 5 \times 10^{-11}$ s^{-1}, assuming the poloidal field B_P on $\mathscr{S}$ is of order 5 G. This shear is roughly what one would expect from the westward drift, the angular velocity of which is $\sim 10^{-10}$ s^{-1}, and from estimates of Benton (1981) for $\partial u_r/\partial r$.

We finally notice that, for large α, our estimate of the field gradient at the surface of the core does not separately depend on σ_a and α as long as the mantle time-constant is fixed, i.e., $\sigma_a \propto \alpha^2$. To see this we need only recall that $\bar{\Gamma} \propto \alpha, \eta_a \propto \alpha^{-2}$ and, according to (55), $q_2 \propto \alpha$.

5. Conclusions

If the core were a perfect conductor, both the changing field at the Earth's surface and the electric currents induced in the mantle would be determined completely by the rearrangement of the lines of force emerging from the core by the motions, **u**, on its surface, and the associated electromagnetic couple, Γ, on the mantle would be completely determined. To each flow **u** consistent with the observed geomagnetic field, a corresponding couple could be computed.

We have demonstrated how this programme may be carried through on the geophysically reasonable hypothesis that the electromagnetic time-constant of the mantle is short compared with the time scale over which the inducing, core generated, field is changing. In this case, Γ may be divided naturally into two parts, Γ_S and Γ_T, associated with poloidal and toroidal induced fields in the mantle, produced, respectively, by toroidal and poloidal induced currents. The couple Γ_S can be computed directly from the observed **B** and $\dot{\mathbf{B}}$, without reference to the core surface motions. This was done in section 2. The couple Γ_T is more

difficult to estimate, and it is necessary to make a hypothesis about **u**. We supposed in section 3 that **u** is dominated by a zonal latitude-dependent shearing of which the westward drift, ζ_0, is the most important factor determining Γ_T, the z-component of Γ_T. In fact, we wrote $\mathbf{u} = u_\phi \hat{\boldsymbol{\phi}}$ where

$$u_\phi(\theta) = \zeta(\theta) a \sin\theta, \quad \zeta = \sum_{l=0}^{A} \zeta_l P_l(\theta) \tag{63}$$

Using data supplied by Barraclough and the conductivity model (8) and (9), we found that during most of this century Γ_T is ~ 10 times larger than Γ_S and of the same sign. We also found that their sum, Γ, *fluctuated* by typically 10^{18} Nm over decade periods as required by observations of the length of the day (Morrison, 1979). However, the *average* value of Γ was also typically -1.5×10^{18} Nm, a value too large to be reconciled with the observations, which demonstrate that the mean unbalanced couple on the mantle is only of order -5×10^{16} Nm. There appears to be no way to save the model in the framework of flux rearrangement from a perfectly conducting core. It is true that the unbalanced couple of -1.5×10^{18} Nm could be removed by flux rearrangement: the angular velocity, ζ_0, of the core surface relative to the mantle would adjust itself by Lenz's law until Γ is brought to values of the order of -5×10^{16} Nm. The resulting westward drift, ζ_0, would, however, then be too small to be consistent with observation. Moreover, the *fluctuations* in Γ would then also be ~ 30 times too small. We conclude that another couple must be acting, and in section 4 we proposed that this is an electromagnetic couple produced by flux leakage from the core.

To obtain a sufficiently small mean electromagnetic couple, we found in section 4 that the radial gradient of toroidal field at the core surface, responsible for the flux diffusion and the balancing torque, must be of the order of 0.5 G km^{-1}. This, in fact, is what might be induced plausibly from the poloidal field of the Earth by a zonal shear of order 5×10^{-11} s^{-1}, i.e., one of the same (or smaller) order as the westward drift angular velocity relative to the mantle. (A core conductivity of order $5 \times 10^5\ \Omega^{-1}$ m^{-1} was assumed here.) The sense of this shear is opposite to that between core and mantle, as one would expect intuitively if its leakage is to off-set the couple associated with flux rearrangement by the westward drift. Indeed it is natural to reverse the picture as follows. The gradient of the toroidal field created by the dynamo in the core causes toroidal flux to leak into the mantle, creating a couple that accelerates the mantle eastwards relative to the core. As the relative motion (seen as westward drift!) increases, the drag of the poloidal field through the mantle (flux rearrangement!) grows as does the opposing couple (Lenz's law). Eventually a balance is reached, which is thereafter maintained, at least in the average as the geomagnetic field fluctuates.

Amongst the by-products of this investigation, we should remark that the inclusion of $l > 0$ terms in (63), i.e., core surface motions over and above the solid body rotation (ζ_0), considerably improves the agreement between the calculated and observed secular variation, without altering materially the couple Γ_T on the mantle. There seemed to be clear indications in ζ_2 of a long period oscillation which we identified tentatively with the 60 y torsional oscillation of the core proposed by Braginsky (1970). We look forward to the day when more comprehensive observations of the Earth's field, and methods of analysis that give convergent fields on the core surface (i.e., independent of truncation level, N, e.g., Shure et al. (1982), Gubbins (1983)) lead to firmer magnitudes for the flow fields and couples that are also independent of N.

Acknowledgement

We are grateful to David Barraclough for providing data prior to their publication.

References

Alldredge, L.R., 1977. Deep mantle conductivity. J. Geophys. Res., 82: 5427–5431.

Backus, G.E., 1968. Kinematics of geomagnetic secular variation in a perfectly conducting core. Philos. Trans. R. Soc. London, Ser. A: 263: 239–266.

Backus, G.E., 1982. The electric field produced in the mantle by the dynamo in the core. Phys. Earth Planet. Inter., 28: 191–214.

Backus, G.E., 1983. Application of mantle filter theory to the magnetic jerk of 1969. Geophys. J.R. Astron. Soc., 74: 713–746.

Benton, E.R., 1981. A simple method for determining the vertical growth rate of vertical motion at the top of the Earth's outer core. Phys. Earth Planet. Inter., 24: 242–244.

Benton, E.R. and Muth, L.A., 1979. On the strength of electric currents and zonal magnetic fields at the top of Earth's core: methodology and preliminary estimates. Phys. Earth Planet. Inter., 20: 127–133.

Benton, E.R. and Whaler, K.A., 1983. Rapid diffusion of the poloidal geomagnetic field through the weakly conducting mantle: a perturbation solution. Geophys. J.R. Astron. Soc., 75: 77–100.

Booker, J.R., 1969. Geomagnetic data and core motions. Proc. R. Soc. London, Ser. A: 309: 27–40.

Braginsky, S.I., 1970. Torsional magnetohydrodynamic vibrations in the Earth's core and variations in day length. Geomagn. Aeron., 10: 3–12 Eng. pp. 1–8.

Braginsky, S.I., 1983. On short geomagnetic secular variations. Geomagn. Aeron., 23: 106–112.

Braginsky, S.I. and Fishman, V.M., 1976. Electromagnetic interaction of the core and mantle when electrical conductivity is concentrated near the core boundary. Geomagn. Aeron., 16: 907–915, Eng. pp. 601–606.

Bullard, E.C., 1949. Electromagnetic induction in a rotating sphere. Proc. R. Soc. London, Ser. A: 199: 413–443.

Bullard, E.C., Freedman, C., Gellman, H. and Nixon, J., 1950. The westward drift of the Earth's magnetic field. Philos. Trans. R. Soc. London, Ser. A: 243: 67–92.

Gubbins, D., 1983. Geomagnetic field analysis I—Stochastic inversion. Geophys. J.R. Astron. Soc., 73: 641–652.

Hodder, B.M., 1981. Geomagnetic secular variation since 1901. Geophys. J.R. Astron. Soc., 65: 763–776.

Kolomiytseva, G.I., 1972. Distribution of electrical conductivity in the mantle of the Earth, according to data on secular geomagnetic field variations. Geomagn. Aeron., 12: 1082–1085, Eng. pp. 938–941.

McDonald, K.L., 1957. Penetration of the geomagnetic secular field through a mantle with variable conductivity. J. Geophys. Res., 62: 117–141.

Morrison, L.V., 1979. Re-determination of the decade fluctuations in the rotation of the Earth in the period 1861–1978. Geophys. J.R. Astron. Soc., 58: 349–360.

Munk, W. and Revelle, R., 1952. On the geophysical interpretation of irregularities in the rotation of the Earth. Mon. Not. R. Astron. Soc., Geophys. Suppl., 6: 331–347.

Richmond, A.D., 1969. Relation of the westward drift of the geomagnetic field to the rotation of the Earth's core. J. Geophys. Res., 74: 3013–3018.

Roberts, P.H., 1967. An introduction to magnetohydrodynamics. Longmans, Green and Co., London.

Roberts, P.H., 1972. Electromagnetic core–mantle coupling. J. Geomagn. Geoelectr., 24: 231–259.

Roberts, P.H. and Scott, S., 1965. On analysis of the secular variation, 1. A hydromagnetic constraint: theory. J. Geomagn. Geoelectr., 17: 137–151.

Rochester, M.G., 1960. Geomagnetic westward drift and irregularities in the Earth's rotation. Philos. Trans. R. Soc. London, Ser. A: 252: 531–555.

Roden, R.B., 1963. Electromagnetic core–mantle coupling. Geophys. J., 7: 361–373.

Shure, L., Parker, R.L. and Backus, G.E., 1982. Harmonic splines for geomagnetic modelling. Phys. Earth Planet. Inter., 28: 215–229.

Stix, M., 1982. On electromagnetic core–mantle coupling. Geophys. Astrophys. Fluid Dynam., 21: 303–313.

Voorhies, C.V. and Benton, E.R., 1982. Pole-strength of the Earth from MAGSAT and magnetic determination of the core radius. Geophys. Res. Lett., 9: 258–261.

Whaler, K.A., 1982. Geomagnetic secular variation and fluid motion at the core surface. Philos. Trans. R. Soc. London, Ser. A: 306: 235–246.

Geophys. Astrophys. Fluid Dynamics, Vol. 44, pp. 181–187
Reprints available directly from the publisher
Photocopying permitted by license only

ON TOPOGRAPHIC CORE-MANTLE COUPLING

PAUL H. ROBERTS

Institute of Geophysics and Planetary Physics, University of California, Los Angeles, CA 90024, USA

(*Received 5 January 1988; in final form 29 March 1988*)

The purpose of this note is two-fold: to draw attention to a perplexing difficulty connected with topographic core-mantle coupling, and to suggest tentatively an explanation. The difficulty is an apparent conflict between the most comprehensive theory of the coupling so far attempted (Anufriev and Braginsky, 1975a, b, 1977a, b) and recent explicit calculations based on magnetic and seismic information (Speith *et al.*, 1986). It is argued that asymmetric deviations from Anufriev and Braginsky's basically axisymmetric model of the underlying core flow are capable of resolving the difficulty.

KEY WORDS: Geophysical fluid dynamics, core-mantle coupling.

The search for an explanation of the decade fluctuations in the length of the day has recently focussed heavily on the role of the topographic torque, $\mathbf{\Gamma}$, exerted by the core on the mantle:

$$\mathbf{\Gamma} = \int_{CMB} p\mathbf{r} \times d\mathbf{A}_{CMB}. \tag{1}$$

Here p is the pressure, $\mathbf{r}$ is the radius vector from the geocenter, O, and $d\mathbf{A}_{CMB}$ is the vector element of surface area, directed along the normal to the core-mantle boundary ("CMB") from core to mantle. Were the CMB a sphere (radius c, say), $\mathbf{r}$ and the element of surface area ($d\mathbf{A}_0$, say) would be parallel, and the integral (1) would vanish. Topographic coupling relies on the fact that the CMB is non-spherical, with equation

$$r = c - h(\theta, \phi), \tag{2}$$

say, where θ is colatitude and ϕ is longitude. Since $h \ll c$, (1) may be written as

$$\mathbf{\Gamma} = c \int_{r=c} p\nabla h \times d\mathbf{A}_0 = -c \int_{r=c} h\nabla p \times d\mathbf{A}_0. \tag{3}$$

We shall focus on the component of $\mathbf{\Gamma}$ parallel to the angular velocity, $\mathbf{\Omega}$, of the Earth (Speith *et al.*, 1986):

$$\Gamma_z = -\int_{r=c} p(\partial h/\partial\phi)\, dA_0 = \int_{r=c} h(\partial p/\partial\phi)\, dA_0. \tag{4}$$

The magneto-geostrophic equation,

$$2\rho\boldsymbol{\Omega}\times\mathbf{v}=-\nabla p+\mathbf{J}\times\mathbf{B}+\rho\mathbf{g}, \tag{5}$$

can be used to link Γ to the core velocity $\mathbf{v}$. Here ρ is the density of the core (assumed constant), $\mathbf{g}$ is the acceleration due to gravity, $\mathbf{v}$ is the core velocity, $\mathbf{B}$ is the magnetic field, $\mathbf{J}=\mu^{-1}\nabla\times\mathbf{B}$ is the electric current density, and μ is the permeability; SI units are used. Equation (5) is obtained from the momentum equation by discarding the inertial terms (because they should be negligible for processes on the slow time scales of primary interest), and by omitting viscous friction, which is important only in a very thin Ekman layer at the surface of the core. Immediately beneath this boundary layer, $\mathbf{v}$ is almost horizontal, and will be written as $\mathbf{v}_s$, the so-called "core surface velocity"; this velocity can be inferred from the observed magnetic secular variation by supplementing the frozen-flux approximation by further hypotheses, e.g. that the equation of tangential geostrophy applies:

$$2\rho(\boldsymbol{\Omega}\times\mathbf{v}_s)_H=-\nabla_H p, \tag{6}$$

where the suffix H stands for "horizontal". Equation (6) follows from (5) on the plausible assumption that the electrical conductivity of the mantle is so small compared with the electrical conductivity (σ, say) of the core, and the magnetic field and electric currents on the CMB are therefore so weak compared with their counterparts deep in the core, that the Lorentz force, $\mathbf{J}\times\mathbf{B}$ is negligible on the CMB. This should be true even if the ratio of Lorentz to Coriolis forces deep in the core, i.e. the magnetic coupling parameter,

$$\mathscr{C}=B_T^2/2\Omega\rho\mu U_s c, \tag{7}$$

is of order unity, as is often thought to be the case. In (7), B_T is a typical sub-surface field strength, and U_s is a typical magnitude of $\mathbf{v}_s$.

When both magnetic and buoyancy forces are negligible, (5) reduces to the geostrophic equation, basic to the classic theory of rotating fluids,

$$2\rho\boldsymbol{\Omega}\times\mathbf{v}=-\nabla p, \tag{8}$$

which implies (the Proudman-Taylor theorem) that disturbances have a long-range influence in the direction of $\boldsymbol{\Omega}$. The most famous illustration of this is the Taylor column, carried along with a body towed slowly through a rotating tank. This long-range influence led Hide (1969) to propose that the effects of CMB topography on core flow would be large and that, in a first approximation,

$$c^{-1}\,\partial p/\partial\phi=O(2\Omega\rho U_s), \tag{9}$$

independently of h (assumed, however, large compared with the Ekman layer thickness). Using (9) to estimate the integral (4), we find

$$\Gamma_z = O(2\Omega\rho U_s c^3 h). \tag{10}$$

Hide recognized that (10) might be an overestimate. Since p is single-valued, $\partial p/\partial\phi$ must take both signs in the integrand of (4), so that some cancellations must be expected during integration (4). Hide estimated however that modest CMB-topography (less that 1 km) would create a torque of the required order of magnitude (i.e. about 1 Hadley unit $= 10^{18}$ Nm).

Hide's conclusions were not those of Anufriev and Braginsky (1975a, b; 1977a, b—in what follows these papers will be denoted by AB1–AB4, respectively; see also Moffatt, 1978). Anufriev and Braginsky based their discussion on the dominant axisymmetric terms,

$$p = \bar{p}(r,\theta), \qquad \mathbf{B} = \bar{B}_\phi(r,\theta)\hat{\boldsymbol{\phi}}, \qquad \mathbf{v} = \bar{v}_\phi(r,\theta)\hat{\boldsymbol{\phi}}, \tag{11}$$

of the nearly symmetric geodynamo model (Braginsky, 1964), in which (11) provides the leading terms in expansions,

$$p = \bar{p}(r,\theta) + p'(r,\theta,\phi) + \ldots, \tag{12}$$

$$\mathbf{B} = \bar{B}_\phi(r,\theta)\hat{\boldsymbol{\phi}} + \mathbf{B}'(r,\theta,\phi) + \ldots, \tag{13}$$

$$\mathbf{v} = \bar{v}_\phi(r,\theta)\hat{\boldsymbol{\phi}} + \mathbf{v}'(r,\theta,\phi) + \ldots, \tag{14}$$

in powers of $R_m^{-1/2}$, where

$$R_m = \mu\sigma U_s c \tag{15}$$

is the magnetic Reynolds number of the core, i.e.

$$p'/\bar{p} = O(R_m^{-1/2}), \tag{16}$$

and similarly for $\mathbf{B}$ and $\mathbf{v}$. In (11)–(14), $\hat{\boldsymbol{\phi}}$ is the unit vector in the direction of increasing ϕ, the overbars denote axisymmetric fields, and the primes denote purely asymmetric fields (i.e. those that lack ϕ-independent parts). The magnetic Reynolds number of the core is typically quoted as about 100, i.e.

$$R_m \gg 1, \tag{17}$$

so making (11) a plausible first approximation to (12)–(14) and $\bar{B}_\phi$ a reasonable estimate of B_T. It was with fields of the type (11), or in the case of AB1–3 planar analogues of those fields, that Anufriev and Braginsky modeled the core. Except in AB1, they took the toroidal field to be zero on the CMB, corresponding to an insulating mantle.

The effect of a (non-axisymmetric) bump on the CMB is to "deflect" the meridional pressure gradient, $\nabla\bar{p}$ of model (11) into the ϕ-direction, so creating by (4) a nonzero Γ_z. If $\mathscr{C} = O(1)$ however, the disturbance created by a bump is spread

out in the ϕ-direction by the "elasticity" of the toroidal field lines, and spread into the core no deeper than

$$D = O(\lambda \mathscr{C}^{-1/2}), \tag{18}$$

where λ is the lateral scale of the bump; see AB2. [We will not consider "small" bumps, defined as those whose magnetic Reynolds number $\mu\sigma U_s\lambda$ is $O(1)$ or less; for these, see AB3.] Not surprisingly, the effect of the pressure disturbances created by large bumps is no longer independent of h as in (10) but is proportional to h. In AB2 it is shown that, when $\mathscr{C} = O(1)$, both estimates (9) and (10) are reduced by a factor of order

$$f = (\lambda/c^2\mathscr{C})h, \tag{19}$$

which is plausibly less than 10^{-4}. [Take for instance, $h = 1\,\text{km}$, $\lambda = 10^3\,\text{km}$, $B_T = 0.03T(300\ \text{Gauss})$, and $U_s = 4.10^{-4}\,\text{m s}^{-1}$. Then $\mathscr{C} = 0.03$ and $f = 3.10^{-5}$.] Realizing that these results, derived from a planar model, are suspect for very large bumps, those for which the penetration depth (18) is of order c, Anufriev and Braginsky investigated spherical models in cases where $\mathscr{C}$ is of order $(\lambda/c)^2$ or smaller. Their analysis confirmed the relevance of the reduction factor f [see AB3, equation (29a)]. Consistent with its subsidiary role in the $R_m^{-1/2}$ expansions (12)–(14), the buoyancy force [$\rho\mathbf{g}$ in (5)] was not included by Anufriev in their analyses; the effects of mantle conduction on the topographic torque were also excluded.

That a drastic reduction in Γ_z below estimate (10) was a theoretical possibility was known prior to the work of Anufriev and Braginsky, but for a very different reason. When (8) holds, there is complete cancellation in (4):

$$\Gamma_z \equiv 0. \tag{20}$$

The relevance of this conclusion was contested by Hide (1970), who reported laboratory experiments (unpublished to this day), performed jointly with C. G. Collier, in which "boundary layer separation" occurred in the flow over a bump in a rotating tank. This destroyed the "fore-and-aft" symmetry of the pressure field with respect to the bump, so that the cancellation in the integrand of (4) became incomplete; the resulting Γ_z was once more of order (10). This conclusion was re-inforced by the numerical work of James (1980). It seemed that, at worst, (20) indicates only that (10) is to some extent a modest overestimate. Moreover, the presence of buoyancy and Lorentz forces in (5) will in any case prevent (20) from strictly applying.

The question of whether, and to what extent, (20) is significant has little to do with the objections of Anufriev and Braginsky. Experiments, whether they be

performed in the laboratory or on a computer, have little relevance to the magnetic environment of the upper core, unless they include the effects of Lorentz forces, which so far they have not done. Quite apart from the possible suppression of boundary layer separation be magnetic fields, a reduction of Γ_z arises because, as for a stratified fluid, the magnetic field, as it strengthens with depth into the core, increasingly opposes the penetration of the disturbance created by a bump. That disturbance is therefore spread out in the ϕ-direction and is confined to the uppermost layer of the core, of thickness D. The magnetic field curtails, by the large factor f, the $\partial p/\partial\phi$ created from $\bar{p}$ by the bump, and the questions of whether cancellation (20) is complete or not, and whether a bump does or does not shed eddies, becomes the question of whether very small quantities in (4) cancel or not. Irrespective of the answer, Γ_z is drastically reduced. The subsequent findings of Speith *et al.* seems all the more extraordinary.

Speith *et al.* (1986) obtained p from (6) as described above, and derived h from seismic tomography. They then evaluated Γ_z explicitly using (4). Clearly, this procedure is fraught with danger; the significant cancellations likely to occur were p and h precisely known (see above) are unlikely to arise when the p and h deduced from such long chains of reasoning are employed. Undeterred by the likelihood of significant error from this source, Speith *et al.* evaluated (4) and obtained a Γ_z of the order of one Hadley unit, but which was if anything too large rather than too small. Seemingly there is a paradox to be unravelled. The main purpose of this note is to suggest the following simple way of reconciling the apparently irreconcilable.

When Anufriev and Braginsky evaluate the torque (4), they are interested only in the pressure field that the bump itself creates from its interaction with the prevailing zonal flow. When Speith *et al.* evaluate torque (4), they are indifferent to the source of the latitudinal pressure variations; they make no attempt to relate p to h. They implicitly include the self-created pressure of the bump as a probably minor ingredient in a total pressure variation that is plausibly dominated by contributions from core magnetoconvection.

Anufriev and Braginsky had core convection in mind, but only in the limited sense of their basically axisymmetric model (11), in which to leading order the dominant unperturbed pressure, $\bar{p}$, is axisymmetric. Since $\partial\bar{p}/\partial\phi = 0$, the torque (4) is zero. It is only because, in their next approximation, the bump creates from $\bar{p}$ an asymmetric part of order $f\bar{p}$, that Γ_z is nonzero in their calculation. In reality, however, irrespective of CMB topography, the flow regime in the core has, like the pressure field with which it is associated, an asymmetric part not modeled by Anufriev and Braginsky; e.g. see (12). We may regard their analyses as establishing that this pressure field is not significantly affected by bumps, but when bumps are present it will in general, by (4), create from them a nonzero Γ_z. Within the framework of expansion (12)–(14), torque (4) and estimate (10) are reduced, as in (16), by a factor of order $R_m^{-1/2}$, i.e. only of order 0.1. This observation may provide a first tentative step towards resolving the apparent conflict between the two approaches.

Some vestiges of the Anufriev–Braginsky difficulty remain, however. Apart from the secular deceleration of the earth's rotation produced by the tides, the average

angular acceleration of the mantle is zero. This may be understood in the following way. Since the total momentum of the core-mantle system is constant, when a change in the magnetohydrodynamic state of the core excites a topographic torque Γ_z on the mantle, it also creates an equal but opposite torque on the core. The resulting growing change in the relative motion between the mantle and the surface of the core is associated with a restoring torque (Lenz's law) that grows until eventually it completely cancels out the exciting torque, Γ_z, so returning the net torque to zero. This restoring torque is associated with a change in the zonal part of $\mathbf{v}_s$, and this is *precisely* the type of flow that creates a tiny topographic torque according to Anufriev and Braginsky. This suggests that topographic restoring torques are too weak to be influential, i.e. that the restoring torques are non-topographic, and most likely they are magnetic in origin.

Our speculation that the pressure field exciting a topographic torque is associated with core convection has a further implication. Convection patterns necessarily drift in longitude, and as they do so the torque they produce by interaction with the (fixed) mantle topography will change secularly from one sign to the other and back. The restoring torque will necessarily do the same, in order that the net torque preserves its zero average. As indicated above, the zonal differential motion between mantle and core surface would therefore repeatedly reverse on the same slow timescale. This behavior is in sharp contrast to consequences of an alternative theory of core-mantle coupling, namely that the torques are electromagnetic in origin. The exciting torque is, on that view, the result of leakage of predominantly axisymmetric meridional currents into the mantle from the core, where they have been generated by the geodynamo. [See, for example, Stix and Roberts (1984), Section 4.] It is therefore essentially of one sign (even for both polarity states of the geomagnetic field), and this implies that the relative zonal motion (westward drift?) of mantle and core surface also does not change as the convection pattern drifts secularly.

Acknowledgements

Thanks are due to Stanislav Braginsky for criticising this Note in draft. I am grateful to Curt Voorhies for informing me more fully of the details of Speith *et al.* (1986) prior to its full publication. The contents of this note formed part of a presentation made by the author during the Sessions on "The Physics and Chemistry of the Core-Mantle Boundary" held at the 1987 Fall Meeting of the American Geophysical Union; see Roberts (1987). This work was supported by the National Science Foundation under grant EAR 86-12937.

References

Anufriev, A. P. and Braginsky, S. I., "Effect of a magnetic field on the stream of liquid rotating at a rough surface," *Magnetohydrodynamics* **4**, 461–467 (1975a). (This is referred to in the text as "AB1".)

Anufriev, A. P. and Braginsky, S. I., "Influence of irregularities of the boundary of the earth's core on the velocity of the liquid and on the magnetic field," *Geomagn. & Aeron.* **15**, 754–757 (1975b). (This is referred to in the text as "AB2".)

Anufriev, A. P. and Braginsky, S. I., "Influence of irregularities of the boundary of the earth's core on fluid velocity and the magnetic field. II," *Geomagn. & Aeron.* **17,** 78–82 (1977a). (This is referred to in the text as "AB3".)

Anufriev, A. P. and Braginsky, S. I., "Effect of irregularities of the boundary of the earth's core on the speed of the fluid and on the magnetic field. III," *Geomagn & Aeron.* **17,** 492–496 (1977b). (This is referred to in the text as "AB4".)

Braginsky, S. I., "Self-excitation of a magnetic field during the motion of a highly conducting fluid," *Sov. Phys. JETP* **20,** 726–735 (1964).

Hide, R., "Interaction between the earth's liquid core and solid mantle," *Nature* **222,** 1055–1056 (1969).

Hide, R., "On the earth's core-mantle interface," *Q. J. R. Meteorol. Soc.* **96,** 579–590 (1970).

James, I. N., "The forces due to topographic flow over shallow topography," *Geophys. and Astrophys. Fluid Dyn.* **14,** 225–250 (1980).

Moffatt, H. K., "Topographic coupling at the core-mantle interface," *Geophys. & Astrophys. Fluid Dyn.* **9,** 279–288 (1978).

Roberts, P. H., "The torque exerted by the core on the mantle," *EOS* **68** (44), 1481–1482 (1987).

Speith. M. A., Hide, R., Clayton, R. W., Hager, B. H. and Voorhies, C. V., "Topographic coupling of core and mantle, and changes in length of day," *EOS* **67** (44), 908 (1986).

Stix, M. and Roberts, P. H., "Time-dependent electromagnetic core-mantle coupling," *Phys. Earth and Planet. Inter.* **36,** 49–60 (1984).

ELSEVIER

Physics of the Earth and Planetary Interiors 91 (1995) 63–75

PHYSICS OF THE EARTH AND PLANETARY INTERIORS

A three-dimensional convective dynamo solution with rotating and finitely conducting inner core and mantle

Gary A. Glatzmaier [a,*], Paul H. Roberts [b]

[a] *Institute of Geophysics and Planetary Physics, Los Alamos National Laboratory, Los Alamos, NM 87545, USA*
[b] *Institute of Geophysics and Planetary Physics, University of California, Los Angeles, CA 90024, USA*

Received 10 November 1994; accepted 3 March 1995

Abstract

We present the first three-dimensional (3D), time-dependent, self-consistent numerical solution of the magnetohydrodynamic (MHD) equations that describe thermal convection and magnetic field generation in a rapidly rotating spherical fluid shell with a solid conducting inner core. This solution, which serves as a crude analog for the geodynamo, is a self-sustaining supercritical dynamo that has maintained a magnetic field for three magnetic diffusion times, roughly 40 000 years. Fluid velocity in the outer core reaches a maximum of 0.4 cm s^{-1}, and at times the magnetic field can be as large as 560 gauss. Magnetic energy is usually about 4000 times greater than the kinetic energy of the convection that maintains it. Viscous and magnetic coupling to both the inner core below and the mantle above cause time-dependent variations in their respective rotation rates; the inner core usually rotates faster than the mantle and decadal variations in the length of the day of the mantle are similar to those observed for the Earth. The pattern and amplitude of the radial magnetic field at the core–mantle boundary (CMB) and its secular variation are also similar to the Earth's. The maximum amplitudes of the longitudinally averaged temperature gradient, shear flow, helicity, and magnetic field oscillate between the northern and southern hemispheres on a time scale of a few thousand years. However, only once in many attempts does the field succeed in reversing its polarity because the field in the inner core, which has the opposite polarity to the field in most of the outer core, usually does not have enough time to reverse before the field in the outer core changes again. One successful magnetic field reversal occurs near the end of our simulation.

1. Introduction

A fundamental goal of geophysics is a coherent understanding of the geodynamo, the mechanism that generates the Earth's magnetic field in the liquid, electrically conducting, outer core. As described by Parker (1955), east–west fluid flow shears poloidal (meridional) magnetic field generating new toroidal (east–west) field while helical fluid flow twists toroidal magnetic field generating new poloidal field. However, a more detailed self-consistent description of the magnetohydrodynamics of the Earth's core is required to answer many fundamental questions about the structure and evolution of the geomagnetic field.

Several attempts have been made to develop self-consistent convective dynamo models. Fearn and Proctor (1987) developed a dynamo model

* Corresponding author.

SSDI 0031-9201(95)03049-2

that neglected the inertial and viscous forces in the momentum equation (because they are expected to be small in the Earth's liquid core) and used Taylor's constraint (Taylor, 1963) to determine the geostrophic flow; however, no converged solutions were found with this method. Zhang and Busse (1989, 1990) studied convective dynamos near the onset of convection using a severely truncated model that is constrained to be symmetric with respect to the equator and either two or four times periodic in longitude (opposed to fully three dimensional) and periodic in time (opposed to fully time dependent). However, as those workers pointed out about their convergence checks, their 'solutions obtained for $N_T = 4$ differ significantly in their quantitative aspects from those obtained for $N_T = 3$' (Zhang and Busse, 1989) and they 'tend to diverge when the Rayleigh number is increased too far from its critical value or when high Taylor numbers T are approached' (Zhang and Busse, 1990).

There have been a few successful attempts numerically to simulate convective dynamos with self-consistent, fully nonlinear, time-dependent, 3D models. Convective dynamos have been simulated in planar (Cartesian) geometry to study local dynamo action by Meneguzzi and Pouquet (1989), Brandenburg et al. (1990), Nordlund et al. (1992) and St. Pierre (1993). Full global geometry was included in the original 3D spherical shell models of the solar dynamo by Gilman and Miller (1981), Gilman (1983), and Glatzmaier (1984, 1985a,b) and in a recent model by Kageyama et al. (1995). However, of these, only the recent models by St. Pierre and Kageyama et al. generate magnetic energy an order of magnitude greater than kinetic energy and therefore begin to approach the strong-field regime appropriate for studying the geodynamo. In the other models, the generated magnetic energy was several orders of magnitude less than the kinetic energy.

In this paper we present the first self-consistent 3D numerical simulation of a convective dynamo in spherical geometry that generates a magnetic field with energy at least three orders of magnitude greater than the kinetic energy of the convection. Although our simulation is still far from a realistic simulation of the geodynamo, the parameters, except fluid viscosity, are appropriate for the Earth and we begin to make meaningful comparisons with observations and measurements.

2. The model

The anelastic magnetohydrodynamic (MHD) model and numerical method have been described in detail elsewhere Glatzmaier, (1995). The preliminary solution presented here was obtained using the same numerical method but a simpler set of equations for which the reference state density $\bar{\rho}$ and temperature $\bar{T}$ and the viscous $\bar{\nu}$, thermal $\bar{\kappa}$ and magnetic $\bar{\eta}$ diffusivities are assumed constant, the density perturbation is proportional to only the temperature perturbation (using a constant coefficient of volume expansion $\bar{\alpha}$), and viscous and Joule heating are neglected, i.e. using the Boussinesq approximation. For the present solution, only thermal convection is modeled and it therefore approximates the combined effects of compositional and thermal convection. In addition, the effects on the outer core of luni-solar precession of the mantle and the effects of core–mantle boundary (CMB) topography were not represented in this simulation.

The time-dependent (unbarred) perturbation variables are functions of the spherical coordinates (r, θ, ϕ) in a frame of reference rotating with angular velocity $\boldsymbol{\Omega} = \Omega \hat{\boldsymbol{z}}$. We scale the temperature T and pressure p perturbations (relative to the reference state) and the fluid velocity $\boldsymbol{v}$ and magnetic field $\boldsymbol{B}$ by $Q/(c_p \bar{\rho} \bar{\kappa} d)$, $2\Omega \bar{\rho} \bar{\kappa}$, $\bar{\kappa}/d$, and $(2\Omega \mu_0 \bar{\rho} \bar{\eta})^{1/2}$, respectively, where Q is the specified outward heat flow at the inner core boundary (ICB), c_p is the specific heat capacity of the core fluid, d is the depth of the outer core ($R_{CMB} - R_{ICB}$), and μ_0 is the (constant) magnetic permeability of both the liquid outer core and solid inner core ($\mu_0 = 4\pi$ when $\boldsymbol{B}$ is in gauss and c.g.s. units are used). We solve the following scaled Boussinesq equations for the outer fluid core:

$$\nabla \cdot \boldsymbol{v} = 0 \tag{1}$$

$$\nabla \cdot \boldsymbol{B} = 0 \tag{2}$$

$$0 = -\nabla p + \mathrm{Ra}\, g' T \hat{\boldsymbol{r}} + \boldsymbol{v} \times \hat{\boldsymbol{z}} + \mathrm{E}\nabla^2 \boldsymbol{v} + P(\nabla \times \boldsymbol{B}) \times \boldsymbol{B} \tag{3}$$

$$\frac{\partial \boldsymbol{B}}{\partial t} = \nabla \times (\boldsymbol{v} \times \boldsymbol{B}) + \boldsymbol{P}\nabla^2 \boldsymbol{B} \tag{4}$$

$$\frac{\partial T}{\partial t} = -\boldsymbol{v} \cdot \nabla T + \nabla^2 T \tag{5}$$

Here $g' = \bar{g}/g_0$, the gravitational acceleration scaled by its value at the CMB, and is approximately proportional to radius. (It is not exactly proportional to radius in the liquid core because we assume the density of the solid core is 16% greater than that of the liquid core.) The dimensionless parameters are the modified Rayleigh number,

$$\mathrm{Ra} = \frac{g_0 \bar{\alpha} Q}{2\Omega c_p \bar{\rho} \bar{\kappa}^2}$$

the Ekman number,

$$\mathrm{E} = \frac{\bar{\nu}}{2\Omega d^2}$$

and the ratio of the diffusivities,

$$P = \frac{\bar{\eta}}{\bar{\kappa}}$$

For the inner solid core, we solve only Eqs. (2) and (4) with the (time-dependent) solid body rotation of the inner core determined at the ICB.

The inertial terms in the momentum equation, (3), were dropped because their expected contribution was small compared with the other forces in the momentum equation. The advantage of this approximation is that it eliminates the Courant condition on the numerical timestep owing to the Alfvén velocity (St. Pierre, 1992), which would require a much smaller timestep than that determined by just the usual Courant condition owing to the convective velocity. This is a significant advantage, as the radial magnetic field, unlike the radial velocity, does not go to zero at the boundaries where the radial grid is the finest. However, we find that our timestep still needs to be roughly ten times smaller than what the usual Courant condition would dictate during most of the simulation to maintain numerical stability. The reason for this, as pointed out by Hollerbach and Ierley (1991), is that the model equations are very stiff. That is, the magnetic field evolves in a way that usually maintains a small magnetic torque on the boundaries. However, once in a while when the magnetic torque does become large, a large zonal flow is temporarily needed to produce a correspondingly large viscous torque; and at this time a small timestep is required.

Hollerbach and Jones (1993b, 1995) have demonstrated with their 2D axisymmetric intermediate dynamo simulations that magnetic coupling to a solid conducting inner core stabilizes the magnetic field generated in the outer core. Our model provides for magnetic coupling to the inner core below and to the mantle above. Magnetic coupling to the inner core is achieved by solving for the time-dependent magnetic field that diffuses into the inner metallic core. Boundary conditions force the three components of the magnetic field and the horizontal components of the electric field to be continuous at the ICB. We treat the mantle differently because most of the mantle is a very poor electrical conductor except for a shallow layer just above the CMB. Magnetic coupling to this lower-mantle region is implemented by using a thin conducting shell approximation and assuming a perfect insulator exterior to the thin shell. This results in a boundary condition on the magnetic field at the CMB involving the radial integral of the electrical conductivity in the lower mantle.

Although viscous forces are probably relatively small in the outer core they have been retained in the momentum equation, (3), to damp small spatial scales and, as we have neglected the inertial terms, to provide simple torques to balance any net magnetic torques that develop at the ICB and CMB.

The boundary conditions on the velocity at the ICB and CMB are impermeable, no-slip conditions. This results in viscous coupling to the inner core and mantle. The solid inner core below and the solid mantle above are also allowed to rotate about any axis through the center so that the total angular momentum of the mantle, outer core and inner core remains zero in the rotating frame of reference. As we neglect the inertial terms for the outer core, we chose to neglect

them also for the inner core and mantle. Consequently, the sum of the magnetic and viscous torques parallel to the rotation axis of the reference frame vanishes on the ICB and CMB (Gubbins, 1981; Hollerbach and Jones, 1993a), and the sums of the small off-axis components of these torques at the ICB and CMB balance the Coriolis torques owing to the moments of inertia of the inner core and mantle, respectively. This allows the model to simulate changes in the length of the day owing to core dynamics.

An isothermal boundary condition is placed on the temperature at the CMB, and a spherically symmetric radial heat flux boundary condition is applied at the ICB. These provide the superadiabatic temperature drop across the outer core that drives the convection. There is no prescribed internal heating. The prescribed heat flux at the ICB is intended to mimic the precipitation of light (compositionally buoyant) material that would result from iron plating onto the solid inner core (Braginsky, 1963).

As described by Glatzmaier (1995), the three thermodynamic variables, three components of mass flux, and three components of magnetic field are expanded in spherical harmonics to resolve their horizontal structures and in Chebyshev polynomials to resolve their radial structures. A full Chebyshev space is employed in the outer core and a half Chebyshev space in the inner core for the magnetic field. A poloidal–toroidal representation automatically satisfies Eqs. (1) and (2) for the divergence-free mass flux and magnetic field. A spectral transform method allows the nonlinear terms to be calculated in physical space whereas all spatial derivatives and time-integrations are performed in spectral space. The time-integration treats the nonlinear terms explicitly with an Adams–Bashforth method and the linear terms implicitly with a Crank–Nicolson method. The system of equations is solved for each timestep with a Chebyshev collocation method in the following order. After updating the magnetic field (4) and temperature (5), the new Lorentz force and buoyancy force are computed and used to balance the Coriolis, pressure gradient and viscous forces in (3) in the process of solving for the velocity and pressure. As all degrees, l, of the spherical harmonic coefficients of velocity and pressure (for a given spherical harmonic order, m) are coupled, they are solved simultaneously in a large block banded matrix. The major difference between this model (Glatzmaier, 1995) and that of the earlier solar dynamo model (Glatzmaier, 1984) is that in the earlier model the Coriolis terms are treated explicitly so each spherical harmonic component of the equations can be solved independently. Therefore, instead of solving a large block banded matrix for each m, the small block matrices for each l and m can be solved separately, which requires much less memory and cpu time for a given numerical resolution. The new method is preferred when one wishes to study problems such as the geodynamo for which the Coriolis and Lorentz forces are the two dominant forces and the viscous force is relatively small. That is, for numerical accuracy and stability considerations, it is better to solve for the velocity implicitly through the Coriolis force instead of solving for it only through the viscous force based on the small explicit difference between the two dominant forces.

Our model has been validated by comparing several of our numerical solutions with fluid dynamic experiments and with solutions obtained by others. Numerical solutions from a modified version of our convection model have produced good qualitative agreement with the results of a rotating, hemispherical, convection experiment flown on SpaceLab 3 aboard a Space Shuttle in 1985 (Hart, et al., 1986). Another version of our model produced solutions in agreement with a centrifuge convection experiment conducted at Johns Hopkins University (Cardin and Olson, 1992; Glatzmaier and Olson, 1993). Our benchmark solutions of 3D thermal convection showed excellent agreement with those obtained by another model (Bercovici et al., 1989). An axisymmetric (alpha–omega intermediate) version of our dynamo model, with a finitely conducting and rotating inner core, produced a benchmark solution in extremely good quantitative agreement (R. Hollerbach, personal communication, 1994) with the periodic intermediate dynamo solution that Hollerbach and Jones (1993b, 1995) obtained

using a different numerical method (Hollerbach, 1994).

The problem we have simulated is defined by Eqs. (1)–(5) and one choice of the key dimensionless parameters: $R_{ICB}/R_{CMB} = 0.35$; $Ra = 5.7 \times 10^7$; $E = 2.0 \times 10^{-6}$; $P = 10$. In addition, to maximize magnetic coupling to the lower mantle, we assume a layer $0.04d$ thick above the CMB that has the same electrical conductivity as the core. To relate our simulation to the Earth, we choose the following geophysically realistic dimensional parameters that satisfy the above values of Ra, E, and P: $\bar{\alpha} = 10^{-5}$ K^{-1}, $c_p = 6.7 \times 10^6$ erg (g K)$^{-1}$, $d = 2.26 \times 10^8$ cm, $g_0 = 1.1 \times 10^3$ cm s^{-2}, $\bar{\rho} = 11$ g cm^{-3}, $\Omega = 7.27 \times 10^{-5}$ s^{-1}, $\bar{\eta} = 3 \times 10^4$ cm^2 s^{-1}, $Q = 5 \times 10^{20}$ erg s^{-1}, $\bar{\kappa} = 3 \times 10^3$ cm^2 s^{-1}. Here $\bar{\kappa}$ represents a sub-grid scale eddy diffusivity and Q represents a combined thermal and compositional buoyancy source. The magnetic diffusion time, normally defined as $\tau_\eta = R_{CMB}{}^2/(\pi^2\bar{\eta})$, becomes 1.3×10^4 years. However, with the above choice of dimensional parameters, the Ekman number of 2×10^{-6} implies a viscosity of $\bar{\nu} = 1.5 \times 10^7$ cm^2 s^{-1}, making the Prandtl number $P_r = \bar{\nu}/\bar{\kappa} = 5000$ and the magnetic Prandtl number $P_m = \bar{\nu}/\bar{\eta} = 500$, which is geophysically unrealistic. The commonly accepted value of the molecular viscous diffusivity, $\bar{\nu} = 10^{-2}$ cm^2 s^{-1} (Gans, 1972), implies $E \approx 10^{-15}$, which is beyond computational reach if one wishes to resolve the Ekman boundary layer. A reasonable eddy viscous diffusivity would be comparable with the eddy thermal diffusivity, 3×10^3 cm^2 s^{-1}. This would make $E \approx 10^{-10}$, but this is also at present out of reach using our current numerical method. It is hoped that our value of $E = 2 \times 10^{-6}$ is small enough to provide the correct trend. That is, although we would like the (eddy) viscous force in the bulk of our simulated fluid core to be about nine orders of magnitude smaller than the other forces in (3), it actually is about six orders of magnitude smaller (except in the boundary layers, where it is comparable with the other forces); so our results may very well be in the correct asymptotic regime.

As we have dropped the inertial terms in (3), our choice of dimensional parameters is not a unique way of satisfying our choice of dimensionless parameters. One could make the following changes using positive scale factors a and c: $\Omega \to a\Omega$, $\bar{\eta} \to c\bar{\eta}$, $\bar{\nu} \to a\bar{\nu}$, $\bar{\kappa} \to c\bar{\kappa}$, $Q \to ac^2Q$ This has the following consequence:

$$P_r \to \frac{a}{c}P_r,\ P_m \to \frac{a}{c}P_m,\ \tau_\eta \to \frac{1}{c}\tau_\eta.$$

Then, for example, setting $a = 2 \times 10^{-4}$ and $c = 1$ would give more acceptable Prandtl numbers at the expense of an unacceptable rotation rate.

Also, our Ekman and Rayleigh numbers have been calculated using the values of the diffusivities 'seen' only by the large spatial scales. We use 'hyper-diffusivities' that increase with the spherical harmonic degree l of the variable they act on in the following way:

$$\bar{\nu}(l) = \bar{\nu}(1 + 0.075l^3)$$

This damps the small scales while allowing the large scales to experience much less diffusion. Physically, a hyper-diffusivity accounts for the presumed greater tendency for small-scale (resolved) eddies to transfer heat and momentum to sub-grid scale eddies because they are closer in size. By calculating the critical Rayleigh number using the hyper-diffusivities but no magnetic field, we estimate that our Rayleigh number is about 2000 times critical for this Ekman number.

In grid space, we use 49 Chebyshev radial levels in the outer core and 17 in the inner core with 32 latitudinal levels and 64 longitudinal levels. In spectral space, all Chebyshev degrees up to 46 are used in the outer core and only even degrees up to 30 in the inner core; and we use a triangular truncation of spherical harmonics, with degrees and orders up to 21. The 'hyper-diffusivities' make it possible to simulate large Rayleigh number, small Ekman number convection with this (relatively low) spatial resolution for the 2×10^6 timesteps that were required to span three magnetic diffusion times, starting from small temperature perturbations and a seed magnetic field.

3. Results

With our choice of scaling we are able to describe our results in dimensional units. The

three magnetic diffusion times that our simulation spans represents approximately 40 000 years. Our maximum fluid velocity, v_{max}, is about 0.4 cm s^{-1}, corresponding to a Rossby number, Ro = $v_{max}/(2\Omega d)$, of roughly 10^{-5}. Temperature perturbations can be as high as $T_{max} = 10^{-2}$ K. Our value of B_{max} is about 560 gauss, which usually occurs near the ICB; and the ratio, $\mathscr{M}/\mathscr{E}$, of the magnetic to kinetic energies is typically about 4000.

It should be noted that re-scaling with parameters a and c as outlined in the previous section would cause the following changes:

$$v_{max} \rightarrow cv_{max},\ T_{max} \rightarrow acT_{max},$$

$$B_{max} \rightarrow (ac)^{1/2} B_{max},$$

$$\mathrm{Ro} \rightarrow \frac{c}{a}\mathrm{Ro},\ \frac{\mathscr{M}}{\mathscr{E}} \rightarrow \frac{a}{c}\frac{\mathscr{M}}{\mathscr{E}}.$$

However, the magnetic Reynolds number, $v_{max}d/\overline{\eta} \approx 3 \times 10^3$, the Elsasser number, $B^2_{max}/(2\Omega\mu_0\bar{\rho}\overline{\eta}) \approx 5 \times 10^2$, the Hartmann number, $B_{max}d/(\mu_0\bar{\rho}\overline{\eta}\overline{\nu})^{1/2} \approx 2 \times 10^4$, the magnetic coupling parameter, $B^2_{max}/(2\Omega\mu_0\bar{\rho}dv_{max}) \approx 0.2$, and the Nusselt number, $\Delta T_0/\Delta T \approx 60$ (where ΔT_0 is the superadiabatic temperature drop that would exist across the outer core were there only thermal diffusion and no convection) all remain unaffected by this re-scaling.

Our solution is a self-sustaining dynamo, for a magnetic field has been maintained for three magnetic diffusion times. This is illustrated in Fig. 1, where the total kinetic energy (relative to the rotating frame of reference) and magnetic energy, integrated over the volume of the outer core, are plotted vs. time for roughly the last half of the simulated time. An even stronger argu-

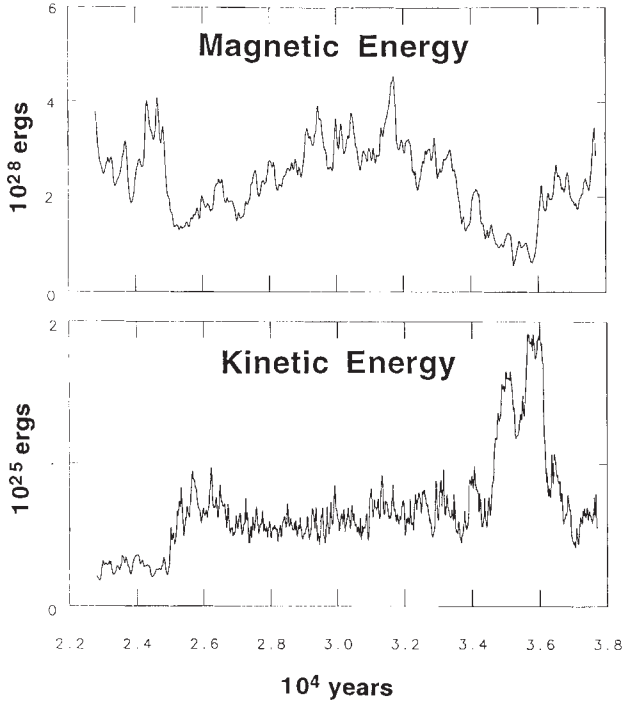


Fig. 1. Magnetic and kinetic energies integrated over the outer core and plotted vs. time for the last part of the simulation.

ment can be made when one computes (integrating over the volume of the outer core and using the magnetic hyper-diffusivity) the characteristic magnetic energy dissipation time proposed by Cattaneo et al. (1991),

$$\tau = \frac{\langle B^2 \rangle}{\langle \overline{\eta} \, | \nabla \times \boldsymbol{B} |^2 \rangle}$$

This was typically about 700 times smaller than our simulated time. Even if this were multiplied by their 'safety factor' of 10–100 to compensate for the dominant length scale, the magnetic field has been maintained for many times longer than their characteristic dissipation time.

As suggested by Hide (1979), another check for a self-sustaining dynamo is that the absolute value of the radial magnetic flux averaged over the CMB does not decrease when averaged over long periods of time. This value in our simulation is typically 5 gauss and varies as much as ± 3 gauss on a time scale of about 10^4 years; but there is no evidence of a decreasing trend in this value.

As seen in Fig. 1, the kinetic and magnetic energies are time dependent. We focus on the last half of the simulation because the solution needed to evolve, adjust, and forget its initial (random) conditions during the first half. It should be noted how the magnetic energy decreases, on the average, between 3.2×10^4 and 3.6×10^4 years and then quickly recovers. The kinetic energy grows to roughly twice its normal value between 3.4×10^4 and 3.6×10^4 years, while the magnetic energy is low. This increase in kinetic energy is mainly due to an increase in the differential rotation. This, in turn, generates more magnetic field which reduces the zonal flow and brings the kinetic energy in the differential rotation back down to normal.

The total energy of the magnetic field within the inner core is usually no more than 10% of that within the outer core, and the total magnetic energy exterior to the outer core is usually less than 1% of the magnetic energy within the outer core. Typically, 30–70% of the total kinetic energy in the outer core is in the axisymmetric part of the east–west flow, i.e. differential rotation; the axisymmetric meridional circulation typically represents less than 5% of the total kinetic energy. Likewise, the axisymmetric part of the east–west magnetic field, i.e. toroidal field, usually represents 30–70% of the total magnetic energy, whereas the axisymmetric meridional field, i.e. poloidal field, typically accounts for 1–20% of the magnetic energy in the outer core. However, the maximum poloidal field strength that occurs in the core is usually comparable with, and sometimes greater than, the maximum toroidal field strength.

The kinetic and magnetic energy spectra and the temperature variance spectrum show drops of several orders of magnitude with spherical harmonic degree (starting at $l = 6$), order, and Chebyshev degree. These large drops in energy and variance indicate that our solution (with the hyper-diffusivities) is numerically well resolved. The kinetic energy usually drops less than the magnetic energy and thermal variance probably owing to the strong effect of the nonlinear Lorentz force (4).

Viscous and magnetic coupling at the ICB and CMB produce time-dependent rotation rates of our inner core and mantle relative to the frame of reference rotating at $\boldsymbol{\Omega} = \Omega \hat{\boldsymbol{z}}$. The z-axis rotation rate of the inner core is almost always eastward, i.e. prograde relative to the initial non-convecting state. It can be as large as a few times 10^{-9} rad s^{-1} and changes on a time scale of about 500 years. The off-axis rates change amplitude and direction on a shorter time scale but are seldom greater than 10^{-10} rad s^{-1}. The z-axis rotation rate of our mantle (relative to $\boldsymbol{\Omega}$) can be either eastward or westward and is usually a few times 10^{-12} rad s^{-1}, similar to the Earth's decadal variations in the length of the day (Hide and Dickey, 1991). The off-axis rates are roughly 10^{-14} rad s^{-1}.

The 3D structure of our solution is very time dependent. A snapshot of the numerical solution (Figs. 2–4), at time 3.2×10^4 years of Fig. 1, illustrates typical thermal, flow and field structures. Fig. 2 shows the temperature perturbation and the radial component of fluid velocity in a spherical surface (with equal area projections) at mid-depth in the outer core and the radial com-

ponent of the magnetic field at the CMB. The temperature structure at this time is characterized by a large hot plume in the equatorial region and one in each polar region. The radial velocity has a similar pattern in addition to a 'Taylor column' structure at low latitude. These upwellings and downwellings propagate in longitude (some westward, some eastward) with phase velocities comparable with the horizontal fluid ve-

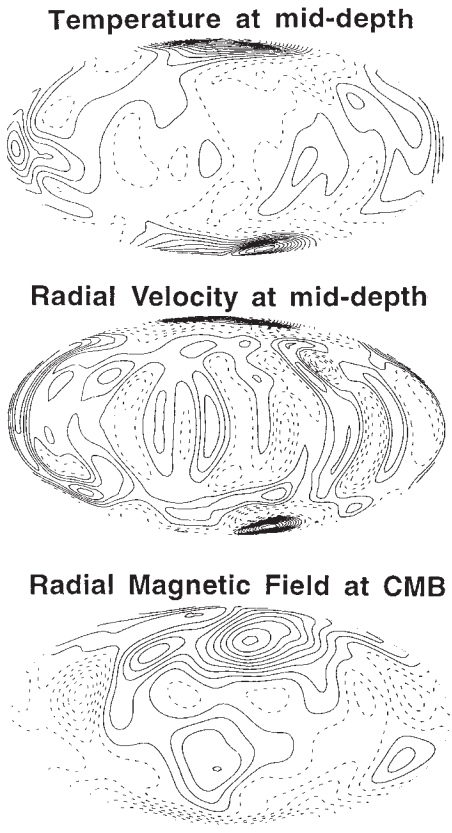


Fig. 2. A snapshot of the temperature perturbation (relative to the mean state) and the radial components of velocity and magnetic field in equal area projections of a spherical surface. The temperature and velocity are at mid-depth in the outer core and the magnetic field is at the CMB. Solid (broken) contours represent positive (negative) temperature perturbation and upward (downward) directed velocity and magnetic field. The temperature perturbation varies between $+4\times10^{-3}$ and -1×10^{-3} K, velocity between $+0.2$ and -0.1 cm s^{-1} and magnetic field between $+38$ and -25 gauss.

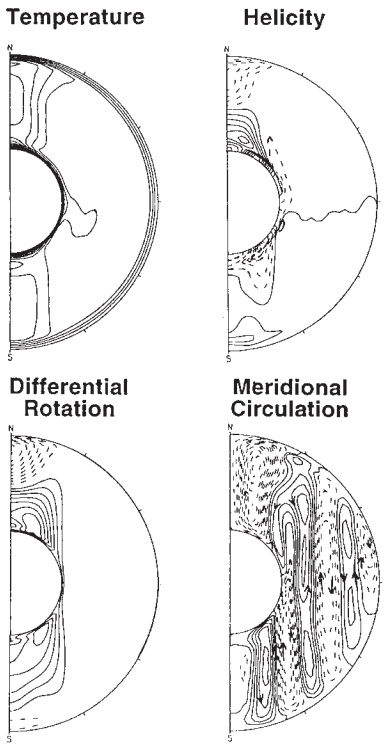


Fig. 3. A snapshot of the longitudinally averaged temperature perturbation, helicity, differential rotation (i.e. eastward angular velocity relative to the rotating frame), and meridional circulation plotted in a meridian plane showing the inner and outer cores. Solid (broken) contours represent positive (negative) values. For the streamlines of the meridional circulation, solid (broken) contours represent counter-clockwise (clockwise) directed flow. The temperature perturbation varies from 0 K at the CMB to 1.8×10^{-2} K at the ICB, helicity in the range $\pm3\times10^{-9}$ cm s^{-2}, angular velocity between $+4.8\times10^{-9}$ and -5.4×10^{-9} rad s^{-1} (or zonal velocity between $+0.30$ and -0.15 cm s^{-1}), and the maximum meridional flow is 0.1 cm s^{-1}.

locities and usually faster in the equatorial region than in the polar regions. However, throughout most of the outer core the fluid velocity and magnetic field are mainly horizontal and approximately aligned with each other. Where relatively large radial components do develop the local flow and field vectors remain aligned, either parallel or anti-parallel to each other, to minimize magnetic field distortion.

Our radial magnetic field at the CMB (Fig. 2) is a mixture of many modes, with the magnetic energy spectrum usually rather flat at the low spherical harmonic degrees, as it is for the Earth's CMB (Langel, 1987). Also, as for the Earth, the largest flux concentrations do not exist exactly over the poles. The maximum radial magnetic field at the CMB is fairly time dependent; it ranges between about 10 gauss and 50 gauss. The low end of this range is approximately today's value for the Earth (Bloxham and Jackson, 1992).

The secular variation of the radial component of our simulated magnetic field at the CMB has properties similar to the Earth's secular variation (Bloxham, and Gubbins, 1985; Bloxham and Jackson, 1992). Some flux patches remain relatively stationary for long periods of time whereas others propagate, sometimes westward, sometimes eastward. It is difficult to assign an average velocity to a magnetic feature because these features are not simple traveling waves. They merge, split, accelerate, change direction, oscillate, decay and grow, and there are usually several different features on the CMB propagating at different speeds and in different directions at the same time. A typical velocity of field lines at the CMB is 0.02 cm s^{-1}, roughly what has been measured on the

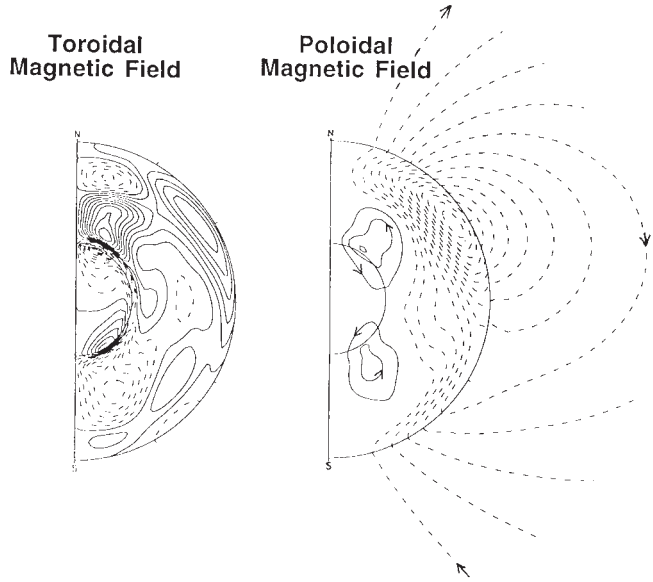


Fig. 4. A snapshot of the longitudinally averaged toroidal (i.e. east–west) magnetic field and longitudinally integrated poloidal (i.e. meridional) magnetic field in a meridian plane. The solid (broken) toroidal field contours represent eastward (westward) directed field; whereas solid (broken) poloidal lines of force represent counter-clockwise (clockwise) directed field. The toroidal field varies between +130 and −170 gauss and the maximum poloidal field is 210 gauss.

Earth (Stacey, 1922). This is a little smaller than our horizontal fluid velocity at $0.05d$ below the CMB, which can reach 0.1 cm s^{-1}, similar to estimates of the horizontal fluid velocity obtained from an inverse problem (Voorhies, 1986; Bloxham, 1989; Whaler, 1990; Jackson et al., 1993) using the radial component of the magnetic field at the CMB and making several assumptions (e.g. frozen flux, tangentially geostrophic flow, steady flow).

When our solution is averaged in longitude the resulting axisymmetric parts of the temperature and velocity show less time dependence than the 3D structures. Their profiles (again at the same timestep) are displayed in meridian planes in Fig. 3 which shows that most of the axisymmetric flow is taking place inside the 'tangent cylinder', the cylinder drawn parallel to the rotation axis and tangent to the ICB. Relative to the region inside the tangent cylinder, the region outside is nearly isothermal (except for the boundary layers) owing to the non-axisymmetric (Taylor column) convection there.

The angular rotation rate of the inner core (Fig. 3) is eastward, i.e. prograde, relative to the rotating frame of reference and to the mantle; however, there are stronger eastward zonal flows inside the tangent cylinder just above the ICB. In addition, there are westward jets inside the tangent cylinder near the CMB. Unlike the differential rotation in solar dynamo models (Gilman, 1983; Glatzmaier, 1984), which is maintained mainly by the convergence of the nonlinear Reynolds stress, here the strong zonal flow inside the tangent cylinder is maintained by the Coriolis force resulting from the meridional flow toward the rotation axis near the ICB and away from the axis near the CMB. There is also a weak westward thermal wind outside the tangent cylinder that usually peaks with the temperature in the equatorial region.

The longitudinally averaged helicity, the dot product of velocity and vorticity, is also shown in Fig. 3. It is fairly small and uniform outside the tangent cylinder, negative in the northern hemisphere and positive in the southern hemisphere. It is much larger inside the tangent cylinder and, in the northern hemisphere, is typically positive (i.e. right-handed) near the ICB and negative (left-handed) near the CMB. If one considers only the radial contribution to helicity from the axisymmetric flows, it is clear that where the strong outward flow inside the tangent cylinder has an eastward (solid contours) zonal flow in the northern hemisphere or westward (broken contours) zonal flow in the southern hemisphere the velocity and vorticity are parallel, and so the helicity is positive (solid contours).

During most of the simulation, the longitudinally averaged temperature and helicity and the differential rotation and meridional circulation have respective profiles that are qualitatively similar to those illustrated in the snapshot of Fig. 3. However, these profiles tend to become amplified in one hemisphere for a while and then in the other, with an oscillation 'period' of a few thousand years. For example, in Fig. 3 they are somewhat greater in the northern hemisphere.

The snapshot (at the same timestep) of the toroidal magnetic field, i.e. the axisymmetric eastward and westward directed field, and the poloidal magnetic field, i.e. the axisymmetric meridional lines of force, is illustrated in Fig. 4. At this time there exists strong eastward directed field (solid contours) inside the tangent cylinder in the northern hemisphere and westward directed field (broken contours) in the southern hemisphere with somewhat smaller oppositely directed toroidal fields in the inner core and near the CMB. The poloidal field is mainly a combination of dipolar and quadrupolar modes. It should be noted how the radial gradient of the zonal flow (Fig. 3) inside the tangent cylinder shears the existing poloidal field maintaining the toroidal field there (Fig. 4), and (see Fig. 1 of Glatzmaier (1985a)) how the right-handed helical flow (solid contours, Fig. 3) twists the existing eastward toroidal field (solid contours, Fig. 4) inside the tangent cylinder in the northern hemisphere maintaining the counter-clockwise (solid contours, Fig. 4) poloidal field there. The oppositely directed poloidal field is generated outside the tangent cylinder in the northern hemisphere near the CMB because helicity is negative (left-handed) there.

The magnetic field diffuses into the inner core

from the ICB. As the field within the inner core can only undergo diffusion, it changes on a magnetic diffusion time scale that is larger than the advective time scale for field in the outer (convecting) core. In addition, the radial gradient of the zonal flow is usually greatest near the ICB; this effectively generates toroidal field from poloidal field, always providing a strong source of new field at the ICB. However, as found in the axisymmetric intermediate dynamo solutions of Hollerbach and Jones (1993b, 1995), we find that the presence of a finitely conducting inner core diminishes the chaotic tendency the magnetic field otherwise would have in our 3D solutions. The thin (parametrized) conducting layer above the CMB provides magnetic coupling to the mantle but is less effective in stabilizing the evolution of the magnetic field in the upper part of the outer core.

The longitudinally averaged magnetic field in most of the outer core normally has the opposite polarity to that in the inner core (Fig. 4). It also varies on a similar (advective) time scale as the axisymmetric temperature and velocity. A time series shows that this outer core field periodically attempts to reverse its polarity but usually fails and returns to its normal profile because the field in the inner core usually does not have enough time to decay away and allow the reversed polarity to diffuse in before the field in the outer core changes again. At times, it evolves in a manner very similar to one of our axisymmetric intermediate dynamo solutions which undergoes a hemispheric oscillation but never reverses (Fig. 2c of Glatzmaier and Roberts (1993)). However, in that model the prescribed buoyancy force and helicity are time independent. In our present 3D solution, the buoyancy and helicity are determined self-consistently; their time-dependent structures are influenced by the evolution of the velocity and magnetic field.

During the first half of our simulation, the field went through several adjustments; however, our solution maintained the same magnetic polarity from 2.0×10^4 to 3.6×10^4 years. It then quickly reversed. It should be noted (Fig. 1) how the magnetic energy decreases, on the average, by about a factor of four during the 4000 years before the reversal and then quickly recovers after the reversal.

4. Discussion

To learn more about the structure and dynamics of the geodynamo we have computed and analyzed the first 3D self-consistent numerical solution of the MHD equations that describe thermal convection and magnetic field generation in a rotating spherical fluid shell with a conducting solid inner core. Our solution is a self-sustaining dynamo that has maintained a magnetic field for three magnetic diffusion times. The pattern, amplitude, and secular variation of our radial magnetic field at the CMB and the changes in the length of the day are similar to those observed on the Earth. In addition, the horizontal flow just below the CMB has an amplitude like that inferred from inversions of the Earth's CMB field.

Our solution has several interesting properties. In agreement with traditional assumptions for the geodynamo, total magnetic energy is at least three orders of magnitude greater than total kinetic energy, and the total energy in the (axisymmetric) toroidal field is usually several times greater than that in the (axisymmetric) poloidal field. However, contrary to traditional assumptions, the maximum poloidal field strength in our solution is usually comparable with the maximum toroidal field strength. We also find that the maximum amplitudes of the longitudinally averaged temperature gradient, shear flow, and helicity occur inside the tangent cylinder and oscillate between the northern and southern hemispheres on a time scale of a few thousand years. These axisymmetric structures and their oscillatory time dependence have new implications for the alpha and omega profiles prescribed in mean field kinematic and intermediate dynamo models.

The solution also shows that the maximum of the longitudinally averaged magnetic field in the outer core oscillates between the two hemispheres with the maxima in the temperature and flow. However, only once in many attempts does the field succeed in reversing its polarity because the field in the inner core, which has the opposite

polarity to the field in most of the outer core, usually does not have enough time to reverse before the field in the outer core changes again. In addition, magnetic energy decreased for several thousand years before the reversal that did occur in our simulation and then quickly recovered after the reversal.

Several model improvements will be needed to obtain more realistic simulations of the geodynamo. Abandoning the Boussinesq approximation and using a more realistic reference state with the anelastic approximation would be relatively simple improvement (Glatzmaier, 1995). Compositional convection needs to be included (Braginsky, 1963). Viscous and thermal eddy diffusivities need to be tensor diffusivities to capture their local anisotropic properties (Braginsky and Meytlis, 1991; St. Pierre, personal communication, 1994). Lateral variations in the radial heat flux at the CMB owing to large temperature variations in the lower mantle certainly must influence the structure and dynamics of thermal convection in the outer core and possibly the magnetic reversal frequency (Hide, 1967; Loper and McCartney, 1986; Bloxham and Gubbins, 1987). Topographic-coupling should take the place of the non-axisymmetric viscous coupling (Kuang and Bloxham, 1993). The effects on core dynamics of luni-solar precession of the mantle also need to be tested in a 3D model (Malkus, 1968; Loper, 1975; Vanyo, 1991). Each of these improvements has the potential to alter dramatically the character of the dynamo solution. However, each will require simulating a significant fraction of a magnetic diffusion time to analyze its effects, and just one simulation like the one presented here requires a tremendous amount of computing resources.

Acknowledgements

The computing resources for this simulation were provided by the NSF Pittsburgh Supercomputing Center via grant MCA94P016P. Support for this research was provided by the Institute of Geophysics and Planetary Physics at Los Alamos and by Los Alamos LDRD Grant 93202. We would also like to thank the referees for helpful suggestions.

References

Bercovici, D., Schubert, G., Glatzmaier, G.A. and Zebib, A., 1989. Three-dimensional thermal convection in a spherical shell. J. Fluid Mech., 206: 75–104.

Bloxham, J., 1989. Simple models of fluid flow at the core surface derived from geomagnetic field models. Geophys. J. Int., 99: 173–182

Bloxham, J. and Gubbins, D., 1985. The secular variation of the Earth's magnetic field. Nature, 317: 777–781.

Bloxham, J. and Gubbins, D., 1987. Thermal core–mantle interactions. Nature. 325: 511–513.

Bloxham, J. and Jackson, A., 1992. Time-dependent mapping of the magnetic field at the core–mantle boundary. J. Geophys. Res., 97: 19537–19563

Braginsky, S.I., 1963. Structure of the F layer and reasons for convection in the Earth's core. Sov. Phys. Dokl., 149: 8–10.

Braginsky, S.I. and Meytlis, V.P., 1991. Local turbulence in the Earth's core. Geophys. Astrophys. Fluid Dyn., 55: 71–87.

Brandenburg, A., Nordlund, A., Pulkkinen, P., Stein, R.F. and Tuominen, I., 1990. Simulation of turbulent cyclonic magneto-convection. Astron. Astrophys., 232: 277–291.

Cardin, P. and Olson, P., 1992. An experimental approach to thermochemical convection in the Earth's core. Geophys. Res. Lett., 19: 1995–1998.

Cattaneo, F., Hughes, D.W. and Weiss, N.O., 1991. What is a stellar dynamo?. Mon. Not. R. Astron. Soc., 253: 479–484.

Fearn, D.R. and Proctor, M.R.E., 1987. On the computation of steady self-consistent spherical dynamos. Geophys. Astrophys. Fluid Dyn., 38: 293–325.

Gans, R.F., 1972. Viscosity of the Earth's core. J. Geophys. Res., 77: 360–366.

Gilman, P.A., 1983. 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.

Gilman, P.A. and Miller, J., 1981. Dynamically consistent nonlinear dynamos driven by convection in a rotating spherical shell. Astrophys. J. Suppl., 46; 211–238.

Glatzmaier, G.A., 1984. Numerical simulations of stellar convective dynamos. I. The model and method, J. Comput. Phys., 55: 461–484.

Glatzmaier, G.A., 1985a. Numerical simulations of stellar convective dynamos. II Field propagation in the convection zone. Astrophys. J., 291: 300–307.

Glatzmaier, G.A., 1985b. Numerical simulations of stellar convective dynamos. III At the base of the convection zone. Geophys. Astrophys. Fluid Dynamics 31: 137–150.

Glatzmaier, G.A., 1995. Global three-dimensional numerical models of planetary dynamos, in preparation.

Glatzmaier, G.A. and Olson, P., 1993. Highly supercritical

thermal convection in a rotating spherical shell: centrifugal vs. radial gravity. Geophys. Astrophys. Fluid Dyn., 70: 113–136.

Glatzmaier, G.A. and Roberts, P.H., 1993. Intermediate dynamo models. J. Geomagn. Geoelectr., 45: 1605–1616.

Gubbins, D., 1981. Rotation of the inner core. J. Geophys. Res., B, 86: 11695–11699.

Hart, J.E., Glatzmaier, G.A. and Toomre, J., 1986. Space-laboratory and numerical simulations of thermal convection in a rotating hemispherical shell with radial gravity. J. Fluid Mech., 173: 519–544.

Hide, R., 1967. Motions of the Earth's core and mantle, and variations of the main geomagnetic field. Science, 157: 55–56.

Hide, R., 1979. Dynamo theorems. Geophys. Astrophys. Fluid Dyn., 14: 183–186.

Hide, R. and Dickey, J.O., 1991. Earth's variable rotation. Science, 253: 629–637.

Hollerbach, R., 1994. Magnetohydrodynamic Ekman and Stewartson layers in a rotating spherical shell. Proc. R. Soc. London, A, 444: 333–346.

Hollerbach, R. and Ierley, G.R., 1991. A modal α^2-dynamo in the limit of asymptotically small viscosity. Geophys. Astrophys. Fluid Dyn., 56: 133–158.

Hollerbach, R. and Jones, C.A., 1993a. A geodynamo model incorporating a finitely conducting inner core. Phys. Earth Planet Inter., 75: 317–327.

Hollerbach, R. and Jones, C.A., 1993b. Influence of the Earth's inner core on geomagnetic fluctuations and reversals. Nature, 365: 541–543.

Hollerbach, R. and Jones, C.A., 1995. On the magnetically stabilizing role of the Earth's inner core. Phys. Earth Planet. Inter., 87: 171–181.

Jackson, A., Bloxham, J. and Gubbins, D., 1993. Time-dependent flow at the core surface and conservation of angular momentum in the coupled core-mantle system. Geophys. Monogr. Am. Geophys. Union, 72: (IUGG Vol. 12) 97–107.

Kageyama, A., Sato, T., Watanabe, K., Horiuchi, R., Hayashi, T., Todo, Y., Watanabe, T.H. and Takamara, H., 1995. Computer simulation of a magnetohydrodynamic dynamo II. Phys. Plasmas, in press.

Kuang, W. and Bloxham, J., 1993. On the effect of boundary topography on flow in the Earth's core. Geophys. Astrophys. Fluid Dyn., 72: 161–195.

Langel, R.A., 1987. The main field. In: J.A. Jacobs (Editor), Geomagnetism, Vol. 1. Academic Press, London, pp. 249–512.

Loper, D.E., 1975. Torque balance and energy budget for the precessionally driven dynamo. Phys. Earth Planet. Inter., 11: 43–60.

Loper, D.E. and McCartney, K., 1986. Mantle plumes and the periodicity of magnetic field reversals. Geophys. Res. Lett., 13: 1525–1528.

Malkus, W.V.R., 1968. Precession of the Earth as the cause of geomagnetism. Science, 160: 259–264.

Meneguzzi, M. and Pouquet, A., 1989. Turbulent dynamos driven by convection. J. Fluid Mech., 205: 297–318.

Nordlund, A., Brandenburg, A., Jennings, R.L., Rieutord, M., Ruokolainen, J., Stein, R.F. and Tuominen, I., 1992. Dynamo action in stratified convection with overshoot. Astrophys. J., 392: 647–652.

Parker, E.N., 1955. Hydromagnetic dynamo models. Astrophys. J., 122: 293–314.

St. Pierre, M.G., 1992. The stability of the magnetostrophic approximation I: Taylor state solutions. Geophys. Astrophys. Fluid Dyn., 67: 99–115.

St. Pierre, M.G., 1993. The strong field branch of the Childress–Soward dynamo. In: M.R.E. Proctor, P.C. Mathews and A.M. Rucklidge (Editors), Theory of Solar and Planetary Dynamos. Cambridge University Press, Cambridge, pp. 295–302.

Stacey, F.D., 1992. Physics of the Earth, 3rd edn. Brookfield Press, Brisbane.

Taylor, J.B., 1963. The magneto-hydrodynamics of a rotating fluid and the Earth's dynamo problem. Proc. R. Soc. London, Ser. A, 274: 274–283.

Vanyo, J.P., 1991. A geodynamo powered by luni-solar precession. Geophys. Astrophys. Fluid Dyn., 59: 209–234.

Voorhies, C.V., 1986. Steady flows at the top of Earth's core derived from geomagnetic field models. J. Geophys. Res., 91: 12444–12466.

Whaler, K.A., 1990. A steady velocity field at the top of the Earth's core in the frozen-flux approximation—errata and further comments, Geophys. J. Int., 102: 507–109.

Zhang, K.K. and Busse, F.H., 1989. Convection driven magnetohydrodynamic dynamos in rotating spherical shells. Geophys. Astrophys. Fluid Dyn., 49: 97–116.

Zhang, K.K. and Busse, F.H., 1990. Generation of magnetic fields by convection in a rotating spherical fluid shell of infinite Prandtl number. Phys. Earth Planet. Inter., 59: 208–222.

A three-dimensional self-consistent computer simulation of a geomagnetic field reversal

Gary A. Glatzmaier* **& Paul H. Roberts†**

* Institute of Geophysics and Planetary Physics, Los Alamos National Laboratory, Los Alamos, New Mexico 87454, USA
† Institute of Geophysics and Planetary Physics, University of California, Los Angeles, California 90024, USA

A three-dimensional, self-consistent numerical model of the geodynamo is described, that maintains a magnetic field for over 40,000 years. The model, which incorporates a finitely conducting inner core, undergoes several polarity excursions and then, near the end of the simulation, a successful reversal of the dipole moment. This simulated magnetic field reversal shares some features with real reversals of the geomagnetic field, and may provide insight into the geomagnetic reversal mechanism.

A FUNDAMENTAL goal of geophysics is a coherent understanding of the structure and dynamics of the Earth's interior. An integral part of this understanding must be a model of the Earth's magnetic field that reproduces its salient features: a field that is maintained for many magnetic diffusion times, is dominantly dipolar at the surface with a dipole axis that on the average lies close to the geographic axis of rotation, and exhibits secular variation with occasional excursions and reversals. The only plausible candidate for such a model is the dynamo model, in which new magnetic field is continually being generated by the shearing and twisting fluid motions within the Earth's liquid, electrically conducting, outer core[1,2]. According to ref. 3, Albert Einstein considered understanding the origin of the Earth's magnetic field as being one of the five most important unsolved problems in physics. Today we know that a nonlinear three-dimensional model of the magnetohydrodynamics (MHD) of the Earth's core is needed to explain the structure and the long-term evolution of the geomagnetic field.

There are two principal reasons for this. First, it is widely perceived that linear mathematical analysis, which has provided great insights into basic dynamo processes, is unlikely to contribute as fundamentally to the understanding of the geodynamo because the Earth operates in the so-called 'strong-field' regime. Analytical methods are successful in studying 'weak-field' regimes, in which the magnetic forces can be treated as a perturbation in the dynamics. However, in the strong-field regime the nonlinear magnetic Lorentz force is as large as the Coriolis force and therefore cannot be treated as a perturbation. A nonlinear numerical computation is therefore required to simulate the MHD.

The second reason arises from Cowling's theorem[4], which shows that a self-sustained magnetic field produced by a dynamo cannot be axisymmetric. But most numerical models of dynamos have been 'mean-field' kinematic models, which require a prescription of the axisymmetric effects of some hypothetical three-dimensional convection and solve for the evolution of only the two-dimensional axisymmetric parts of the magnetic fields[5]. Intermediate mean-field models[6-11] go somewhat further by also solving for the axisymmetric zonal flow and meridional circulation, but still require a prescribed two-dimensional structure of some hypothetical averaged helical flow (the alpha effect[2]) and usually a prescribed structure for the buoyancy force or for the zonal thermal wind shear it produces (the omega effect[1]). They are therefore not self-consistent because the solutions strongly depend on how one chooses to prescribe the alpha and omega structures. The simplest way to generate a three-dimensional magnetic field is to drive it with a prescribed three-dimensional velocity profile[12-14]; but in this approach the (kinematic) velocity is not a self-consistent convective solution with nonlinear feedback from the Lorentz force. Travelling-wave solutions can also be generated[15] but lack mode-dependent temporal behaviour. Magneto-convection simulations in three dimensions are closer to self-consistency because the time-dependent thermodynamic, velocity, and magnetic fields are all solved in three dimensions with nonlinear feedback[16-19]; but the main part of the magnetic field is maintained via boundary conditions. These approaches have been used because of their simplicity and relatively small computing resource requirements; but they are limited because either they do not produce self-consistent solutions of the full three-dimensional MHD equations or, in magneto-convection, they do not produce self-sustaining magnetic fields and are therefore not suitable for studying field reversals. They have been and continue to be extremely useful for providing a relatively cheap way of testing the influence of new physics[9] or studying secular variation of the field on short timescales[19].

Self-consistent numerical simulations of three-dimensional convective dynamos have been computed in planar (cartesian) geometry to study local dynamo action[20-23] and in global (spherical) geometry[24-29] to study the solar dynamo. Of these, however, the only models that generate magnetic energy an order of magnitude greater than kinetic energy and therefore begin to approach the strong-field regime appropriate for studying the geodynamo are the recent models by St Pierre[23] and Kageyama *et al.*[29]. In the other models, the generated magnetic energy is several orders of magnitude less than the kinetic energy.

Here we present a fully self-consistent three-dimensional numerical simulation of a convective strong-field dynamo in a spherical shell with a finitely conducting inner core[30]. A magnetic field is maintained for more than three magnetic diffusion times and has energy at least three orders of magnitude greater than the kinetic energy of the convection that maintains it. The exciting feature that we focus on in this paper is a reversal of the dipole polarity that occurs near the end of our simulation. The simulation required over 2,000,000 computational time steps that, over a period of more than a year, took more than 2,000 CPU hours on a Cray C-90, which is why we have decided to report on our results now instead of waiting for our simulation to span a much longer period of time. With only one reversal simulated we cannot yet say anything about the statistical behaviour of reversals in our model. Some minor changes in our model would also bring it closer to geophysical reality. But because our simulation is self-consistent and maintains a field that resem-

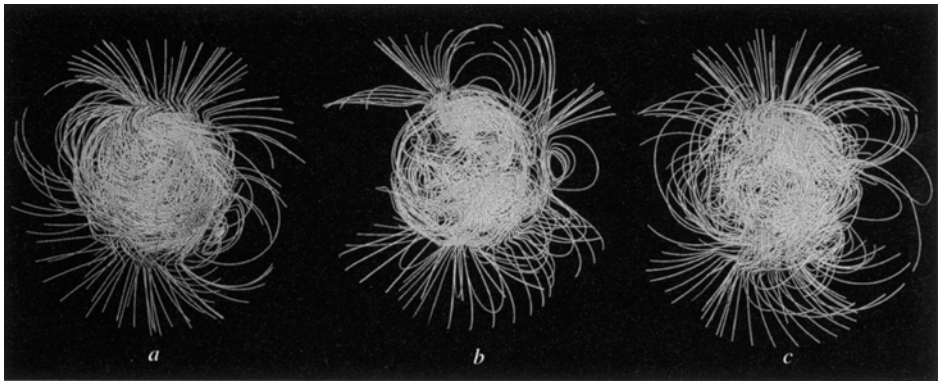


FIG. 1 The three-dimensional (3D) magnetic field structure portrayed through lines of force plotted out to the surface of our modelled Earth. Snapshots are displayed: *a*, before the reversal (9,000 years before the end of our simulation); *b*, midway through the transition as seen at the surface (that is, when the axial dipole part of the field at the surface goes through zero, about 4,000 years before the end); *c*, after the reversal (at the end of our simulation). Lines are yellow (blue) where the radial component of the field is directed outward (inward). The rotation axis is vertical. One hundred lines of force are plotted in each snapshot, so the relative field intensity is not represented in this figure.

bles the Earth's in many respects, we believe that it provides a plausible description of the geomagnetic field and the way it reverses.

Model description

The numerical model (G.A.G., manuscript in preparation) solves the nonlinear MHD equations that govern the three-dimensional structure and evolution of an electrically conducting fluid undergoing thermal convection in a rapidly rotating spherical shell, our model of the Earth's outer fluid core. A specified heat flux at the inner core boundary (ICB) drives thermal convection in the fluid core. This convection, influenced by the rotation of the core, twists and shears magnetic field, generating new magnetic field to replace that which diffuses away. The field diffuses into a solid, electrically conducting, inner core providing magnetic torque between the inner and outer cores. Magnetic torque also exists between the outer fluid core and a solid mantle above through a thin conducting layer at the core mantle boundary (CMB). The rest of the mantle is assumed to be an insulator, so the field above this layer is a source-free potential field. Time-dependent rotation rates of the solid inner core and solid mantle are determined by the net torques at the ICB and CMB, respectively.

The model prescriptions[30] (mass, dimensions, rotation rate and material properties) are Earth-like, except the applied heat flux at the ICB, which is somewhat higher to compensate for the lack of compositional buoyancy sources in this version of our model, and the viscosity of the fluid core, which owing to computational limitations was chosen to be larger than the Earth's. But because the viscous forces in our model (outside the thin viscous boundary layers) are already about six orders of magnitude smaller than the Coriolis and Lorentz forces, our solution is probably in the correct asymptotic regime for the Earth. In comparison with two other strong-field dynamos, Kageyama *et al.*[29] assume that the convecting fluid is a compressible gas with a viscosity more than two orders of magnitude greater than our viscosity (when scaled appropriately, using the Earth's rotation rate and core radius), and St Pierre[23] uses a viscosity slightly larger than in our model.

The electrical conductivity of the solid inner core is probably much the same as that of the fluid outer core but, in our initial test calculations (those done before the simulation we are reporting on here), we assumed for simplicity an insulating inner core. The result was a chaotic magnetic field that, unlike the Earth's, reversed its dipole polarity roughly every thousand years. Hollerbach and Jones[9] then demonstrated, in their two-dimensional mean-field dynamo model, that a finitely conducting solid inner core provides a degree of stability to the magnetic field. Their alpha-omega dynamo generates a field that never reverses, although the magnetic energy undergoes periodic changes. The lack of field reversals in their solution is probably because of the particular prescription of their alpha and omega driving terms that are held constant in time. But the reversal inhibition they demonstrated and the reversal mechanism they suggested motivated us to include a finitely conducting inner core in our three-dimensional convective model. Again, in comparison, Kageyama *et al.*[29] do not have a finitely conducting inner core and therefore have no magnetic coupling between their outer and inner cores, and the model by St Pierre[23] is planar.

Our present solution, with a finitely conducting inner core, spans >40,000 years, more than three magnetic diffusion times, with no indication that it will decay away, which is suggestive evidence that our solution is a self-sustaining convective dynamo. The solution begins with random small-scale temperature perturbations and a seed magnetic field. After an initial period of adjustment (~10,000 years) during which the dipole part of the field gradually becomes dominant, our time-dependent solution maintains its dipole polarity until near the end of the simulation, when it reverses in little more than 1,000 years and then maintains the new dipole polarity for roughly the remaining 4,000 years of the simulation.

Field structure during the reversal

To illustrate the magnetic field reversal mechanism, we display the change in the structure of the field during the last 9,000 years of our simulation with three snapshots in each of Figs 1–4. The three-dimensional field structure is portrayed in Fig. 1 via lines of force at 9,000 years before the end of our simulation (Fig.

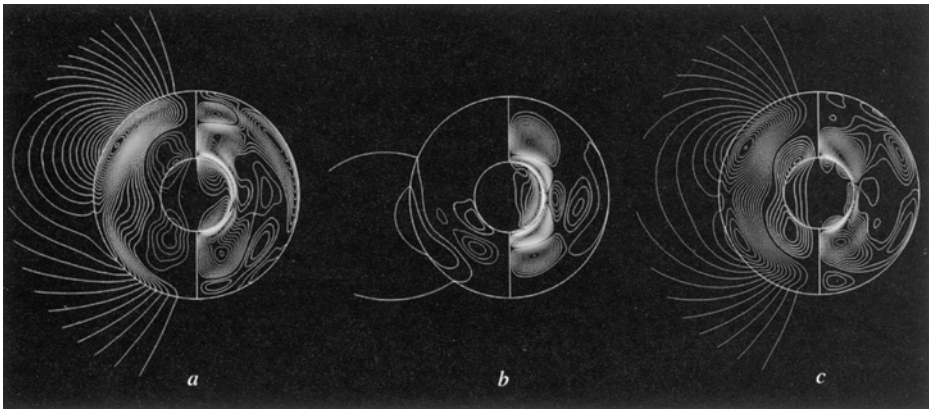


FIG. 2 The longitudinal average of the 3D magnetic field displayed with contours of the toroidal (east–west) part of the field plotted in the right hemispheres and lines of force of the meridional (poloidal) part of the field plotted in the left hemispheres. Snapshots are displayed: *a*, before the reversal (9,000 years before the end of our simulation); *b*, midway through the transition as seen at the ICB (that is, when the axial dipole part of the field at the ICB goes through zero, about 5,000 years before the end); *c*, after the reversal (at the end of our simulation). The north (south) geographic pole is at the top (bottom) of each plot. The outer circular boundary is the core–mantle boundary and the inner circular boundary is the inner core boundary. Red (blue) contours are eastward (westward)-directed toroidal field. Green (yellow) lines of force are clockwise (anticlockwise)-directed poloidal field that are plotted out to the surface. Note that for the series of plots in this figure the relative field intensity is represented in terms of the number of toroidal field contours and the density of poloidal field lines.

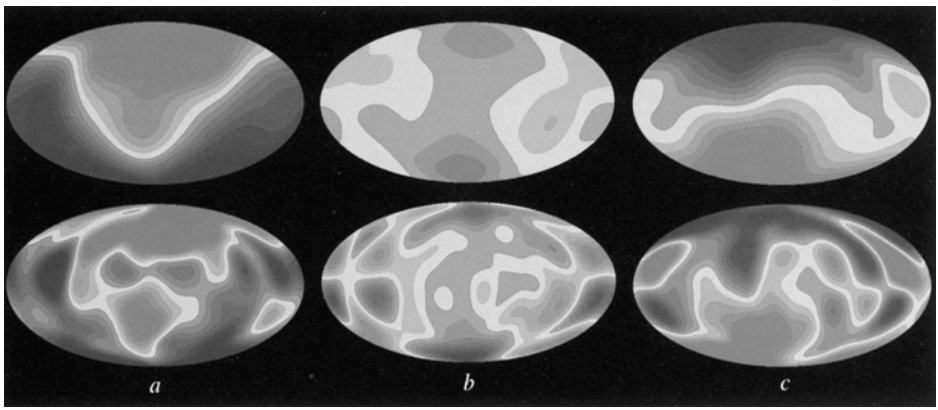


FIG. 3 The radial component of the magnetic field plotted in a Hammer projection of the core–mantle boundary (lower plots) and of the surface (upper plots). Snapshots *a*, *b* and *c* are displayed at the same times as those in Fig. 1. Red (blue) contours represent outward (inward)-directed field. The relative intensities at the different times are also reflected in the colour contours. But the field intensity at the surface has been multiplied by a factor of ten to obtain colours in the surface plots that are comparable to those in the CMB plots.

1*a*), at roughly the middle of the polarity transition as seen at the surface (~4,000 years before the end; Fig. 1*b*), and at the end of our simulation (Fig. 1*c*). We begin plotting the lines of force at the surface of our modelled Earth. They penetrate inward through the insulating mantle and then into the outer fluid core where the field is generated. The transition from the relatively smooth structure of the potential field outside the core to the much more complicated and intense field structure inside the core is quite striking. The maximum field intensity usually occurs near the ICB and is typically between 30 and 50 mT. The field at the surface has a dominantly dipolar structure before (Fig. 1*a*) and after (Fig. 1*c*) the reversal, with the dipole axis nearly aligned with the rotation axis, which is vertical in Fig. 1. During the polarity transition (Fig. 1*b*) the field structure at the

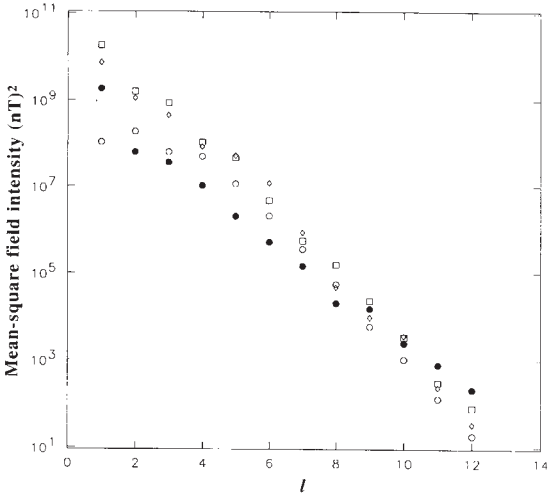


FIG. 4 The mean-square field intensity over the surface plotted as $(l+1)\Sigma_m[(g_l^m)^2+(h_l^m)^2]$, where g_l^m and h_l^m are the traditional Gauss coefficients for the potential field outside the core, and l and m are the degree and order, respectively, in the spherical harmonic expansion. Spectra are displayed at the same three times as those in Figs 1 and 3: squares before, open circles during, and diamonds after the reversal. The spectrum for the present-day Earth, based on MAGSAT data[34], is plotted as filled circles.

surface is much more complicated and the dipole axis passes through the equatorial plane.

To illustrate the field structure within the inner and outer cores more easily, we plot in Fig. 2 the longitudinal average of the three-dimensional field at 9,000 years before the end (Fig. 2*a*), at roughly the middle of the polarity transition as seen at the ICB (~5,000 years before the end; Fig. 2*b*), and at the end of our simulation (Fig. 2*c*). The right side of each plot shows contours of the east–west (toroidal) part of the field; the left side shows lines of force of the meridional (poloidal) part of the field. There are typically two main toroidal field concentrations, one in each hemisphere, in opposite directions and usually inside the imaginary cylinder tangent to the inner core where large zonal flows shear poloidal field, so generating toroidal field. Toroidal field also diffuses into the inner core from the ICB, where it is generated when poloidal field is sheared by the inner core as it moves differentially with respect to the fluid just outside the ICB.

The longitudinally averaged poloidal field (left sides of Fig. 2) typically has two dipolar polarities: one in the outer part of the fluid core, which is also the dipole polarity observed at the surface, and the opposite polarity in the inner part of the fluid core and the inner solid core. Poloidal field is generated by helical flow that twists toroidal field; in the Northern Hemisphere, for example, the time-dependent helicity of the flow (that is, the correlation between velocity and vorticity) is usually right-handed and much larger inside the 'tangent cylinder' and left-handed outside[30]. This illustrates one of the limitations of models that require, unlike ours, a prescription of the fluid flow instead of solving for it self-consistently. For example, the helical flow structure prescribed by Hollerbach and Jones[9] has no radial or time dependence and the field has only a single dipole polarity that never reverses. By contrast, it is the interaction of the two dipolar polarities present in our solution that plays an essential role in our reversal mechanism in a way that brings to mind the simple interaction between the two disks of the Rikitake dynamo model[31].

A movie of our simulation shows how the field in the fluid outer core is continually attempting to reverse its axial dipole polarity on a short timescale (~100 years) corresponding to convective overturning but usually fails because the field in the solid inner core, which can only change on a longer diffusive timescale (~1,600 years), usually does not have enough time to diffuse away before it is regenerated at the ICB. It also shows how the axial quadrupole part of the outer field tends to reverse its polarity on a roughly thousand-year timescale, causing a hemispheric (non-reversing dipole) oscillation in the structure and intensity of the outer poloidal field[10] because the sum of an axial dipole and an axial quadrupole results in an enhanced axial poloidal field in one hemisphere and a diminished poloidal field in the other. Once in many attempts the three-dimensional configurations of the buoyancy, flow and magnetic fields in the outer core are right for a long enough period of time for the inner core axial dipole field to diffuse away (Fig. 2*b*), thus allowing the reversed axial dipole polarity in the outer core to diffuse into the inner core. If this one model is representative of the Earth, it suggests that the strong nonlinear feedbacks in three dimensions and the different timescales of the fluid and solid cores are responsible for the Earth's stochastic reversal record.

When the reversal occurs in our simulation, the energy of the total field in the core is about one-quarter of its typical value. Once the new field polarity has become established, the magnetic energy in the fluid core quickly recovers. During the transition, eastward and westward toroidal field are alternately generated several times in both hemispheres at the ICB before the reversed polarity finally becomes established. Notice how the toroidal field is asymmetric with respect to the Equator before (Fig. 2*a*) and after (Fig. 2*c*) the transition but is symmetric midway through the transition (Fig. 2*b*). In our simulation, the toroidal field reverses first, then the inner poloidal field that penetrates the solid core, and finally, somewhat later, the outer poloidal field that appears at the Earth's surface reverses. This entire process takes (depending on how one defines the beginning and end of the reversal) a little more than 1,000 years, roughly the characteristic magnetic diffusion timescale for the inner core.

Figure 3 shows snapshots of just the radial component of the field plotted at the CMB (lower plots) and at the surface (upper plots) at the same three times as those depicted in Fig. 1. Because the dipolar part of the pattern decays the most slowly with radius, its contribution is greater at the surface than at the CMB. The dipole parts of these fields are nearly axial before (Fig. 3*a*) and after (Fig. 3*c*) the transition and equatorial midway through

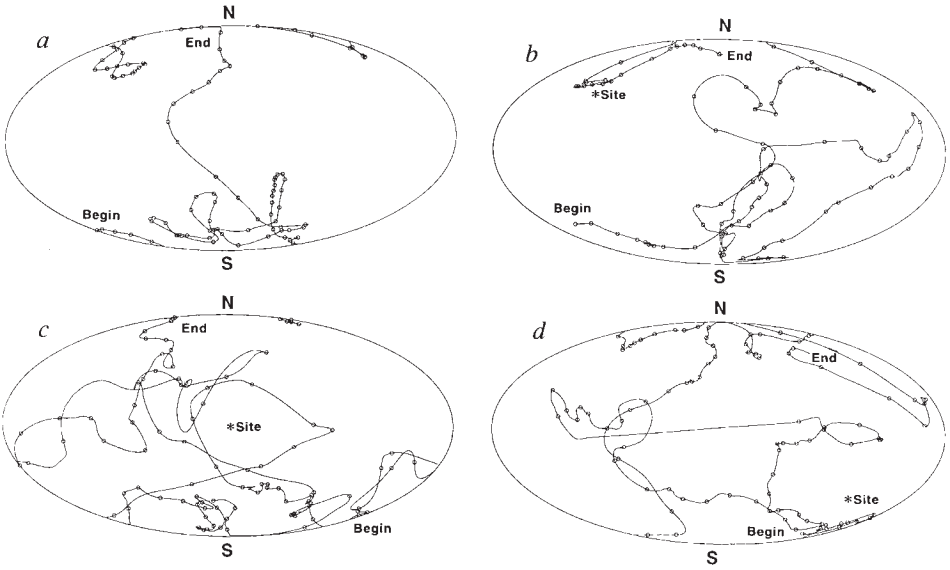


FIG. 5 The true geomagnetic pole (*a*) and the virtual geomagnetic poles (*b–d*) computed using the field directions at three surface sites: *a*, +40° latitude and 60° longitude, *b*, 0° latitude and 180° longitude, and *c*, −40° latitude and 300° longitude. The pole, with the inward-directed field, is plotted in a Hammer projection of the surface with markers at 100-year intervals during the last 9,000 years of our simulation.

the transition (Fig. 3*b*). But the dipole part of the field decreases more during the transition than the other modes do; therefore the equatorial dipole part of our transitional field is less dominant than the axial dipole part typically is before and after the transition. For example, the large radial fields at the poles in Fig. 3*b* are the result of a relatively strong quadrupole mode.

The CMB plots of Fig. 3*a*, *c* are qualitatively similar to the Earth's present field projected onto the CMB[32] with multimode contributions and several flux concentrations. The maximum intensity of the radial component of the field at the CMB in Fig. 3*a*, *c* is ~3.0 mT; the maximum intensity in Fig. 3*b* is 0.8 mT. These values can be compared with the maximum intensity of the Earth's present radial field at the CMB (ref. 32), which is ~1.0 mT.

Our simulated field structure at the surface can be compared with the Earth's present surface field stucture by plotting the mean-square field intensity over the surface as a function of the spherical harmonic degree, with the traditional Gauss coefficients[33,34]. The spectra (displayed only out to spherical harmonic degree 12) at the same three times as those used in Figs. 1 and 3 are plotted in Fig. 4. The spectra corresponding to the snapshots long before (squares) and long after (diamonds) the transition show the dipolar ($l=1$) contribution dominant over the quadrupolar ($l=2$) and octupolar ($l=3$) contributions. These two spectra indicate that we could be driving the convection a little too hard because our magnetic energies for the low degrees are higher than those of the present-day Earth[33,34], which are plotted as filled circles in Fig. 4. Also, because we have no

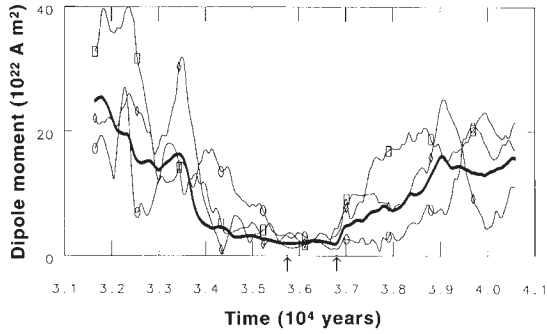


FIG. 6 The true dipole moment (heavy curve) and virtual dipole moments (marked by squares, circles and diamonds) computed using the field at the three sampling sites used in Fig. 5 during the last 9,000 years of our simulation. The times indicated are from the beginning of our simulation. The left arrow marks the mid-point in the transition as seen at the ICB, and the right arrow marks the mid-point as seen at the surface.

crustal magnetic fields in our model, the slopes of our spectra do not become shallower with increasing degree above $l = 12$ (not shown), in contrast with the Earth's[33,34]. Our spectrum taken midway through the transition (open circles) shows that the dipole contribution has decreased more than the other modes during the reversal, as is also apparent in Fig. 1. During our entire simulation, magnetic energy is continually being exchanged in a chaotic manner between the modes, especially back and forth between the dipole ($l = 1$) and quadrupole ($l = 2$) modes on a few-thousand-year timescale. Could the decrease in the Earth's dipole moment over the past 150 years[34] be a similar adjustment?

The reversal sampled at different sites

To compare our simulated reversal with the palaeomagnetic reversal record we now describe how our reversal appears as observed locally at different sampling sites at the surface. We have analysed directional plots (declination and inclination) of our surface field at 50 sites distributed over the surface spanning the last 9,000 years of the simulation. These plots (not shown) indicate that the evolution of the local field direction depends greatly on the sampling location; but they all show that a reversal has occurred. By monitoring the rate of change of our surface field direction at these 50 sites we find the largest rate (at one site, at one time) to be 0.1° per day. But most sites have a maximum rate of ~0.01° per day during the reversal, which is larger than the rate observed in the secular variation of the present-day geomagnetic field[34] but much less than the impulse of 6° per day inferred from the Steens Mountain record[35], the largest rate seen in the palaeomagnetic reversal record.

By knowing the local direction of the surface field at a sampling site (that is, declination and inclination) and assuming the field to be purely dipolar, one can determine the geographical location of the virtual geomagnetic pole (VGP)[3]. In our model we can isolate just the dipolar part of the field (that is, spherical harmonics of degree $l = 1$) and compute the geographical location of the true geomagnetic pole (TGP), a luxury that palaeomagnetists do not enjoy owing to the relatively sparse set of sampling sites and the uncertainty in correlating the ages of samples from the different sites. The TGP path (Fig. 5*a*) during the last 9,000 years of our simulation shows several excursions (or aborted reversals)[36] occurring at different longitudes before and after the successful dipole reversal occurs. The transition itself, seen in our TGP path, takes only ~1,200 years (depending on how one defines the beginning and end of the transition) compared with estimates that range from 1,200 years to the more likely ~5,500 years, seen in the palaeomagnetic reversal record[35,37,38].

We have plotted in Fig. 5*b–d* the VGP paths over the surface during the last 9,000 years of our simulation at three arbitrary sampling sites. Although these VGP paths show that a reversal has occurred, the details are strongly dependent on the sampling site (marked by an asterisk) and do not resemble the TGP path. For example, the long, nearly horizontal, 100-year segment of the VGP path in Fig. 5*d* corresponds to an 'impulse' that is also seen in the directional plot at the same site but is not seen in the other VGP paths or in the TGP path. In addition, our estimate of the transition time would be significantly longer than 1,200 years if we based it only on these VGP paths.

These plots can be compared with VGP transition paths or patches from the palaeomagnetic record, which some argue occur at preferred longitudes[36,39–42] and others interpret in other ways[43–47]. But because the boundary conditions at the ICB and CMB in our present simulation are spherically symmetric, we do not and would not expect to see our excursions and transition occur along preferred longitudes[41,48,49], although we do intend to investigate this intriguing issue with modified versions of our model.

If, in addition to measuring the inclination and declination, one measures the field intensity at a given sampling site, the virtual dipole moment (VDM) of the assumed dipolar field can be computed[3]. In Fig. 6 we plot the VDM computed with the field at our same three sampling sites during the last 9,000 years of our simulation. Also plotted in Fig. 6 is our true dipole moment (TDM). The VDMs provide crude approximations to the TDMs; they correctly reflect the general decrease of the TDM before and during the transition and the recovery after the transition, which is similar to what is seen in palaeointensity records[3]. An asymmetric 'saw-tooth' pattern is also seen in filtered palaeointensity records[50] but the typical time constant of the cycles is ~500,000 years, much longer than the time of our simulation spans. It is interesting that our TDM decreases to nearly its minimum value almost 2,000 years before the mid-point in the transition as seen at the ICB (the left arrow in Fig. 6). It remains low until the mid-point in the transition as seen at the surface (the right arrow in Fig. 6), when it begins its recovery. It is typically between 10×10^{22} and 20×10^{22} A m^2 during most of the simulation, compared with the Earth's present value of $\sim 8 \times 10^{22}$ A m^2. It reaches a minimum of 1.6×10^{22} A m^2 midway through the transition, which is ~10% of its usual value compared with the average drop to 25% seen in plots of the VDM for the Earth[3].

The relatively poor correlations in Figs 5 and 6 between the three sampling sites are due to the fact that the dipole contribution in our simulated magnetic field is not always strongly dominant (Fig. 4). This illustrates how uncertain the details of the VGPs and VDMs could be if the Earth's field were not strongly dipolar during its reversals as some palaeomagnetic reversal records suggest[44,51].

Model limitations and needed improvements

With these initial results from our self-consistent global three-dimensional simulation of the geodynamo we have attempted to bridge the palaeomagnetic and dynamo modelling communities. But with so far only one reversal simulated, our explanations and conclusions about geomagnetic reversals are speculative. Certainly, more analysis is required to improve our understanding of the geodynamo and its reversal mechanism. In addition, several model improvements, which could alter our results and conclusions, will be required for more realistic simulations in the future. For example, more detailed thermodynamics, including compositional convection, is needed. The viscous, thermal and compositional eddy diffusivities should be variable and anisotropic. The effects of topography at the CMB, heterogeneous heat flux and electrical conductivity at the CMB, and luni-solar precession of the mantle could also provide interesting new insights. □

Received 16 June; accepted 23 August 1995.

1. Elsasser, W. M. *Phys. Rev.* **72,** 821–833 (1947).
2. Parker, E. N. *Astrophys. J.* **122,** 293–314 (1955).
3. Merrill, R. T. & McElhinny, M. W. *The Earth's Magnetic Field* (Academic, London, 1983).
4. Cowling, T. G. *Mon. Not. R. astr. Soc.* **94,** 34–48 (1934).
5. Roberts, P. H. & Soward, A. M. *A. Rev. Fluid Mech.* **24,** 459–512 (1992).
6. Braginsky, S. I. & Roberts, P. H. *Geophys. astrophys. Fluid Dyn.* **38,** 327–349 (1987).
7. Olson, P. *Geophys. Res. Lett.* **16,** 613–616 (1989).
8. Bareng hi, C. F. & Jones, C. A. *Geophys. astrophys. Fluid Dyn.* **60,** 211–243 (1991).
9. Hollerbach, R. & Jones, C. A. *Nature* **365,** 541–543 (1993).
10. Glatzmaier, G. A. & Roberts, P. H. *J. Geomagl. Geoelectr.* **45,** 1605–1616 (1993).
11. Nakajima, T. & Roberts, P. H. *Proc. R. Soc. Lond. A* **448,** 1–28 (1995).
12. Pekeris, C. L., Accad., Y. & Shkoller, B. *Phil. Trans. R. Soc. Lond. A* **275,** 425–461 (1973).
13. Kumar, S. & Roberts, P. H. *Proc. R. Soc. Lond. A* **314,** 235–258 (1975).
14. Gubbins, D. & Sarson, G. *Nature* **368,** 51–55 (1994).
15. Zhang, K. K. & Busse, F. H. *Phys. Earth planet. Inter.* **59,** 208–222 (1990).
16. Arter, W. *Geophys. astrophys. Fluid Dyn.* **31,** 311–344 (1985).
17. Matthews, P. C. in *Solar and Planetary Dynamos* (eds Proctor, M. R. E., Matthews, P. C. & Rucklidge, A. M.) 211–218 (Cambridge Univ. Press, 1993).
18. Fearn, D. R., Proctor, M. R. E. & Sellar, C. C. *Geophys. astrophys. Fluid Dyn.* **77,** 111–132 (1994).
19. Olson, P. & Glatzmaier, G. A. *Phys. Earth planet. Inter.* (in the press).
20. Meneguzzi, M. & Pouquet, A. *J. Fluid Mech.* **205,** 297–318 (1989).
21. Brandenburg, A., Nordlund, A., Pulkkinen, P., Stein, R. F. & Tuominen, I. *Astr. Astrophys.* **232,** 277–291 (1990).
22. Nordlund, A. *et al. Astrophys. J.* **392,** 647–652 (1992).
23. St Pierre, M. G. in *Solar and Planetary Dynamos* (eds Proctor, M. R. E., Matthews, P. C. & Rucklidge, A. M.) 295–302 (Cambridge Univ. Press, 1993).
24. Gilman, P. A. & Miller, J. *Astrophys. J. Suppl.* **46,** 211–238 (1981).
25. Gilman, P. A. *Astrophys. J. Suppl.* **53,** 243–268 (1983).

26. Glatzmaier, G. A. *J. comput. Phys.* **55,** 461–484 (1984).
27. Glatzmaier, G. A. *Astrophys. J.* **291,** 300–307 (1985).
28. Glatzmaier, G. A. *Geophys. astrophys. Fluid Dyn.* **31,** 137–150 (1985).
29. Kageyama, A. *et al. Phys. Plasmas* **2,** 1421–1431 (1995).
30. Glatzmaier, G. A. & Roberts, P. H. *Phys. Earth planet. Inter.* **91,** 63–76 (1995).
31. Rikitake, T. *Proc. Camb. phil. Soc.* **54,** 89–105 (1966).
32. Bloxham, J. & Gubbins, D. *Nature* **317,** 777–781 (1985).
33. Cain, J. C., Wang, Z., Schmitz, D. R. & Meyer, J. *Geophys. J.* **97,** 443–447 (1989).
34. Langel, R. in *Geomagnetism* (ed. Jacobs, J. A.) Vol. 1, 249–512 (Academic, San Diego, 1987).
35. Coe, R. S., Prevot, M. & Camps, P. *Nature* **374,** 687–692 (1995).
36. Hoffman, K. A. *Nature* **359,** 789–794 (1992).
37. Opdyke, N. D., Kent, D. V. & Lowrie, W. *Earth planet. Sci. Lett.* **20,** 315–324 (1973).
38. Kristjansson, L. *Geophys. J. R. astr. Soc.* **80,** 57–71 (1985).
39. Tric, E. *et al. Phys. Earth planet. Inter.* **65,** 319–336 (1991).
40. Clement, B. M. *Earth planet. Sci. Lett.* **104,** 48–58 (1991).
41. Laj, C., Mazaud, A., Weeks, R., Fuller, M. & Herrero-Bervera, E. *Nature* **351,** 447 (1991).
42. Ratcliff, C. D., Geissman, J. W., Perry, F. V., Crowe, B. M. & Zeitler, P. K. *Science* **266,** 412–416 (1994).
43. Langereis, C. G., van Hoof, A. A. M. & Rochette, P. *Nature* **358,** 226–230 (1992).
44. Valet, J.-P., Tucholka, P., Courtillot, V. & Meynadier, L. *Nature* **356,** 400–407 (1992).
45. McFadden, P. L., Barton, C. E. & Merrill, R. T. *Nature* **361,** 342–344 (1993).
46. Prevot, M. & Camps, P. *Nature* **366,** 53–57 (1993).
47. McFadden, P. L. & Merrill, R. T. *J. geophys. Res.* **100,** 307–316 (1995).
48. Runcorn, S. K. *Nature* **356,** 654–656 (1992).
49. Clement, B. M. & Stixrude, L. *Earth planet. Sci. Lett.* **130,** 75–85 (1995).
50. Valet, J.-P. & Meynadier, L. *Nature* **366,** 234–238 (1993).
51. Clement, B. M. & Kent, D. V. *Geophys. Res. Lett.* **18,** 81–84 (1991).

ACKNOWLEDGEMENTS. The computing resources for this simulation were provided by the NSF Pittsburgh Supercomputing Center. Support for this research was provided by the Institute of Geophysics and Planetary Physics and the LDRD program at Los Alamos.

ELSEVIER

Physica D 97 (1996) 81–94

PHYSICA D

An anelastic evolutionary geodynamo simulation driven by compositional and thermal convection

Gary A. Glatzmaier [a], Paul H. Roberts [b,*]

[a] *Institute of Geophysics and Planetary Physics, Los Alamos National Laboratory, Los Alamos, NM 87545, USA*
[b] *Institute of Geophysics and Planetary Physics, University of California, Los Angeles, CA 90095, USA*

Abstract

We have extended our geodynamo simulation 40 000 years using a more realistic representation of the thermodynamics and convection. The anelastic approximation replaces the Boussinesq approximation; and compositional buoyancy, in addition to thermal buoyancy, drives convection in the fluid outer core. Boundary conditions at the inner core boundary model the freezing of the heavy constituent onto the solid inner core and the release of the light constituent into the fluid outer core. The resulting simulated magnetic field has a strongly dipole dominated structure outside the core, similar to the Earth's field, so far displaying no tendency to reverse its dipole polarity. The non-dipolar structure of the field at the core–mantle boundary is also quite similar to the Earth's, including the rate of its general westward drift. As in our original simulation, the solid inner core typically rotates about 1°/yr faster than the mantle, in agreement with recent seismic studies of the Earth. This three-dimensional self-consistent solution seems to be simulating the stable regime of the geodynamo between reversals.

1. Introduction

The first fully self-consistent three-dimensional (3D) numerical simulation of the "geodynamo" [1–3], the mechanism in the Earth's core that generates the geomagnetic field, has demonstrated that a magnetic field with an intensity, structure, and time dependence somewhat similar to the Earth's can be maintained by a relatively simple Boussinesq model of thermal convection that prescribes the Earth's mass, dimensions, rotation rate, and as much as possible, realistic material properties and heat flux. In this paper we describe a 40 000-year continuation of the original geodynamo simulation that employs a more realistic representation of the thermodynamics and convection. The Boussinesq approximation is replaced by the anelastic approximation of the magnetohydrodynamic equations; and compositional buoyancy, in addition to thermal buoyancy, drives convection.

The Earth's solid inner core slowly grows as the heavy constituent in the fluid alloy outer core freezes onto it. Verhoogen [4], in 1961, first proposed the release of latent heat in this process to be a source of thermal buoyancy that drives convection in the outer core. In 1963, Braginsky [5] first pointed out that the release of light constituent in this process would provide a source of compositional buoyancy that might be more important in driving convection. Here we investigate the combined effects of these two buoyancy sources in an anelastic simulation of the geodynamo that accounts for the evolution of the core's entropy and composition.

* Corresponding author.

PII S0167-2789(96)00100-5

The anelastic approximation of the fluid dynamic equations [6–11] allows for radially dependent thermodynamic variables (which the Boussinesq approximation considers constant) without requiring prohibitively small numerical time steps that would be needed to resolve acoustic waves in a fully compressible representation. The anelastic approximation is based on the assumption that fluid flow velocities are small compared with acoustic velocities. This is a very good approximation for the Earth's outer core where acoustic velocities are seven orders of magnitude greater than fluid flow velocities. The equations describe 3D time-dependent perturbations relative to a radially dependent reference state. Various versions of the anelastic approximation have been used to simulate convection and magnetic field generation in stars [12–18] and convection in the Earth's mantle [19–22].

2. The anelastic model

Our anelastic model of the geodynamo was developed from an original version of a solar dynamo model [15]. It has been modified in many ways since then, most recently due to the influence of [11].

Here we fit our no-flow non-magnetic reference state of the Earth's fluid, iron-rich, outer core to the *Preliminary Reference Earth Model, PREM* [23], constraining it to be a hydrostatic, adiabatic, well-mixed binary alloy. Therefore, the generalized potential $\overline{\Pi}$ (the sum of specific enthalpy and gravitational potential), the specific entropy, $\overline{S}$, and the mass fraction of light constituent, $\overline{\xi}$, are all independent of radius, r (in addition to being independent of colatitude, θ, and longitude, ϕ). That is, from the definitions of specific enthalpy and gravitational potential, we have

$$\frac{\mathrm{d}\overline{\Pi}}{\mathrm{d}r} = \overline{T}\frac{\mathrm{d}\overline{S}}{\mathrm{d}r} + \bar{\mu}\frac{\mathrm{d}\overline{\xi}}{\mathrm{d}r} + \frac{1}{\bar{\rho}}\frac{\mathrm{d}\bar{p}}{\mathrm{d}r} + \frac{\mathrm{d}\overline{U}}{\mathrm{d}r} = 0$$

because $\mathrm{d}\overline{S}/\mathrm{d}r = 0$ since the reference state is adiabatic, $\mathrm{d}\overline{\xi}/\mathrm{d}r = 0$ since it is well-mixed, and $\mathrm{d}\bar{p}/\mathrm{d}r = -\bar{\rho}\,\mathrm{d}\overline{U}/\mathrm{d}r$ since it is hydrostatic. Here $\overline{T}$, $\bar{\mu}$, $\bar{\rho}$, $\bar{p}$, and $\overline{U}$ are the reference state temperature, chemical potential, density, pressure, and gravitational potential. Although $\overline{\Pi}$, $\overline{S}$, and $\overline{\xi}$ are independent of space (in the outer core), they are time dependent because of the slow secular cooling of the Earth.

The mean radii of the inner core boundary (ICB) and core–mantle boundary (CMB) are set to the Earth's values, $r_{\mathrm{ICB}} = 1.222 \times 10^6$ m and $r_{\mathrm{CMB}} = 3.480 \times 10^6$m, respectively; and the reference state fluid densities at these boundaries are set to PREM values, $\bar{\rho}_{\mathrm{ICB}} = 12.166 \times 10^3$ kg/m^3 and $\bar{\rho}_{\mathrm{CMB}} = 9.903 \times 10^3$ kg/m^3, respectively. We then employ the two-parameter polynomial,

$$\bar{\rho} = c_1(1 - c_2 r^2)$$

to fit these constraints, where the constants, c_1 and c_2, are determined from r_{ICB}, r_{CMB}, $\bar{\rho}_{\mathrm{ICB}}$, and $\bar{\rho}_{\mathrm{CMB}}$. Next we set the mass of the solid inner core (SIC) to the PREM value, $m_{\mathrm{SIC}} = 9.839 \times 10^{22}$ kg, and obtain the reference state gravitational acceleration at the ICB, $\bar{g}_{\mathrm{ICB}} = Gm_{\mathrm{SIC}}/r_{\mathrm{ICB}}{}^2$. The gravitational equation, $\mathrm{d}(r^2\bar{g})/\mathrm{d}r = 4\pi G\bar{\rho}r^2$, is then integrated through the fluid outer core (using the above polynomial fit) to get the gravitational acceleration,

$$\bar{g} = \frac{\mathrm{d}\overline{U}}{\mathrm{d}r} = c_3/r^2 + c_4 r(1 - c_5 r^2),$$

where the constants, c_3, c_4, and c_5, are in terms of r_{ICB}, r_{CMB}, m_{SIC}, c_1, c_2, and the gravitational constant, G. The gravitational potential, $\overline{U}$, in the outer core is obtained by integrating the above equation in radius, using the PREM value $\overline{U}_{\mathrm{ICB}} = 4.402 \times 10^{-2}$ m^2/s^2. The reference state pressure, $\bar{p}$, is then obtained by integrating the hydrostatic equation, $\mathrm{d}\bar{p}/\mathrm{d}r = -\bar{\rho}\bar{g}$, and taking the PREM value $p_{\mathrm{ICB}} = 3.289 \times 10^{11}$ N/m^2. Note that using this procedure, starting with the above two-parameter fit for $\bar{\rho}$, we get values for $\bar{g}$ and $\bar{p}$ at the CMB that agree with PREM to within 0.2% and 0.4%, respectively. Of course we could have obtained better agreement with PREM had we used a higher order polynomial fit for $\bar{\rho}$; but we chose not to because it would make very little difference to our dynamical solution and because many of the other material properties we assume are much more uncertain.

Our reference state (adiabatic) temperature is $\overline{T} = \overline{T}_{\mathrm{ICB}}(\bar{\rho}/\bar{\rho}_{\mathrm{ICB}})^{\gamma}$, where we assume a constant Grüneisen parameter in the fluid outer core, $\gamma = 1.35$ [24], and $\overline{T}_{\mathrm{ICB}} = 5300$ K [11]. We assume the

reference state chemical potential, $\bar{\mu}$, has the same radial dependence as the reference state gravitational potential:

$$\bar{\mu} = c_6 \left(\overline{U} - \overline{U}_{\mathrm{CMB}} \right),$$

where $c_6 = -0.6$ [11] and $\bar{\mu}_{\mathrm{CMB}}$ has been set to zero (see discussion after Eq. (11)).

The following thermodynamic derivatives are needed for the equation of state that gives the density perturbation, ρ, as a function of the perturbations in specific entropy S, pressure p, and mass fraction of light constituent ξ:

$$\left(\frac{\overline{\partial \rho}}{\partial p} \right)_{S,\xi} = -\frac{1}{\bar{\rho} \bar{g}} \frac{\mathrm{d}\bar{\rho}}{\mathrm{d}r},$$

$$\left(\frac{\overline{\partial \rho}}{\partial S} \right)_{p,\xi} = \frac{\bar{\rho}}{\bar{g}} \frac{\mathrm{d}\overline{T}}{\mathrm{d}r},$$

$$\left(\frac{\overline{\partial \rho}}{\partial \xi} \right)_{S,p} = \frac{\bar{\rho}}{\bar{g}} \frac{\mathrm{d}\bar{\mu}}{\mathrm{d}r} = c_6 \bar{\rho}.$$

The following additional thermodynamic derivatives are used to obtain the temperature perturbation for diagnostics:

$$\left(\frac{\overline{\partial T}}{\partial p} \right)_{S,\xi} = -\frac{\gamma \overline{T}}{\bar{\rho}^2 \bar{g}} \frac{\mathrm{d}\bar{\rho}}{\mathrm{d}r},$$

$$\left(\frac{\overline{\partial T}}{\partial S} \right)_{p,\xi} = \frac{\overline{T}}{c_p},$$

$$\left(\frac{\overline{\partial T}}{\partial \xi} \right)_{S,p} = -\frac{h^{\xi}}{c_p}.$$

We assume c_p and h^{ξ} are constants equal to 840 J/(kg K) and -5×10^6 J/kg, respectively [11].

We also specify the radially dependent magnetic diffusivity, $\bar{\eta}$, and turbulent diffusivities of entropy, $\bar{\kappa}^S$, light constituent, $\bar{\kappa}^{\xi}$, and viscosity, $\bar{\nu}$; however, for the solution presented here, we choose them to be constants, without the boundary layer enhancements they had in the original simulation [1,2]. The magnetic diffusivity (in both the inner and outer cores) and the turbulent entropy and light constituent diffusivities are set to 2 m^2/s [11]; the turbulent viscous diffusivity is set to 1450 m^2/s. For computational reasons the Ekman number (i.e., the ratio of the viscous force to Coriolis force) is only 2×10^{-6} in our simulation, which is about 700 times larger than it should be even for a turbulent viscosity in the Earth's outer core. We have chosen to use the correct rotational and magnetic time scales by using the correct values of the boundary radii, r_{ICB} and r_{CMB}, magnetic diffusivity, $\bar{\eta}$, and angular velocity, $\Omega = 7.29 \times 10^{-5}\ \mathrm{s}^{-1}$, for the Earth's fluid core since the two major forces in the outer core are the Coriolis and Lorentz forces. Consequently, the only major discrepancy in our model is the larger than realistic turbulent viscosity. However, as discussed in [1–3], the viscous forces (outside the viscous boundary layers) are already more than five orders of magnitude smaller than the Coriolis and Lorentz forces in our simulation; so the fact that they should be two to three orders of magnitude smaller is likely not important to the dynamics of the solution. This will have to be tested with higher resolution simulations that can cope with smaller viscosity; however, it is encouraging that our present simulation maintains a magnetic field with an intensity, structure, and time dependence similar to the Earth's.

We continue to use hyper-diffusivities (mainly for computational reasons), which damp the smaller scales more than the larger scales. However, compared with our original simulation [1,2] we have reduced the amount of extra damping by a factor of two and by more than a factor of four for the magnetic diffusivity.

Our reference state also includes the conductive heat flux down the adiabat:

$$-c_p \bar{\rho} \bar{\kappa}^T \frac{\partial \overline{T}}{\partial r},$$

where the thermal diffusivity is taken to be

$$\bar{\kappa}^T = c_7 \left(\frac{\overline{T}}{c_p \bar{\rho} \bar{\eta}} \right)$$

as suggested in [11], with $c_7 = 0.02\ \mathrm{kg\,m^3/(K^2\,s^4)}$. This makes $\bar{\kappa}^T$ more than five orders of magnitude smaller than $\bar{\kappa}^S$. However, the superadiabatic temperature gradient,

$$\frac{\overline{T}}{c_p} \frac{\partial S}{\partial r},$$

is roughly five orders of magnitude smaller then the adiabatic gradient; so the turbulent diffusive heat flux is comparable to the conductive heat flux down the adiabat. The divergence of this reference state heat flux down the adiabat provides a cooling term in the entropy equation (7). On the other hand, since temperature perturbations relative to the adiabatic reference state temperature (in our simulation) are only a few times 10^{-4} K, we neglect the small conductive heat flux due to the gradient of these temperature perturbations. We also assume no radioactive heating in the core.

Having established the radially dependent reference state (the "barred" variables), the following anelastic magnetohydrodynamic 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 radially dependent reference state, in a frame of reference rotating at constant angular velocity $\boldsymbol{\Omega}$:

$$\boldsymbol{\nabla} \cdot \bar{\rho}\mathbf{v} = 0, \tag{1}$$

$$\boldsymbol{\nabla} \cdot \mathbf{B} = 0, \tag{2}$$

$$\boldsymbol{\nabla}^2 U = 4\pi G\rho, \tag{3}$$

$$\rho = \left[\left(\frac{\overline{\partial \rho}}{\partial S} \right)_{\xi,p} S + \left(\frac{\overline{\partial \rho}}{\partial \xi} \right)_{S,p} \xi + \left(\frac{\overline{\partial \rho}}{\partial p} \right)_{S,\xi} p \right] = \left[C\bar{\rho} + \left(\frac{\overline{\partial \rho}}{\partial p} \right)_{S,\xi} p \right], \tag{4}$$

$$\bar{\rho}\frac{\partial \mathbf{v}}{\partial t} = -\boldsymbol{\nabla} \cdot (\bar{\rho}\mathbf{v}\mathbf{v}) - \bar{\rho}\boldsymbol{\nabla}(p/\bar{\rho} + U) - C\bar{\rho}\bar{g}\hat{\mathbf{r}} + 2\bar{\rho}\mathbf{v} \times \boldsymbol{\Omega} + \boldsymbol{\nabla} \cdot (2\bar{\rho}\bar{\nu}(\overleftrightarrow{\mathbf{e}} - \tfrac{1}{3}(\boldsymbol{\nabla} \cdot \mathbf{v})\overleftrightarrow{\delta})) + \frac{1}{\mu_0}(\boldsymbol{\nabla} \times \mathbf{B}) \times \mathbf{B}, \tag{5}$$

$$\frac{\partial \mathbf{B}}{\partial t} = \boldsymbol{\nabla} \times (\mathbf{v} \times \mathbf{B}) - \boldsymbol{\nabla} \times (\bar{\eta}\boldsymbol{\nabla} \times \mathbf{B}), \tag{6}$$

$$\bar{\rho}\frac{\partial S}{\partial t} = -\boldsymbol{\nabla} \cdot (\bar{\rho}S\mathbf{v}) + \boldsymbol{\nabla} \cdot (\bar{\rho}\bar{\kappa}^S \boldsymbol{\nabla} S) + \frac{1}{\overline{T}r^2}\frac{\mathrm{d}}{\mathrm{d}r}\left(r^2 c_p \bar{\rho}\bar{\kappa}^T \frac{\mathrm{d}\overline{T}}{\mathrm{d}r} \right) + \frac{\bar{\eta}}{\mu_0 \overline{T}}|\boldsymbol{\nabla} \times \mathbf{B}|^2 + \frac{\bar{g}}{\overline{T}} \times \left[\bar{\kappa}^S \left(\frac{\overline{\partial \rho}}{\partial S} \right)_{\xi,p} \frac{\partial S}{\partial r} + \bar{\kappa}^\xi \left(\frac{\overline{\partial \rho}}{\partial \xi} \right)_{S,p} \frac{\partial \xi}{\partial r} \right] - \bar{\rho}\dot{\bar{S}}, \tag{7}$$

$$\bar{\rho}\frac{\partial \xi}{\partial t} = -\boldsymbol{\nabla} \cdot (\bar{\rho}\xi\mathbf{v}) + \boldsymbol{\nabla} \cdot (\bar{\rho}\bar{\kappa}^\xi \boldsymbol{\nabla}\xi) - \bar{\rho}\dot{\bar{\xi}}. \tag{8}$$

Here, $\overleftrightarrow{\mathbf{e}}$ is the rate of strain tensor and μ_0 is the magnetic permeability of free space. This set of anelastic equations describes (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.

In Eq. (5) we use the formulation obtained in [11] that conveniently combines (without any approximation other then being anelastic) the contributions to buoyancy from the pressure perturbation and the gradient of the gravitational potential perturbation into a gradient of a "reduced pressure", $p/\bar{\rho} + U$. The remaining buoyancy force is due to the "co-density" defined as

$$C\bar{\rho} = \left[\left(\frac{\overline{\partial \rho}}{\partial S} \right)_{\xi,p} S + \left(\frac{\overline{\partial \rho}}{\partial \xi} \right)_{S,p} \xi \right].$$

As described in [11], this formulation clearly illustrates how the buoyancy sources that drive convection are due to the entropy S and light constituent ξ perturbations. It also includes the effects of the gravitational potential perturbation at no extra computational cost since Eq.(3) does not have to be solved unless the actual pressure perturbation, p, is desired.

In our original simulation [1,2] we neglected the inertial term in the momentum equation for the fluid outer core and for the solid body rotation of the inner core and mantle. Here, as in our recent Boussinesq simulation [3], we retain the largest part of the inertial term, i.e., the axisymmetric east–west fluid flow in the outer core and the variable rotation about $\boldsymbol{\Omega}$ of the inner core and mantle. Therefore, as long as we use the correct values of Ω, r_{ICB}, r_{CMB}, and $\bar{\eta}$, which we do,

our solution cannot be arbitrarily scaled as discussed in [1].

If we would include centrifugal acceleration with the reference state gravitational acceleration, surfaces of constant reference state variables would be spheroidal geopotential surfaces. However, since these deviate so little from spherical surfaces in the Earth's core and since the destruction of spherical symmetry of the reference state introduces many complications having little to do with the subject under study, we neglect centrifugal forces in this model. We also neglect the Poincaré force in (5) that would be present if we were accounting for the luni–solar precession of the Earth's mantle.

The heating terms in Eq. (7) also deserve some discussion. Diffusive heat flux due to unresolved turbulence is driven by the entropy gradient; whereas the conductive heat flux down the adiabat is driven by the reference state temperature gradient. Although these two diffusive heat fluxes are comparable in our simulation, the divergence of the turbulent heat flux is almost two orders of magnitude greater than the divergence of conductive heat flux. The divergence of the convective heat flux (due to the resolved large scales) is comparable to the divergence of the turbulent heat flux (due to the parametrized unresolved small scales); they typically have opposite signs at a given location and are largest near the boundaries. Joule heating, the divergence of the conductive heat flux, and the heating due to the terms involving the radial gradients of S and ξ are all roughly three to ten times smaller than the net heating/cooling due to the sum of the divergences of convective and turbulent heat fluxes. We neglect viscous heating because it is even smaller.

The heating term that involves the radial gradients of S and ξ in (7) is proportional to the radial diffusive flux of buoyancy and naturally occurs because of the internal energy balance (see [11], Section 2). A convenient way to see this is to notice that since

$$\left(\frac{\overline{\partial\rho}}{\partial S}\right)_{\xi,p} = \frac{\bar\rho}{\bar g}\frac{\mathrm{d}\overline{T}}{\mathrm{d}r} \tag{9}$$

we have

$$\overline{T}\nabla\cdot\left(\bar\rho\kappa^S\nabla S\right) + \bar g\kappa^S\left(\frac{\overline{\partial\rho}}{\partial S}\right)_{\xi,p}\frac{\partial S}{\partial r} = \nabla\cdot\left(\overline{T}\bar\rho\kappa^S\nabla S\right)$$

(which is the way this has been represented in the heat equation for some models of solar [15] and mantle [21] convection) and that since

$$\left(\frac{\overline{\partial\rho}}{\partial\xi}\right)_{S,p} = \frac{\bar\rho}{\bar g}\frac{\mathrm{d}\bar\mu}{\mathrm{d}r} \tag{10}$$

we have

$$\bar\mu\nabla\cdot\left(\bar\rho\kappa^\xi\nabla\xi\right) + \bar g\kappa^\xi\left(\frac{\overline{\partial\rho}}{\partial\xi}\right)_{S,p}\frac{\partial\xi}{\partial r} = \nabla\cdot\left(\bar\mu\bar\rho\kappa^\xi\nabla\xi\right).$$

These expressions establish that, taken together, the terms are contributions from the net flux of internal energy into a fluid volume, as is required for conservation of total energy in the volume.

The $\dot{\bar S}$ term in Eq. (7) and the $\dot{\bar\xi}$ term in Eq. (8) have no spatial dependencies and represent the gradual decrease with time of the reference state entropy and increase of the reference state light constituent, respectively. Being independent of space, they do not affect the dynamical solution; they serve only to continually subtract spatially independent parts, $\Delta\overline{S}$ and $\Delta\overline{\xi}$, from the evolving perturbations S and ξ. Other than maintaining a reference state close to the mean state, the exact choice of these incremental changes is arbitrary. Instead of using long time averaged values [11], we choose to calculate, at each time step, what they need to be to maintain

$$\int_{\mathrm{FOC}}\bar\rho S\,\mathrm{d}V = 0 \text{ and } \int_{\mathrm{FOC}}\bar\rho\xi\,\mathrm{d}V = 0,$$

where the volume integral is over the fluid outer core (FOC). That is, after updating S via (7) and ξ via (8) without the $\dot{\bar S}$ and $\dot{\bar\xi}$ terms, we compute

$$\Delta\overline{S} = \int_{\mathrm{FOC}}\bar\rho S\,\mathrm{d}V \Big/ \int_{\mathrm{FOC}}\bar\rho\,\mathrm{d}V$$

and

$$\Delta\overline{\xi} = \int_{\text{FOC}} \bar{\rho}\xi dV \Big/ \int_{\text{FOC}} \bar{\rho}\, dV,$$

and subtract these from S and ξ; so effectively $\dot{\overline{S}} = \Delta\overline{S}/\Delta t$ and $\dot{\overline{\xi}} = \Delta\overline{\xi}/\Delta t$, where Δt is the time step. In our simulation, $\dot{\overline{\xi}}$ is typically $5 \times 10^{-20}\,\text{s}^{-1}$ and $\dot{\overline{S}}$ is typically -10^{-16} W/(kg K); the $-\bar{\rho}\dot{\overline{S}}$ heating in Eq.(7) is typically five times smaller than the Joule heating.

To determine $\Delta\overline{\Pi}$, we make use of the easily derived result

$$\Delta\overline{\Pi} = \overline{T}_{\text{CMB}}\Delta\overline{S} + \bar{\mu}_{\text{CMB}}\Delta\overline{\xi}, \tag{11}$$

which supposes that the mantle is not evolving, so that the radius of the CMB, together with $\bar{p}_{\text{CMB}}$ and $\overline{U}_{\text{CMB}}$, are time independent. For simplicity, we may choose the zero energy level of $\bar{\mu}$ to be $\bar{\mu}_{\text{CMB}}$, this choice removing the ambiguity in the zero energy level of the enthalpy and $\overline{\Pi}$. Then (11) determines $\Delta\overline{\Pi}$.

We also assume the spherically symmetric part of the reduced pressure vanishes at the CMB. This provides the one boundary condition required to solve the spherically symmetric part of the momentum equation (5) for the spherically symmetric part of the reduced pressure throughout the outer core. Although it is convenient for the diagnostics to compute this time dependent radial profile, it does not feedback onto the dynamical solution.

The changes, $\Delta\overline{\Pi}$, $\Delta\overline{S}$ and $\Delta\overline{\xi}$, could be used to update the reference state, including all "barred" variables; at the same time, since $\bar{\mu}$ depends on $\overline{S}$ and $\overline{\xi}$, $\bar{\mu}_{\text{CMB}}$ would evolve from zero. However, for our simulations, which span only tens of thousands of years, these small changes would have little effect on the dynamics.

We impose impermeable non-slip boundary conditions at the ICB and CMB and allow the SIC and solid mantle to rotate about the z-axis relative to the rotating frame of reference (where $\boldsymbol{\Omega} = \Omega\hat{\mathbf{z}}$). These rotation rates are updated each time step using the net torques on these boundaries and the moments of inertia of the inner core, 5.86×10^{34} kg m^2, and mantle, 7.12×10^{37} kg m^2. The total angular momentum of the inner core, outer core, and mantle is constrained to be zero in the rotating frame.

The magnetic field is solved throughout the inner and outer cores and the field is continuous at the ICB, providing magnetic torque between the inner and outer cores. An infinitely thin, spherically symmetric, electrically conducting layer above the CMB with total inductance of 4×10^6 S [25] provides magnetic coupling between the fluid core and the mantle. The rest of the mantle is considered an electrical insulator; so the external field is a source free potential field.

A major improvement to our geodynamo model is the addition of compositional convection. It is generally accepted that, as the Earth cools, the heavy constituent (mainly iron) in the fluid alloy freezes onto the inner core and, in the process, releases latent heat and the light constituent (probably oxygen, silicon, sulfur), providing thermal and compositional buoyancy sources, respectively [5,11]. In our model we assume the latent heat released at the ICB all diffuses into the fluid core since the turbulent diffusivity in the fluid is several orders of magnitude greater than the conductive diffusivity of the SIC. We simulate this by imposing two boundary conditions at the ICB. According to [11], the growth of the local inner core radius, $R(\theta, \phi, t)$, is proportional to the local turbulent (radial) fluxes of entropy and light constituent at the ICB:

$$\rho_{\text{SIC}}\Delta S\frac{\partial R}{\partial t} = -\left(\bar{\rho}\bar{\kappa}^S\frac{\partial S}{\partial r}\right)_{\text{ICB}}, \tag{12}$$

$$\rho_{\text{SIC}}\Delta\xi\frac{\partial R}{\partial t} = -\left(\bar{\rho}\bar{\kappa}^\xi\frac{\partial \xi}{\partial r}\right)_{\text{ICB}}, \tag{13}$$

where $\rho_{\text{SIC}} = 12.764 \times 10^3$ kg/m^3 is the density of the SIC at the ICB and r_{ICB} is the spherically symmetric part of R. According to [11], $\Delta\xi$, the mass of light constituent released when unit mass of alloy freezes at the ICB, is within 10% of 0.065, the value we adopt, while the corresponding entropy release, $\Delta S = 190$ J/(kg K) is less well determined, being possibly 50% or more in error. These uncertainties are hardly surprising when it is recalled that even the composition of the fluid core is unknown. The radial flux of buoyancy at the ICB,

$$-\left[\left(\frac{\overline{\partial\rho}}{\partial S}\right)_{\xi,p}\bar{\kappa}^S\frac{\partial S}{\partial r}+\left(\frac{\overline{\partial\rho}}{\partial\xi}\right)_{S,p}\bar{\kappa}^\xi\frac{\partial\xi}{\partial r}\right],$$

therefore depends on location and time; however, based on (12) and (13) and on our reference state model specifications so far, the ratio of the compositional to thermal buoyancy fluxes is 3.0 everywhere on the ICB, at all times. Note this number is not a result of our 3D solution but a result of our specification of the reference state and material properties; it is in rough agreement, considering how uncertain many of the material properties are, with the estimate of 4 obtained in another analysis [26]. Conditions (12) and (13), are combined, via the elimination of $\partial R/\partial t$, to give us one of our boundary conditions at the ICB, which is imposed on each spherical harmonic component of S and ξ.

The other condition is an expression of the fact that the ICB is perpetually at the freezing point. This leads to an equation that expresses the time rate of change of R as a linear combination of the time rates of change of p, S and ξ at the ICB. It is argued in [11] that the variation of p_{ICB} has negligible effect on $\partial R/\partial t$, and our numerical work confirms this; the contributions made by $\partial p/\partial t$ are at least two orders of magnitude smaller than those made by $\partial S/\partial t$ and $\partial\xi/\partial t$. The condition then reduces to a relation between the time rates of change of R, S and ξ:

$$\frac{1}{r_{\mathrm{ICB}}}\frac{\partial R}{\partial t}=-\frac{1}{0.03c_p}\frac{\partial S}{\partial t}-70\frac{\partial\xi}{\partial t}, \tag{14}$$

where the coefficients 0.03 and 70 come from an analysis of the freezing process and, due to the unknown chemistry of the alloy, are relatively uncertain. Except for the slow spherically-symmetric evolution of the reference state, the time derivatives of S and ξ are determined by (7) and (8). The time rate of change of $\overline{\xi}$ is simply related to that of r_{ICB} through the flux of light constituent released at the ICB during freezing [11]:

$$\dot{\overline{\xi}}=9.527\times10^{-3}\frac{\dot{r}_{\mathrm{ICB}}}{r_{\mathrm{ICB}}}.$$

Consequently, the spherically symmetric part of (14) obeys

$$\frac{1}{r_{\mathrm{ICB}}}\frac{\partial R}{\partial t}=-\frac{1}{0.05c_p}\frac{\partial S}{\partial t}. \tag{15}$$

It is suggested in [11] that (15) can be used in place of (14) for the complete S and R, and not merely for their spherically symmetric parts. We have indeed found that (15) does not result in significant differences from (14) in our simulations. We have nevertheless preferred to use (14) for the asymmetric parts, which we believe to be more precise than (15).

We assume the CMB is impermeable and stationary and therefore force the radial flux of light constituent to vanish at the CMB by setting the radial gradient of ξ to zero there. Heat, on the other hand, flows from the core to the mantle. We specify the spherically symmetric heat flow at the CMB to be 7.2×10^{12} W, which is 23% of the estimated heat loss from the Earth's mantle (excluding crustal heat loss) and is more than twice as large as estimates [24] of this poorly known quantity. This total radial heat flow at the CMB (on the fluid side) is the sum of the conductive heat flow down the adiabat and the turbulent diffusive heat flow:

$$-4\pi r_{\mathrm{CMB}}{}^2\left(c_p\bar{\rho}\bar{\kappa}^T\frac{\mathrm{d}\overline{T}}{\mathrm{d}r}+\overline{T}\bar{\rho}\bar{\kappa}^S\frac{\partial S}{\partial r}\right)_{\mathrm{CMB}}$$

Based on our specification of the reference state, the flow down the adiabat (the first term) is 4.96×10^{12} W, which accounts for two-third of the heat flow at the CMB. The remaining one-third is due to the superadiabatic temperature gradient (i.e., the entropy gradient), which also drives large-scale thermal convection.

Whereas our boundary conditions on S and ξ at the ICB result in a time dependent, heterogeneous heat flux at the ICB that responds to the dynamic solution in the FOC, our heat flux boundary condition at the CMB serves as a "valve" controlled by the lower mantle that determines the strength of the convection (both thermal and compositional) in the FOC which, together with the rotation rate, ultimately determines the strength of the generated magnetic field. The structure of the convection and field could also be influenced by prescribing a heterogeneous heat flux boundary condition at the CMB [27]; however, for the solution presented here, we have prescribed a homogeneous radial heat flux at the CMB.

The numerical method of solving this system of equations has been briefly described in [1]. One additional complication is that the prognostic equations for S (7) and ξ (8) are solved simultaneously since there is now the additional heating term in (7) that involves ξ and since the boundary conditions at the ICB couple S and ξ. However, these (linear) terms and boundary conditions are decoupled in spherical harmonic degree and order; so a separate matrix equation is solved in radius for each spherical harmonic coefficient of the combined S–ξ vector.

We also now use a third-order accurate Runge–Kutta time integration scheme [28] that treats the linear terms implicitly and the nonlinear terms explicitly. This new scheme is also used for updating the rotation rates of the SIC and mantle and for the prognostic part of the boundary conditions on S and ξ at ICB. Although each computational time step now requires three times as much cpu time, the scheme is more accurate and stable and allows longer time steps.

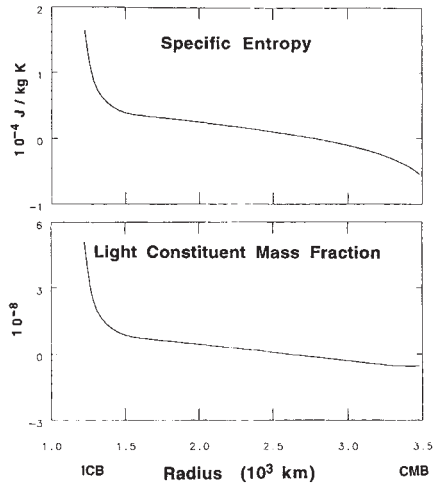


Fig. 1. A snapshot (at 38 000 years into the new simulation) of the radial profile of the symmetric parts of the perturbations (relative to the reference state) of the specific entropy, S, and light constituent mass fraction, ξ.

3. The geodynamo solution

We used our Boussinesq geodynamo solution at 50 000 years into the simulation [3] as the initial condition for the anelastic solution presented here, which spans 40 000 years. To provide a smooth transition, we set the initial 3D structure of the light constituent, ξ, to be essentially the same as that of the entropy, S, at that time. The amplitudes of S and ξ where also adjusted to satisfy the new boundary conditions. As the solution evolved, the structures of S and ξ diverged somewhat due to differences in their equations and boundary conditions. However, since the fluxes of entropy and light constituent at the ICB are locally always in the same proportion, are advected by the same flow, and have the same turbulent diffusivities, the 3D structure of the entropy and light constituent do remain fairly similar, especially near the ICB.

As in our original simulations [1–3], the largest spatial gradients in the axisymmetric part of the entropy, occur near the ICB and inside the imaginary cylinder tangent to the SIC and co-axial with Ω; the region outside this "tangent cylinder" has a somewhat weaker entropy maximum in the equatorial region. The tangent cylinder divides the region outside it where asymmetric Taylor columns develop with axes parallel to Ω from the polar regions inside it where nearly axisymmetric Taylor columns develop with axes co-axial with Ω. This tangent cylinder effect occurs when the viscous force is small compared with the Coriolis force and the Lorentz force is comparable to the Coriolis force [29].

The light constituent mass fraction, ξ, has a similar spatial structure, but with no radial gradient at the CMB. Fig. 1 shows a snap shot of the radial profiles of the spherically symmetric parts of S and ξ. These profiles illustrate how most of the convective driving occurs near the ICB where the radial gradients of S and ξ are most steep; the gradients are relatively small throughout the bulk of the fluid core. Snapshots of S and ξ in a spherical surface at mid-depth in the fluid core (Figs. 2(a) and (b)) illustrate the warmer (high S) and lighter (high ξ) regions near the equator and inside the tangent cylinder, i.e., at high latitude. The

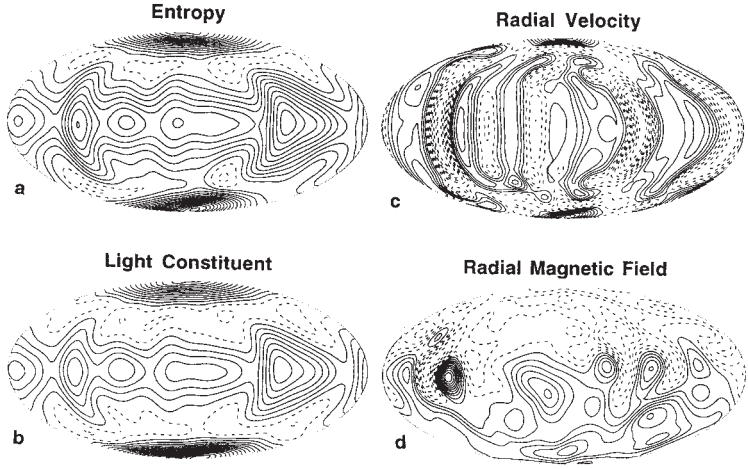


Fig. 2. A snapshot (at the same time as in Fig. 1) of the (a) specific entropy perturbation, (b) light constituent mass fraction perturbation, (c) radial component of the velocity, and (d) radial component of the magnetic field. (a), (b), and (c) are in the spherical surface at mid-depth in the outer core; (d) is at the CMB. For each of these equal area projections, the north (south) geographic pole is at the top (bottom) of the projection, with the equator being a horizontal line through the middle. Solid contours represent positive perturbations relative to the reference state; broken contours represent negative perturbations. The maximum and minimum contours in (a) are 9.0×10^{-5} and -1.2×10^{-5} J/(kg K), in (b) are 2.8×10^{-8} and -0.4×10^{-8}, in (c) are 6.3×10^{-4} and -5.6×10^{-4} m/s, and in (d) are 0.63 and −0.42 mT.

similarity in the structures is quite obvious in these surfaces as it is in the equatorial plane (Figs. 3(a) and (b)) where it is seen how the warm, light fluid is advected outward from the ICB by the flow (Fig. 3(c)). Also notice the steep radial gradients at the ICB in Figs. 3(a) and (b).

There is a tendency for the radial gradient of ξ to become slightly positive at times, i.e., compositionally stable, near the CMB; however, since upward convective heat flux is converted to diffusive heat flux near the CMB, the radial gradient of S there is negative, i.e., thermally unstable, and limits the stable stratification of the light constituent. That is, as the compositional gradient becomes more stable due to the accumulation of light constituent at the top of the fluid core, convective heat flux decreases there and diffusive heat flux increases as the temperature gradient steepens; but the steeper temperature gradient then drives more thermal convection which reduces the stable stratification of the light constituent.

The axisymmetric parts of the fluid flow are also similar to those in our original simulations with eastward flow (relative to the rotating frame of reference) inside the tangent cylinder near the ICB and westward inside the tangent cylinder near the CMB; weak westward flow that is nearly in solid body rotation exists outside the tangent cylinder, as can be seen in Fig. 3(c). The SIC always rotates eastward (relative to the rotating frame), as can be seen in Fig. 3(c), with a typical angular velocity of 1°/yr ($\approx 10^{-5}\,\Omega$); the mantle also usually rotates eastward but with a typical angular velocity of 0.01°/yr ($\approx 10^{-7}\,\Omega$). As in our original simulation [1], the rotation rate of the inner core can be at times as high as 1.8°/yr, in fairly good agreement with recent seismic studies of the Earth [30], which estimate the present rotation rate of the Earth's inner core to be 1.5°/yr faster than the mantle.

Our meridional circulation typically has its greatest amplitude inside the tangent cylinder where it is usually directed outward, parallel to the rotation axis.

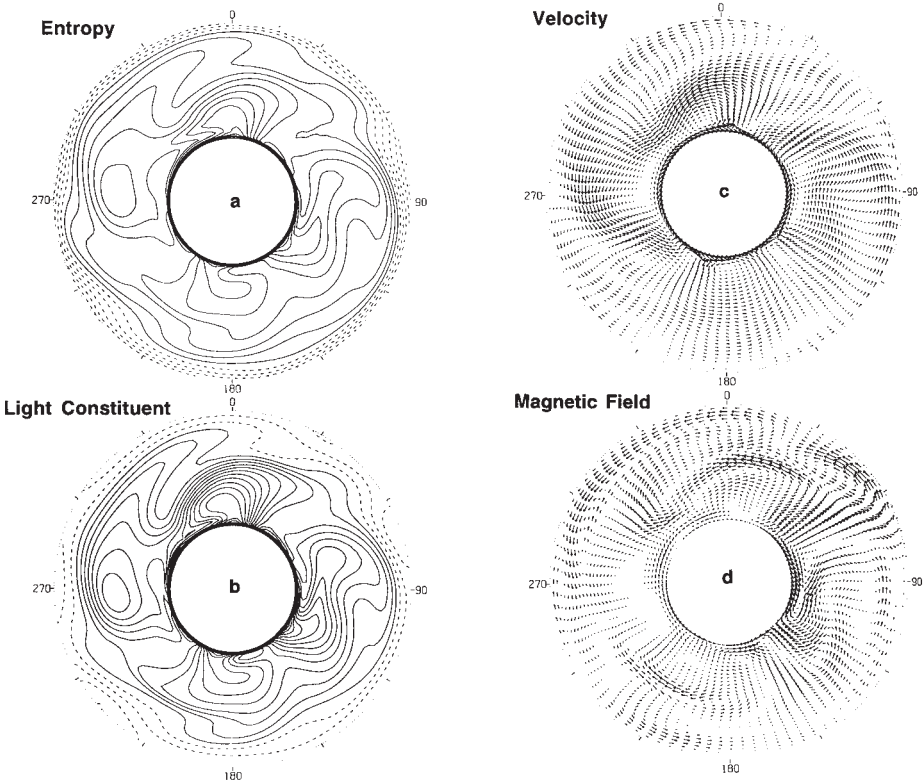


Fig. 3. A snapshot (at the same time as in Fig. 1) of the (a) specific entropy perturbation, (b) light constituent mass fraction perturbation, (c) velocity, and (d) magnetic field in the equatorial plane. The outer circular boundary is the CMB and the inner one is the ICB. These are viewed in the direction of Ω, i.e., from south to north, so the direction of (the eastward) rotation is clockwise. Fields are plotted only in the outer fluid core. Solid (broken) contours represent positive (negative) perturbations relative to the reference state. The maximum and minimum contours in (a) are 12.0×10^{-5} and -6.0×10^{-5} J/(kg K) and in (b) are 3.8×10^{-8} and -0.8×10^{-8}. The maximum vectors (in the equatorial plane) in (c) are 16.9×10^{-4} m/s and in (d) are 6.19 mT.

This is evident in Fig. 2(c) which shows a snapshot of the radial component of the flow in a spherical surface at mid-depth in the fluid core. Weak Taylor column structure is also evident outside the tangent cylinder, i.e., at low latitude in Fig. 2(c). The snapshot of the flow in the equatorial plane (Fig. 3(c)) shows the cross-sections of the Taylor columns and how they are drawn out in longitude in order to minimize flow across the strong east–west component of the magnetic field (Fig. 3(d)) [29]. Figs. 3(c) and (d) illustrate how the flow and the field tend to be aligned locally.

Like our recent Boussinesq simulation of the geodynamo [3], but unlike our first Boussinesq simulation [1,2], the poloidal (meridional) magnetic field in this anelastic simulation has one dipole polarity throughout the inner and outer cores. The axisymmetric part

of the toroidal (east–west) field is anti-symmetric with respect to the equator and, for a given hemisphere, is usually oppositely directed inside and outside the tangent cylinder. Also, as in our more recent Boussinesq simulation, the field has a strong axial dipole component, like the Earth's present field. This is seen in the large $l = 1$ contribution to the mean square magnetic field intensity at the surface (6371 km radius) shown in Fig. 4. This is also very apparent in the snapshot of the radial component of our simulated magnetic field plotted in the spherical surface at the CMB in Fig. 2(d). Several test simulations have demonstrated that the less dominant dipolar structure of our first simulation was due to an enhanced viscosity that was specified in the boundary layer below the CMB. Without this viscous enhancement, the fields generated in our recent Boussinesq simulation [3] and in this anelastic simulation are more strongly dipolar (i.e., more like the Earth's field) and, so far, show no tendency for a magnetic dipole reversal as occurred in [1,2].

As described in [3], the tangent cylinder and ICB play important roles in the maintenance of the strong dipole component of our simulated magnetic field. Helical fluid flow, mainly inside the tangent cylinder and near the ICB [1], twists toroidal field into poloidal field; and, as seen in the snapshot (Fig. 5) of the 3D field structure, poloidal field lines that penetrate the tangent cylinder are stretched into an east–west orientation by the shear of the zonal (east–west) flow on and inside the tangent cylinder, producing large toroidal field wound around the tangent cylinder. This field is expelled as poloidal field where the tangent cylinder intersects the CMB, i.e., in the polar regions, providing a large supply of outward directed field in one polar region and inward directed field in the other, thus maintaining an external potential field having a strong dipolar structure with an axis nearly aligned with the rotation axis.

As seen in Fig. 4, the field intensity at the surface of our anelastic simulation is about three times smaller than the Earth's was in 1980 [31] even though our prescribed heat flow at the CMB may be too large by a factor of two. It is, however, still in the range of the Earth's palaeointensity record [32]. It is also possible that the Poincaré force due to the luni–solar precession of the mantle (which we have not yet included) could provide additional fluid dynamical driving [33,34] that could increase the intensity of the generated magnetic field. As in our Boussinesq simulations, the generated magnetic energy is typically three orders of magnitude greater than the convective kinetic energy (in the rotating frame) that maintains it, which is expected for the Earth's strong-field dynamo. The spectrum is somewhat time dependent; however, as in our recent Boussinesq simulation [3], the intensity of the dipole ($l = 1$) is the least time dependent. The maximum flow (relative to the rotating frame) is typically 2×10^{-3} m/s and the maximum field is typically 20 mT.

The intensity and structure of the non-dipolar part of our magnetic field at the CMB (Fig. 2(d)) are Earth-like [35]. A movie of the radial component of our field at the CMB also shows a "general" westward drift with an Earth-like longitudinal phase velocity of 0.2°/yr (relative to the rotating frame), which, like the Earth's, is most evident in the equatorial region where, at mid-depth in our simulated fluid core, the westward zonal flow peaks at typically 0.4°/yr. As mentioned above, our mantle typically rotates eastward (relative to the rotating frame) at about 0.01°/yr; so, as observed on the surface of our modeled Earth, the westward drift velocity would be slightly larger. Like the Earth's CMB field, not all features participate in the westward drift and the field patterns and phase velocities are time dependent as are the structures of the flow and the field within the fluid core.

4. Conclusions

We have extended our geodynamo simulation 40 000 years using a more realistic representation of the thermodynamics and convection by replacing the Boussinesq approximation with the anelastic approximation and by including compositional convection in addition to thermal convection. The resulting generated magnetic field has a strongly dipole dominated structure outside the core similar to (albeit a little weaker in strength than) the Earth's field; and there has been no tendency for a dipole polarity reversal during this 40 000 years. The non-dipolar structure

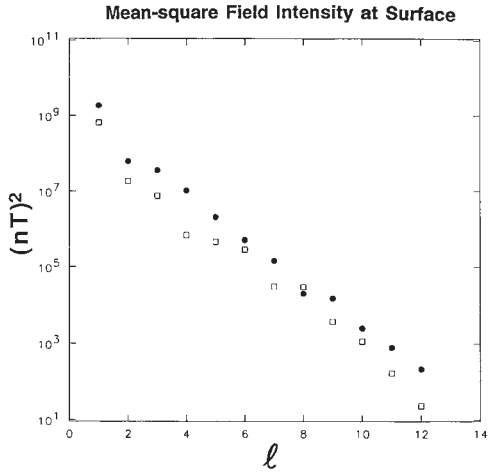


Fig. 4. The mean square magnetic field intensity as measured at the surface, i.e., at radius 6371 km, as a function of spherical harmonic degree, l. Open squares represent our simulated field (at the same time as in Fig. 1); solid circles represent the actual geomagnetic field in 1980 [33].

at the CMB is also quite similar to the Earth's, as is its general westward drift velocity. Recent seismic studies of the Earth [30] also provide preliminary confirmation of our prediction [1] that the solid inner core rotates about 1°/yr faster than the mantle. Our present 3D self-consistent solution seems to be simulating the stable regime of the geodynamo between reversals.

Our time dependent heterogeneous boundary conditions on the entropy and light constituent (of our binary fluid alloy) at the ICB model the freezing of the heavy constituent onto the SIC, which is determined by the dynamical solution in the fluid outer core. Consequently, our solution is controlled by our prescription (7.2×10^{12} W) of the Earth's (poorly known) total radial heat flow at the CMB, which could be too large by a factor of two. Two-third of this is due to thermal conduction down the adiabat. The remaining one-third requires a superadiabatic temperature gradient at the CMB, which drives convection and so limits the degree of a stable compositional gradient there.

If we would instead prescribe the total radial heat flow at the CMB to be less than the conductive heat flow down the adiabat, the region just below the CMB would probably be both thermally and compositionally stable, i.e., the radial gradients of entropy and light constituent would be positive near the CMB. If the region near the CMB were strongly stable, turbulence there would be much weaker because it would be generated non-locally, below the stable region. A thorough investigation of this would require a more sophisticated, spatially and temporally dependent, turbulent diffusivity that would respond to local (and non-local) conditions. It should also be anisotropic, depending on the directions of Ω and the local magnetic field.

Our boundary conditions at the ICB on entropy and light constituent allow us to analyze the rate of growth of the inner core, including its time dependent topography; we will describe these results in a future paper. Future simulations will also need to test the effects of a heterogeneous heat flux and electrical conductivity at the CMB and the effects of luni–solar precession of the mantle.

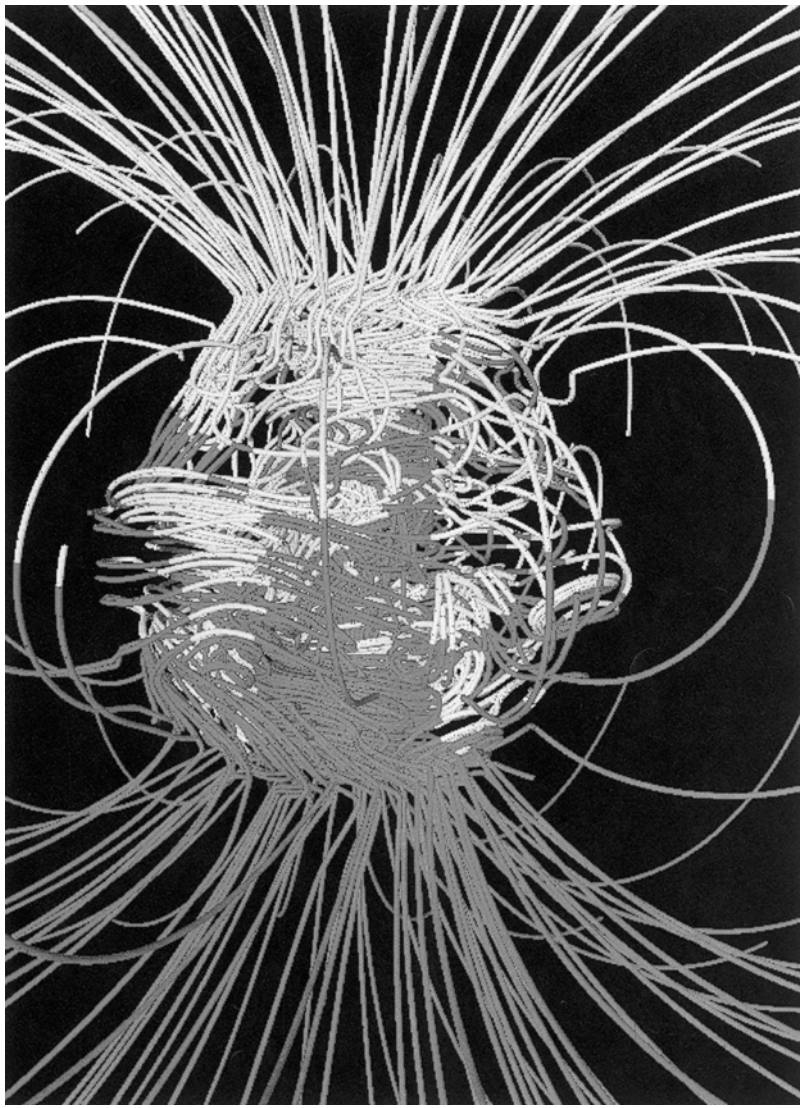


Fig. 5. A snapshot (at the same time as in Fig. 1) of the 3D field portrayed via lines of force that are plotted out to two Earth radii. Gold (blue) lines represent outward (inward) directed magnetic field. The rotation axis is vertical. The transition at the CMB from the smooth, external potential field to the more intense, complicated field inside the core is quite striking. The maximum field intensity is 29 mT.

Acknowledgements

The computing resources that were basic to the research reported here were provided by the NSF Pittsburgh Supercomputing Center under grant MCA94P016P. GAG was supported by Los Alamos LDRD Grant 96149 and PHR by NSF Grant EAR94-06002. This work was conducted under the auspices of the US Department of Energy, supported (in part) by funds provided by the University of California for the conduct of discretionary research by Los Alamos National Laboratory. We dedicate this paper to Fritz Busse in recognition of his long standing interest in the generation of magnetic fields by the Earth and other planets, and in celebration of his 60th birthday.

References

[1] G.A. Glatzmaier and P.H. Roberts, Phys. Earth Planet. Inter. 91 (1995) 63.
[2] G.A. Glatzmaier and P.H. Roberts, Nature 377 (1995) 203.
[3] G.A. Glatzmaier and P.H. Roberts, Phys. Earth Planet. Inter. (1996) in press.
[4] J Verhoogen, Geophys. J. 4 (1961) 276.
[5] S.I. Braginsky, Soviet Phys. Dokl. 149 (1963) 8.
[6] Y. Ogura and N.A. Phillips, J. Atmospheric Sci. 19 (1962) 173.
[7] S.I. Braginsky, Geomag. Aeron. 4 (1964) 698.
[8] D.O. Gough, J. Atmospheric Sci. 26 (1969) 448.
[9] J. Latour, E.A. Spiegel, J. Toomre and J.-P. Zahn, Ap. J. 207 (1976) 233.
[10] G.A. Gilman and G.A. Glatzmaier, Ap. J. Suppl. 45 (1981) 335.
[11] S.I. Braginsky and P.H. Roberts, Geophys. & Astrophys. Fluid Dynam. 79 (1995) 1.
[12] R. Van der Borght, Mon. Not. Roy. Astr. Soc. 173 (1975) 85.
[13] J. Toomre, J.-P. Zahn, J. Latour and E.A. Spiegel, Ap. J. 207 (1976) 545.
[14] A. Nordlund, Astron. Astrophys. 107 (1982) 1.
[15] G.A. Glatzmaier, J. Comp. Phys. 55 (1984) 461.
[16] P.A. Gilman and J. Miller, Ap. J. Suppl. 61 (1986) 585.
[17] G.P. Ginet and R.N. Sudan, Phys. Fluids 30 (1987) 1667.
[18] C.A. Jones and P.H. Roberts, Geophys. & Astrophys. Fluid Dynam. 55 (1990) 263.
[19] G.T. Jarvis and D.P. McKenzie, J. Fluid Mech. 96 (1980) 515.
[20] F. Quareni and D.A. Yuen, in: Mathematical Geophysics, eds. N.J. Vlaar et al., Ch. 11 (1987).
[21] G.A. Glatzmaier, Geophys. & Astrophys. Fluid Dynam. 43 (1988) 223.
[22] H.-P. Bunge, M.A. Richards and J.R. Baumgardner, J. Geophys. Res. (1996) submitted.
[23] A.M. Dziewonski and D.L. Anderson, Phys. Earth Planet. Inter. 25 (1981) 297.
[24] F. Stacey, Physics of the Earth, 3rd Ed. (Brookfield, Brisbane, 1992).
[25] T.J. Shankland, J. Peyronneau and J.-P. Poirier, Nature 366 (1993) 453.
[26] J.R. Lister and B.A. Buffett, Phys. Earth Planet. Inter. 91 (1995) 17.
[27] P. Olson and G.A. Glatzmaier, Philos. Trans. Roy Soc. London Ser. A 354 (1996) 1.
[28] P.R. Spalart, R.D. Moser and M.M. Rogers, J. Comp. Phys. 96 (1991) 297.
[29] P. Olson and G.A. Glatzmaier, Phys. Earth Planet. Inter. 92 (1995) 109.
[30] X. Song and P.G. Richards, Nature (18 July 1996).
[31] R. Langel, in: Geomagnetism, ed. J.A. Jacobs, Vol. 1 (Academic Press, San Diego, 1987) pp. 249–512.
[32] J.P. Valet and L. Meynadier, Nature 366 (1993) 234.
[33] W.V.R. Malkus, Science 160 (1968) 259.
[34] J.P. Vanyo, Geophys. & Astrophys. Fluid Dynam. 59 (1991) 209.
[35] J. Bloxham and D. Gubbins, Nature 317 (1985) 777.

Rotation and Magnetism of Earth's Inner Core

Gary A. Glatzmaier* and Paul H. Roberts

Three-dimensional numerical simulations of the geodynamo suggest that a super-rotation of Earth's solid inner core relative to the mantle is maintained by magnetic coupling between the inner core and an eastward thermal wind in the fluid outer core. This mechanism, which is analogous to a synchronous motor, also plays a fundamental role in the generation of Earth's magnetic field.

Three-dimensional (3D) numerical simulations of the geodynamo, the mechanism in Earth's core that generates the geomagnetic field, showed that a magnetic field with an intensity, structure, and time dependence similar to that of Earth's can be maintained by a convective model (*1*, *2*). The convection, which takes place in the fluid outer core surrounding the solid inner core, twists and shears magnetic field, continually generating new magnetic field to replace that which diffuses away. The model we describe here (*2*) assumes the mass, dimensions, and basic rotation rate of Earth's core, an estimate of the heat flow out of the core, and, as far as possible, realistic material properties (*3*).

Convection is driven by thermal and compositional buoyancy sources that develop at the inner core boundary as the Earth cools and iron alloy solidifies onto the inner core (*4*). For simplicity, we modeled the core as a binary alloy (*5*). Our boundary conditions at the inner core boundary constrain the local fluxes of the latent heat and

G. A. Glatzmaier, Institute of Geophysics and Planetary Physics, Los Alamos National Laboratory, Los Alamos, NM 87545, USA. E-mail: gag@lanl.gov
P. H. Roberts, Institute of Geophysics and Planetary Physics, University of California, Los Angeles, CA 90095, USA. E-mail: roberts@math.ucla.edu

*To whom correspondence should be addressed.

the light constituent to be proportional to each other and to the local cooling rate. The flux of the light constituent provides three times as much buoyancy as the latent heat flux. However, the vigor of the convection, the strength of the magnetic field, and the angular velocity of the inner core are all ultimately determined by the thermal boundary condition at the core-mantle boundary. We prescribed a heat flow through it of 7.2 TW, which is geophysically realistic (6), 5.0 TW of this being the heat conducted down the adiabatic part of the temperature gradient.

The inner core in our model is free to rotate about the geographic axis in response to the magnetic and viscous torques to which the outer core subjects it (7). We found that the inner core rotates relative to the mantle in the eastward direction, at an angular velocity of $\Omega_{IC} \sim 2°$ to 3° per year (8). Our prediction of eastward inner core rotation is consistent with two subsequent, independent analyses of seismic data: Song and Richards (9) first estimated $\Omega_{IC} \sim 1°$ per year, and then Su *et al.* (10) estimated $\Omega_{IC} \sim 3°$ per year. Considering the uncertainties in our model and material properties and in the seismic data and inner core anisotropy, these preliminary theoretical and observational results are in reasonable agreement.

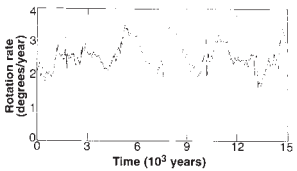


Fig. 1. Eastward angular velocity of the inner core relative to the mantle for the last 15,000 years of our simulation.

For a typical 15,000-year interval, our mean Ω_{IC} is $\overline{\Omega}_{IC} \approx 2.6°$ per year (Fig. 1). The standard deviation about that mean is 0.4°/year, the root mean square of Ω_{IC} is $\Omega_{IC}^{rms} = 0.005°$ per year^2, and the time scale of variations of Ω_{IC} is $\tau_{IC} = \overline{\Omega}_{IC}/\dot{\Omega}_{IC}^{rms} \approx 500$ years. However, there is also a distinct 3000-year period with an amplitude variation of about a ±0.5° per year (Fig. 1).

To explain the sense of rotation of the inner core relative to the mantle, it is helpful to recognize that the inner core plays a central role in the dynamics of the outer core (11). Consider an imaginary circular cylinder coaxial with the rotation axis and touching the inner core at its equator. This "tangent cylinder" divides the outer core into an interior region I and an exterior region E. The interior region consists of a northern polar region I_N and a southern polar region I_S.

Convection in a rapidly rotating body of fluid, like Earth's outer core, takes the form of Taylor columns, which have axes parallel to the rotation axis (12) and fluid flowing in planes perpendicular to it. However, when the fluid flow generates a strong magnetic field that in turn feeds back onto the flow by means of Lorentz forces, the flow structure is somewhat different. In our simulation (2), several nonaxisymmetric magnetically suppressed Taylor columns exist in E, whereas in I, although the flow is also mainly in planes perpendicular to the rotation axis, it is nearly axisymmetric and sheared parallel to the rotation axis. Consequently, heat and light elements are more efficiently transported from the inner core boundary to the core-mantle boundary by the nonaxisymmetric flows in E than by the nearly axisymmetric flow in I, which takes a more indirect, helical route to the core-mantle boundary. As a result, in our simulation (2) the axisymmetric parts of the perturbations of the entropy and light constituent (relative to their constant basic-state values) at a given radius near the inner core boundary are roughly five times greater in I than their respective values in E. The associated buoyancy force in I produces an outward flow along the rotation axis, which (because of mass conservation) requires a flow directed toward the rotation axis near the inner core boundary and away from the axis near the core-mantle boundary. This axisymmetric part of the flow is called the meridional circulation (Fig. 2A). Because the angular momentum of fluid parcels is approximately conserved, fluid flowing toward the rotation axis just outside of the inner core boundary, in I_N and I_S, is spun up; fluid flowing away from the axis near the core-mantle boundary is spun down. The resulting zonal flow in I is eastward near the inner core and westward near the mantle, relative to the fluid in E and to the mantle (Fig. 2B). In atmospheric contexts, these

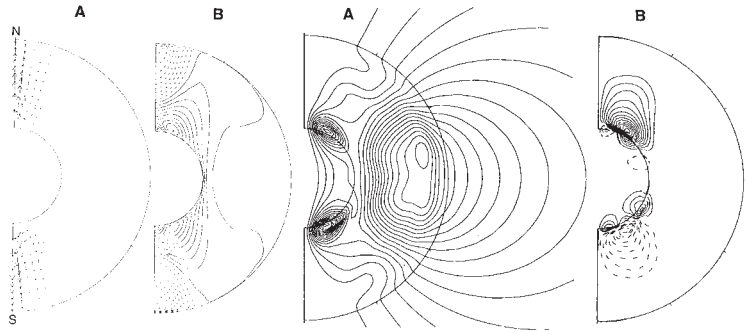


Fig. 2 (left). Snapshots in the meridian plane of (**A**) the axisymmetric meridional fluid flow and (**B**) the axisymmetric zonal fluid flow. The direction of the arrows in (A) are parallel to the flow; their length is proportional to the flow speed. Contours of constant angular velocity ω are shown in (B); solid lines correspond to eastward flow relative to the mantle, and dashed lines, to westward flow. The maximum meridional flow in (A) is 6×10^{-4} m/s; the maximum angular velocity in (B) is 5° per year (contour interval is 0.5° per year). Of particular interest is the contour line adjacent to the inner core boundary, which indicates that the inner core is moving eastward. Note also that the equatorial region of the outer core has a weak westward angular velocity, which helps to conserve total angular momentum of our model Earth and is partially responsible for the typical 0.2° per year westward drift that our magnetic field displays at the core-mantle boundary. **Fig. 3 (right).** Snapshots in the meridian plane of (**A**) the axisymmetric meridional magnetic field and (**B**) the axisymmetric zonal magnetic field. The magnetic lines of force are counterclockwise in (A), and the maximum magnetic intensity is about 25 mT. The contours in (B) are solid for eastward-directed magnetic field and dashed for westward-directed field; the maximum magnetic intensity is about 13 mT (1-mT contour intervals). Of particular interest is the expulsion of zonal field from much of the inner core.

zonal jets would be described as part of a "thermal wind" (*13*), but because strong magnetic fields also exist in the core, especially inside the tangent cylinder, the term "thermal wind" is not precisely correct. However, the strong zonal (east-west) fields (Figs. 3B and 4) are not affected by the zonal velocity because there is little motion across those field lines. The differential zonal flow does shear the strong north-south (meridional) fields (Figs. 3A and 4), but the concentration of these fields along the rotation axis minimizes this effect.

In assessing the effect of these eastward jets on the motion of the inner core, it is helpful to have two physical pictures in mind. The first is based on Alfvén's theorem, which states that magnetic lines of force move with a perfectly conducting medium as though frozen to it. The core is not a perfect electrical conductor, so this theorem does not apply precisely to Earth. Nevertheless, it provides a qualitative description of electromagnetic induction in a moving fluid, in terms of which lines of force passing through the jets tend to be carried eastward with the fluid. These lines of force permeate the inner core (Figs. 3A and 4), and if there is relative motion between the inner core and the jets, the lines of force will lengthen, according to Alfvén's theorem.

The second physical picture describes the dynamical effects of a magnetic field **B** in terms of Faraday-Maxwell stresses, in which magnetic lines of force are in a state of tension. Any relative angular displacement that lengthens the lines of force linking the solid inner core to the fluid outer core above it is resisted by that tension; so, on average, the inner core corotates with the fluid above it. This mechanism has an analogy in electrical engineering: the synchronous motor, where the inner core represents the rotor of an electric motor that is locked synchronously to the eastward propagating magnetic field at the base of the fluid core, created by the convective dynamo operating above it (*1*, *2*).

The "corotation" between the inner core and the fluid lying just above it (*14*) is a generalized view. At any given time, the solid inner core has only one angular velocity, Ω_{IC}, but the angular velocity ω of the fluid jets is a function of position. By corotation we mean that the inner core is rotating at some weighted average of the eastward fluid flow above it such that the net torque on the inner core about the rotation axis is nearly zero. Because the peak zonal flow occurs near the rotation axis, the inner core usually rotates more slowly than the fluid above it in the polar regions and faster than the fluid above it in the equatorial region (Fig. 2B). On areas of the inner core surface where $B_r B_\phi > 0$, the tension of the magnetic field lines increases the eastward Ω_{IC}; where $B_r B_\phi < 0$, it decreases it (*7*). Consider the longitudinally averaged field structure in the snapshot of Fig. 3. The outward-directed field B_r at mid- and high latitude in the northern hemisphere (Fig. 3A) is sheared into eastward-directed field B_ϕ (Fig. 3B); the opposite occurs in the southern hemisphere. However, in both places the product $B_r B_\phi > 0$, which exerts a local, eastward magnetic stress on the inner core. Therefore, the fluid flow and magnetic coupling at the inner core boundary are fundamental to both the rotation of the inner core and the generation of the magnetic field in the outer core.

Magnetic field is also generated by the differential zonal flows (Fig. 2B) in I_N and I_S that shear the meridional fields (Fig. 3A) into zonal fields (Fig. 3B). This field generation maintains a helical magnetic field inside the tangent cylinder (Fig. 4), a combination of the strong meridional field near the rotation axis and the strong zonal field near the tangent cylinder. Recall that the flow is also mainly meridional near the rotation axis and zonal near the tangent cylinder in order to minimize distortion of the field. Our dynamically consistent simulations (*1*, *2*) suggest that the geodynamo develops 3D flow and field structures that produce a delicate, nonlinear balance between the generation of new field by flow structures that twist and shear it and the avoidance of too much resistance to the flow by field structures that are almost force-free in places.

The difference in the structure of the magnetic field between the inner core and the outer core (Fig. 4) may be understood in the following way: The electromagnetic time constant τ_{em} of the inner core is of the order of 10^3 years (*15*), so magnetic fields in the outer core that vary on the time scale of the convection (10^2 years) cannot penetrate far into the inner core. In particular, the time-dependent asymmetric fields, even those of large spatial scale, do not penetrate deeply. However, slowly varying axisymmetric fields, such as those associated with the axial dipole of the Earth, have time to penetrate the inner core completely (Figs. 3A and 4). Because of ohmic dissipation, the associated electric currents deep within the inner core are almost zero, so these slowly varying axisymmetric magnetic fields are good approximations of "potential fields." Zonal fields generated by the global departure of ω from Ω_{IC} do not penetrate far into the inner core because the time scale for reestablishing corotation is very short, and zonal fields created by local departures of ω from Ω_{IC} do not penetrate far into the inner core because of their small spatial scales (Fig. 3B). Therefore, the only fields that persist deep in the inner core are the nearly axisymmetric, potential-like fields, which are nearly parallel to the rotation axis (Fig. 4). These are the fields that provide the magnetic coupling between the inner and outer cores. However, although their maximum strength at the inner core boundary can be as high as 10 mT, they are concentrated in the polar regions (Fig. 4) and so have small moment arms. Therefore, the restoring torques that maintain corotation are not as large as they would otherwise be.

This explanation of why a time-averaged eastward $\overline{\Omega}_{IC}$ is required to bring about a statistically steady state in which

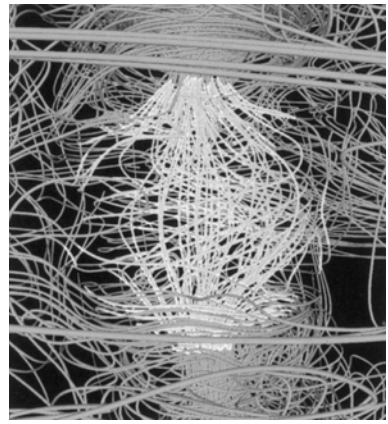


Fig. 4. Snapshot of magnetic lines of force in the core of our simulated Earth. Lines are gold (blue) where they are inside (outside) of the inner core. The axis of rotation is vertical in this image. The field is directed inward at the inner core north pole (top) and outward at the south pole (bottom); the maximum magnetic intensity is about 30 mT.

the time average of the magnetic torque on the inner core $\overline{\Gamma}_B$ is zero ignores the time-average of the viscous torque $\overline{\Gamma}_\nu$ on the inner core (*7*). With a nonzero fluid viscosity, $\overline{\Omega}_{IC}$ adjusts itself so that $\overline{\Gamma}_B + \overline{\Gamma}_\nu = 0$, implying that, because $\overline{\Gamma}_B$ drags the inner core eastward, $\overline{\Gamma}_\nu$ must be acting westward. However, even though the fluid viscosity in our models (*1*, *2*) is several orders of magnitude greater than is likely for the real Earth, the viscous torque on the inner core has little effect.

To demonstrate its effect, we replace our no-slip condition on the zonal flow at the inner core boundary by a viscous stress-free condition, which makes $\Gamma_\nu = 0$, and then continue the integration (*16*). The contours of constant ω near the inner core no longer resemble those in Fig. 2B but, instead, all intersect the inner core boundary. Apart from this, the solution strongly resembles the original one (*2*). The mean Ω_{IC} is still eastward, and the magnitude of Γ_B about its zero time-average is on the order of 10^{16} N m. This value is small enough that Ω_{IC} remains relatively constant on a time scale of decades. It is also small in the sense that the large local magnetic stresses nearly cancel out when integrated over the inner core boundary. That is, when we integrate the absolute value of the moment of the magnetic stress $|B_r B_\phi|/\mu_0$ [instead of $B_r B_\phi/\mu_0$ (*7*)] over the inner core boundary, we consistently obtain values three orders of magnitude greater than Γ_B.

In a further experiment, still using the viscous stress-free boundary condition, we instantaneously reset Ω_{IC} to zero and continue the integration (Fig. 5). As the radial field becomes sheared by the differential motion, the restoring magnetic torque quickly increases, attaining a maximum of 6×10^{18} N m within a couple months. After about 2 years, corotation is completely reestablished (Fig. 5), and the magnitude of Γ_B is again on the order of 10^{16} N m with a zero mean value.

This simple test illustrates the "synchronous motor" mechanism and how short a time is required to spin up Earth's inner core

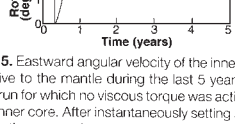


Fig. 5. Eastward angular velocity of the inner core relative to the mantle during the last 5 years of a test run for which no viscous torque was acting on the inner core. After instantaneously setting Ω_{IC} to zero, the magnetic torque reestablishes corotation within about 2 years.

by magnetic torque alone compared with its viscous spin-up time ($\sim 10^4$ years). This 2-year magnetic spin-up time is also short compared with the magnetic diffusion time for the inner core, $\tau_{em} \sim 10^4$ years (*15*). Another way to look at this is to realize that more than 99% of the energy of the magnetic field is inside the core (10% in the inner core) (*1*, *2*), which is typically 10,000 times more energy than the kinetic energy of the rotation of the inner core relative to the mantle. In addition, our simulation predicted a time-dependent Ω_{IC} (Fig. 1) in rough agreement with recent seismic analyses (*9*, *10*) without our model Earth spinning down; that is, we neglected the lunar tidal forces on Earth and instead constrained the total angular momentum of our inner core, outer core, and mantle to be constant (*7*).

Earth's angular momentum is, however, slowly decreasing because of tidal forces; it has been suggested (*10*) that the superrotation of the inner core may be the result of a viscous time lag relative to the spin down of the mantle. However, this explanation relies only on the small viscous torques at the boundaries of the fluid core and ignores the much larger effects of rotating convection and magnetic coupling. For the reasons stated above, we feel that the synchronous motor mechanism plays a much greater role in determining the superrotation of Earth's inner core than would a viscous time-lag mechanism. Also, unlike the viscous time-lag mechanism, which implies a monotonically decreasing inner core rotation rate, our mechanism (Fig. 1) suggests that Earth's inner core has spun both faster and more slowly at times in the past than it is spinning today and predicts that it will spin both faster and more slowly in the future according to the evolving magnetohydrodynamics of Earth's fluid core.

REFERENCES AND NOTES

1. In our original geodynamo simulation [G. A. Glatzmaier and P. H. Roberts, *Phys. Earth Planet. Inter.* **91**, 63 (1995); *Nature* **377**, 203 (1995)], the core was assumed to be nearly uniform in density and temperature: buoyancy was included in the Boussinesq approximation.
2. Here we discuss the results of our most recent geodynamo simulation [G. A. Glatzmaier and P. H. Roberts, *Physica D* **97**, 81 (1996)], which included thermal and compositional convection in the core within the anelastic approximation.
3. The basic state of our iron-rich fluid core (relative to which the 3D time-dependent perturbations are solved) is hydrostatic and well mixed chemically and thermally, and has a spherically symmetric density profile fitted to the Preliminary Reference Earth Model (PREM) [A. M. Dziewonski and D. L. Anderson, *Phys. Earth Planet. Inter.* **25**, 297 (1981)]. We assumed no radioactive heating in the core (*17*) and that the electrical conductivities σ of the inner and outer cores are the same: 4×10^5 S/m. The eddy viscosity in our simulations (*1*, *2*) is, from computational necessity, about three orders of magnitude larger than what would be expected for small-scale turbulence in Earth's fluid core and many orders of magnitude larger than its molecular viscosity. However, the viscous forces in the bulk of our fluid core are still about five orders of magnitude smaller than the Coriolis and Lorentz forces.
4. Because the melting temperature of core material increases with pressure more rapidly than does the temperature along the adiabat, the core presents the unfamiliar situation of a system that is cooled from the top but freezes from the bottom [J. A. Jacobs, *Nature* **172**, 297 (1953)].
5. For example, see S. I. Braginsky and P. H. Roberts, *Geophys. Astrophys. Fluid Dyn.* **79**, 1 (1995).
6. The rate of heat flow out of Earth's core is poorly known. Global 3D mantle convection simulations [P. J. Tackley, D. J. Stevenson, G. A. Glatzmaier, G. Schubert, *J. Geophys. Res.* **99**, 15877 (1994)] obtained 7.2 TW, the value we use. Stacey (*17*) estimates it to be 3.0 TW on the basis of the assumption that the age of the inner core is the same as the age of the Earth and that the rate of heat flow out of the core has remained constant during this time. However, we find in our (unpublished) simulations that a 3.0-TW boundary condition does not maintain a strong, Earth-like magnetic field. Because our 7.2-TW boundary condition (*2*) does maintain an Earth-like field, the age of Earth's inner core may be considerably less than the age of the Earth itself. Of course, many of the material properties that we assume are quite uncertain. For example, we assume that the mass of the light constituent released when a unit mass of alloy freezes at the inner core boundary is 0.065, and the corresponding entropy released is 190 J/(kg K); these have uncertainties of 10 and 50%, respectively (*5*). More significantly, the presence of radioactivity in the core would provide additional thermal buoyancy and therefore would require less compositional buoyancy to obtain similar convective vigor and magnetic field intensity with the same heat flow through the core-mantle boundary but without such a rapid growth rate of the inner core.
7. To obtain the angular velocity Ω_{IC}, we integrated the equation $C(d\Omega_{IC}/dt) = \Gamma_\nu + \Gamma_B$, where $C = 5.86 \times 10^{34}$ kg m^2 is the polar moment of inertia of the inner core and

$$\Gamma_\nu = \oint_{ICB} \mu \bar{\nu} \frac{\partial \omega}{\partial \bar{r}} \bar{r} \sin^2\theta \, dS$$

$$\Gamma_B = -\frac{1}{\mu_0} \oint_{ICB} B_r B_\phi \, r \sin\theta \, dS$$

are the viscous and magnetic torques on the inner core about the polar axis. Here, $\bar{\mu}$ and $\bar{\nu}$ are the fluid density and eddy viscosity in the basic state and μ_0 is the magnetic permeability; **B**, the magnetic field, and **V**, the fluid velocity (including the angular velocity ω), are obtained by integrating the magnetoconvective equations in the fluid core. The integrals defining Γ_ν and Γ_B are over the inner core boundary (ICB), and $dS = r^2 \sin\theta d\theta d\phi$ is an element of surface area. A similar equation of motion can be integrated for the angular velocity of the mantle Ω_M. However, because we apply no torque at the outer boundary of the mantle (that is, at the surface of our model Earth), the total angular momentum of our inner core, outer core, and mantle must remain constant; so we choose to constrain the total angular momentum to be exactly conserved each time step instead of applying the prognostic equation for Ω_M. All of our equations are actually solved in the rotating reference frame for which the total angular momentum of the inner core, outer core, and mantle is zero; therefore, the angular velocity of the inner core relative to the mantle, plotted in Fig. 1, is $\Omega_{IC} - \Omega_M$. However, Ω_M is typically more than two orders of magnitude smaller than Ω_{IC}.
8. This prediction was made as a result of our original geodynamo simulation (*1*). The amplitude of the eastward rotation of the inner core (relative to the mantle) was not substantially different from that derived from the model described here (*2*), although it fluctuated more strongly in time and was actually briefly westward during the polarity reversal

that that model executed during the simulation. The field threading the inner core was at that time very weak, so that the magnetic torque Γ'_{IS} was small. Note also that our models (*1*, *2*) have assumed phase equilibrium on the inner core boundary; that is, as thermodynamic conditions change, freezing and melting occurs instantaneously to maintain the boundary at the freezing point. No allowance has been made for a finite time of relaxation to such a state. If that relaxation time were long compared with the time scales of interest in our model, the inner core would behave as a solid. As B. A. Buffett [*Geophys. Res. Lett.* **23**, 2279 (1996)] noted, the orientation of the inner core would then plausibly be gravitationally "locked" to that of the mantle by the inner core topography created by mantle inhomogeneities, which we have not included in our models. If Earth's inner core is rotating faster than the mantle, as recent observations suggest (*9*, *10*), a short melting-freezing relaxation time, a "mushy zone" at the top of the inner core (D. E. Loper, private communication), or a low inner core viscosity (B. A. Buffett, private communication) may preclude this gravitational locking effect.

9. X. Song and P. G. Richards, *Nature* **382**, 221 (1996).
10. W. Su, A. M. Dziewonski, R. Jeanloz, *Science* **274**, 1883 (1996).
11. R. Hollerbach and C. A. Jones, *Nature* **365**, 541(1993); *Phys. Earth Planet. Inter.* **87**, 171 (1993).
12. For example, P. H. Roberts [*Philos. Trans. R. Soc. London Ser. A* **263**, 93 (1968)] analyzed convection in a sphere, rapidly rotating as measured by the Taylor number *Ta*. He showed that the critical temperature gradient, as measured by a Rayleigh number *Ra*, at which convection could marginally occur is of order $Ta^{2/3}$ for both axisymmetric and nonaxisymmetric modes of convection, but that the nonaxisymmetric motions, later identified by F. H. Busse [*J. Fluid Mech.* **44**, 441 (1970)] as rolls parallel to the rotation axis, have the smaller critical *Ra*. That such rolls are "preferred" can be qualitatively understood in terms of the Proudman-Taylor theorem: Slow, steady motions in an inviscid rotating fluid are 2D with respect to the rotation axis, so that a towed body in such conditions carries a column of fluid with it, a so-called Taylor column. The convective rolls are also often called, a little imprecisely, Taylor columns.
13. Similar thermal wind and meridional circulation structures were obtained by P. Olson and G. A. Glatzmaier [*Phys. Earth Planet. Inter.* **92**, 109 (1995)] using a fully 3D magnetoconvection model but with an imposed zonal magnetic field and a nonrotating, insulating inner core. Similar flow structures also have been obtained by C. Jones and R. Hollerbach (private communication) with their dynamo model using only one nonaxisymmetric mode.
14. Phrases like "just above the inner core boundary" are used in this discussion for presentational simplicity. The fluid in contact with the inner core boundary moves with it because of the no-slip boundary condition; there is no relative motion at the boundary. More precisely, the key quantities determining the coupling between inner and outer cores are B_r and the derivatives of **V** on the inner core boundary, which control B_λ and hence the stress $B_r B_\lambda / \mu_{ij}$ on the inner core boundary. Also, after writing this report we were given a preprint by J. Aurnou, D. Brito, and P. Olson (*Geophys. Res. Lett.*, in press) that describes a simple, analytic model of inner core rotation that approximates the thermal wind and magnetic coupling present in our geodynamo simulations (*1*, *2*).
15. There is no generally accepted way of defining τ_{mag}. We take $\tau_{mag} = \mu_{ij} \pi R^2 / \pi^2$, where R is the Earth's inner core radius, which is appropriate for a sphere surrounded by an insulator. This gives $\tau_{mag} \approx 2400$ years, which is large compared with typical time scales of the fluid motion. The magnetic fields emerging from the inner core impose time scales of order τ_{mag} on the geodynamo mechanism and stabilize it against polarity reversals, a significant fact first suggested by Hollerbach and Jones (*11*) and confirmed by our simulations (*1*, *2*). The absence of zonal field in an inner core coupled magnetically to a fluid core was noted by S. Braginsky [*Geomagn. Aeron.* **4**, 572 (1964)].
16. The viscous stress-free boundary condition forces $\partial \omega / \partial r$ to vanish on the inner core boundary, so the net viscous torque Γ'_ν vanishes there (*7*). This zero torque results in discontinuities in the horizontal velocity between the solid inner core surface and the fluid just above it, which make nonlinear contributions to the magnetic boundary conditions there.
17. F. D. Stacey, *Physics of the Earth* (Brookfield, Brisbane, Australia, ed. 3, 1992).
18. The computing resources were provided by the Pittsburgh Supercomputing Center under grant MCA94P016P and the Advanced Computing Laboratory at Los Alamos National Laboratory. Different aspects of this work were supported by Los Alamos Laboratory Directed Research and Development grant 96149, University of California Directed Research and Development grant 9636, Institute of Geophysics and Planetary Physics grant 713 and NASA grant NCCS5-147. P.H.R. was supported by NSF grant EAR94-06002. The work was conducted under the auspices of the U.S. Department of Energy, supported (in part) by the University of California, for the conduct of discretionary research by Los Alamos National Laboratory.

13 August 1996; accepted 29 October 1996

Part 2

FLUID MECHANICS AND MHD

JOURNAL OF MATHEMATICAL ANALYSIS AND APPLICATIONS 1, 195—214 (1960)

Characteristic Value Problems Posed by Differential Equations Arising in Hydrodynamics and Hydromagnetics

P. H. ROBERTS

Department of Physics, King's College, Newcastle upon Tyne, England

Submitted by S. Chandrasekhar

SUMMARY

A study is made of the general eigenvalue problem posed by a differential equation whose solutions must satisfy certain specified boundary conditions. It is shown that an adjoint eigenvalue problem can be defined and that, in general, it can be represented by a differential equation of the same order as that of the original system together with boundary conditions as numerous as those originally specified. This adjoint system generally differs from the original system but has the same eigenvalues and, moreover, each of its eigensolutions is orthogonal to every eigensolution of the original set except the one(s) belonging to the same eigenvalue. It is shown that the representation of this adjoint system is not, in general, unique, and, if it can be chosen to be identical with the original system, the problem is self-adjoint. A variational method is given for determining the eigenvalues.

To illustrate the application of the method, three problems are considered which are not self-adjoint: the stability of Couette flow, the onset of convection in a rotating sphere heated within, and self-maintained dynamo action in a conducting fluid of infinite extent.

I. INTRODUCTION

Within the past decade, Chandrasekhar has solved a large number of important stability problems in hydrodynamics and hydromagnetics. These have involved the calculation of characteristic values associated with differential equations that have sometimes been of a very high order. For many of these, he was able to express a characteristic value λ_i as the ratio $R(\phi_i)$ of two positive definite integrals in the corresponding characteristic function ϕ_i and he was thence able to show that, if ϕ_i is

subjected to any small variation $\delta\phi_i$, the first order change δR_i in R_i is zero, and conversely. This forms the basis of a powerful and rapid method of determining the characteristic values: for each ϕ_i, assume a physically reasonable trial function, $\Phi_i(\alpha_1, \alpha_2 \ldots)$, which satisfies all the boundary conditions and all the continuity requirements within the region of integration and which, in addition, depends on a number of adjustable parameters $\alpha_1, \alpha_2, \ldots$ Evaluate $R(\Phi_i)$ and minimize the resulting expression with respect to $\alpha_1, \alpha_2, \ldots$ Then this value of $R(\Phi_i)$ is likely to be a good approximation to λ_i, since a first order error in the choice of Φ_i, gives rise to only a second order error in λ_i.

However, it is not always possible to express λ_i as a ratio of two positive definite integrals. Such characteristic value problems are not self-adjoint (in a sense to be defined presently), and are special objects of study in this paper. In § II, it is shown that they admit a variational principle and the method of solution based on this principle is compared with methods devised by Chandrasekhar. In §§ III, IV, and V illustrative examples are given.

II. Method of Solution

Let $\mathscr{L}$ be a linear differential operator defined within a volume $\mathscr{V}$ of the independent variables. Denote by $\mathscr{B}$ conditions on the boundary of $\mathscr{V}$ which define a characteristic value problem from $\mathscr{L}$, i.e.

$$\mathscr{L}\phi_i = \lambda_i\phi_i. \tag{1}$$

Here λ_i and ϕ_i denote a particular characteristic value and a characteristic function belonging to it. We will suppose that the set of functions $\{\phi_i\}$ spans $\mathscr{V}$. Define from $(\mathscr{L}, \mathscr{B})$ an *adjoint* operator and associated boundary conditions $(\mathscr{L}^*, \mathscr{B}^*)$ by the condition that, if f is any function satisfying $\mathscr{B}$ and f^* is any function satisfying $\mathscr{B}^*$, then

$$\int_{\mathscr{V}} f\mathscr{L}^*f^* = \int_{\mathscr{V}} f^*\mathscr{L}f. \tag{2}$$

The operator $\mathscr{L}^*$ defined in this way may be an abstract operator which cannot be represented simply in a differential form. Whether or not a differential form is possible depends on the suitability or otherwise of the conditions $\mathscr{B}^*$. If one repeatedly integrates the right-hand side of expression (2) by parts, and *chooses* each of the conditions $\mathscr{B}^*$ in turn in such a way that the integrated parts vanish either in virtue of $\mathscr{B}$ or in virtue of $\mathscr{B}^*$, then one must discover for $\mathscr{L}^*$ a ***differential*** form which

is of the same order as $\mathscr{L}$. If, however, $\mathscr{B}^*$ is chosen in some way for which $\mathscr{L}^*$ is not a differential operator, it is nevertheless possible to attach a meaning to $\mathscr{L}^*$ (and, as we shall see later, to find a matrix representation for it), and the analysis we are about to give remains valid, although it is not so directly comprehensible. Naturally the representation $(\mathscr{L}^*, \mathscr{B}^*)$ is not unique and there may even be more than one representation in which $\mathscr{L}^*$ can assume a differential form. If, on choosing $\mathscr{B}^*$ to be $\mathscr{B}$, $\mathscr{L}^*$ is found to be $\mathscr{L}$, the characteristic value problem $(\mathscr{L}, \mathscr{B})$ is said to be *self-adjoint*. If not, the characteristic value problem adjoint to (1) is

$$\mathscr{L}^*\phi_i^* = \lambda_i^*\phi_i^*, \tag{1*}$$

where ϕ_i^* denotes a characteristic function satisfying (1*) and $\mathscr{B}^*$, and λ_i^* is the characteristic values to which it belongs. We will suppose that the set $\{\phi_i^*\}$ spans $\mathscr{V}$.

From equations (1), (1*) and (2), we have

$$\lambda_i \int_{\mathscr{V}} \phi_i\phi_j^* = \int_{\mathscr{V}} \phi_j^*\mathscr{L}\phi_i = \int_{\mathscr{V}} \phi_i\mathscr{L}^*\phi_j^* = \lambda_j^* \int_{\mathscr{V}} \phi_i\phi_j^*, \tag{3}$$

i.e.

$$(\lambda_i - \lambda_j^*) \int_{\mathscr{V}} \phi_i\phi_j^* = 0. \tag{4}$$

Hence, if λ_i and λ_j^* are unequal, we have

$$\int_{\mathscr{V}} \phi_i\phi_j^* = 0; \qquad \int_{\mathscr{V}} \phi_j^*\mathscr{L}\phi_i = 0. \tag{5}$$

Moreover, since $\{\phi_i\}$ spans $\mathscr{V}$, the equation

$$\int_{\mathscr{V}} \phi_i\phi_i^* = 0 \tag{6}$$

would, in conjunction with equation (5), imply that ϕ_i^* is identically zero, which is not so, by definition. Hence equation (4) implies that the sets of characteristic values $\{\lambda_i\}$ and $\{\lambda_i^*\}$ are identical. Moreover, the sets $\{\phi_i\}$ and $\{\phi_i^*\}$ are dual (or bi-orthonormal); that is, each function

of either set is orthogonal to every member of the other set except the one(s) belonging to the same characteristic value. From equation (3) it also follows that

$$\lambda_i = \frac{\int_{\mathscr{V}} \phi_i^* \mathscr{L} \phi_i}{\int_{\mathscr{V}} \phi_i^* \phi_i}. \tag{7}$$

Let $\delta\phi_i$ and $\delta\phi_i^*$ be any two small (independent) variations which are applied to ϕ_i and ϕ_i^* and which satisfy the boundary conditions $\mathscr{B}$ and $\mathscr{B}^*$, respectively. The corresponding change $\delta\lambda_i$ in λ_i is

$$\delta\lambda_i = \frac{1}{\int_{\mathscr{V}} \phi_i^* \phi_i} \left[\left\{ \int_{\mathscr{V}} [\delta\phi_i^* \mathscr{L}\phi_i + \delta\phi_i \mathscr{L}^* \phi_i^*] \right\} - \frac{[\int_{\mathscr{V}} \phi_i^* \mathscr{L}\phi_i]}{[\int_{\mathscr{V}} \phi_i^* \phi_i]} \left\{ \int_{\mathscr{V}} [\phi_i^* \delta\phi_i + \phi_i \delta\phi_i^*] \right\} \right], \tag{8}$$

or, using equation (7),

$$\delta\lambda_i = \left[\int_{\mathscr{V}} \delta\phi_i^* [\mathscr{L}\phi_i - \lambda_i \phi_i] + \int_{\mathscr{V}} \delta\phi_i [\mathscr{L}^* \phi_i^* - \lambda_i \phi_i^*] \right] \Big/ \int_{\mathscr{V}} \phi_i^* \phi_i. \tag{9}$$

Hence, if equations (1) and (1*) are obeyed, $\delta\lambda_i$ is zero to first order for all small, arbitrary variations $\delta\phi_i$ and $\delta\phi_i^*$ satisfying $\mathscr{B}$ and $\mathscr{B}^*$, respectively. Evidently, the converse is also true. This forms the basis of a variational method which is analogous to the method described briefly in §1 and reduces to it if the problem is self-adjoint. Provided the trial functions $\Phi_i(\alpha_1, \alpha_2, \ldots)$ and $\Phi_i^*(\alpha_1^*, \alpha_2^*, \ldots)$ are appropriately chosen and, of course, satisfy the boundary conditions $\mathscr{B}$ and $\mathscr{B}^*$, respectively, the error made in determining λ_i from equation (7) is small, being of second order in the error of Φ_i and Φ_i^*. The error can be reduced by determining the stationary value of λ_i under variations of $\alpha_1, \alpha_2, \ldots$ and $\alpha_1^*, \alpha_2^*, \ldots$ This stationary value will not, of course, be a minimum in general.

It remains to extend our discussion to the case in which the adjoint problem does not necessarily have a differential representation. We can

find a matrix representation by taking two orthogonal bases $\{f_\alpha\}$ and $\{f_\alpha^*\}$ of $\mathscr{V}$ which satisfy $\mathscr{B}$ and $\mathscr{B}^*$, respectively. We associate with ϕ an infinite column vector $\{a_\beta\}$ by means of the expansion

$$\phi = \sum_{\beta=1}^{\infty} a_\beta f_\beta, \tag{10}$$

and we represent the operator $\mathscr{L}$ and the identity operator $\mathscr{I}$ by the infinite matrices $L_{\alpha\beta}$ and $I_{\alpha\beta}$ defined by

$$L_{\alpha\beta} = \int_{\mathscr{V}} f_\alpha^* \mathscr{L} f_\beta, \qquad I_{\alpha\beta} = \int_{\mathscr{V}} f_\alpha^* f_\beta. \tag{11}$$

Since $\{f_\alpha\}$ and $\{f_\alpha^*\}$ are orthogonal bases, we find that equation (1) reduces, in virtue of equations (10) and (11), to the matrix equation

$$\sum_{\beta=1}^{\infty} L_{\alpha\beta} a_\beta = \lambda \sum_{\beta=1}^{\infty} I_{\alpha\beta} a_\beta, \qquad \alpha = 1, 2, \ldots \tag{12}$$

The existence of solutions to these equations implies the existence of solutions, having the *same* characteristic values, to the equations

$$\sum_{\beta=1}^{\infty} L_{\beta\alpha} a_\beta^* = \lambda \sum_{\beta=1}^{\infty} I_{\beta\alpha} a_\beta^*, \qquad \alpha = 1, 2, \ldots \tag{12*}$$

The characteristic values themselves are the roots of the equation

$$||L_{\alpha\beta} - \lambda I_{\alpha\beta}|| = 0, \tag{13}$$

where the vertical double lines at each side denote the infinite determinant whose element in row α and column β is given by the expression between them. Equations (12*) are the matrix representation of the adjoint system, the characteristic functions of which are

$$\phi^* = \sum_{\beta=1}^{\infty} a_\beta^* f_\beta^*. \tag{10*}$$

One of the methods used by Chandrasekhar for problems which are not self-adjoint is that of solving equation (13) approximately. To find any particular characteristic value, he picks the base function f_1 (say)

which is physically the most plausible and adds a few (say, $n - 1$) of the adjacent functions $f_2, f_3, \ldots f_n$ (say) to form a finite approximation to ϕ:

$$\phi = \sum_{\beta=1}^{n} a_\beta f_\beta. \tag{14}$$

Equation (13) now reduces to an $n \times n$ determinant and, after solving this for the appropriate characteristic value $\lambda(n)$, he examines the behaviour of $\lambda(n)$ as a function of n. If $\lambda(n)$ appears to converge to a limit $\lambda(\infty)$, say, he presumes that this is a good approximation to the required characteristic value. Very often, the convergence is rapid and leads speedily to a good estimate of λ. That this is so is not surprising. We can regard equation (14) and the corresponding approximation

$$\phi^* = \sum_{\beta=1}^{n} a_\beta^* f_\beta^* \tag{14*}$$

as trial functions for the variational principle based on equation (7). The best values of $\{a_\alpha\}$ and $\{a_\alpha^*\}$ are those for which

$$\partial\lambda/\partial a_\alpha = \partial\lambda/\partial a_\alpha^* = 0, \qquad \alpha = 1, 2, \ldots n, \tag{15}$$

that is, those for which

$$\sum_{\beta=1}^{n} L_{\alpha\beta} a_\beta = \lambda \sum_{\beta=1}^{n} I_{\alpha\beta} a_\beta, \qquad \alpha = 1, 2, \ldots n, \tag{16}$$

$$\sum_{\beta=1}^{n} L_{\beta\alpha} a_\beta^* = \lambda \sum_{\beta=1}^{n} I_{\beta\alpha} a_\beta^*, \qquad \alpha = 1, 2, \ldots n. \tag{16*}$$

Elimination of a_β or a_β^* leads to the determinant equation of order n which determines $\lambda(n)$.

It is appropriate here to describe another method frequently used by Chandrasekhar. Its success depends on the introduction of auxiliary variables with the aid of which λ may be expressed as the ratio of two positive definite integrals despite the fact that the problem is not self-adjoint. In essence, it is based on the possibility of factorizing $\mathscr{L}(=\mathscr{N}\mathscr{M})$ into two non-commuting operators $\mathscr{M}, \mathscr{N}$ each of which is self-adjoint even though their product is not. If such a division is possible, the

boundary conditions $\mathscr{B}$ can be separated into two groups $\mathscr{C}$ and $\mathscr{D}$ in such a way that the differential equation

$$\mathscr{N}\psi = \phi \tag{17}$$

together with $\mathscr{D}$ determine ψ from ϕ uniquely. The characteristic value problem can now be expressed as the solution of the differential equation

$$\mathscr{M}\phi = \lambda\psi \tag{18}$$

subject to the boundary conditions $\mathscr{C}$. The statement we have made that $\mathscr{M},\mathscr{N}$ are self-adjoint means in this context that, if f_1 and f_2 are any two functions satisfying $\mathscr{C}$, and g_1 and g_2 are any two functions satisfying $\mathscr{D}$, then

$$\int_{\mathscr{V}} f_1 \mathscr{M} f_2 = \int_{\mathscr{V}} f_2 \mathscr{M} f_1; \qquad \int_{\mathscr{V}} g_1 \mathscr{N} g_2 = \int_{\mathscr{V}} g_2 \mathscr{N} g_1. \tag{19}$$

According to equation (18),

$$\lambda = \int_{\mathscr{V}} \phi \mathscr{M} \phi \Big/ \int_{\mathscr{V}} \psi \mathscr{N} \psi, \tag{20}$$

and so, on a small variation $\delta\phi$, the corresponding change in λ is

$$\delta\lambda = 2\left[\int_{\mathscr{V}} \delta\phi \mathscr{M}\phi - \lambda \int_{\mathscr{V}} \psi \mathscr{N} \delta\psi\right] \Big/ \int \psi \mathscr{N} \psi, \tag{21}$$

and, according to the definition of ψ, the corresponding change in $\delta\psi$ satisfies (cf. equation (17))

$$\mathscr{N}\delta\psi = \delta\phi. \tag{22}$$

Hence

$$\delta\lambda = 2\int_{\mathscr{V}} \delta\phi(\mathscr{M}\phi - \lambda\psi) \Big/ \int_{\mathscr{V}} \psi \mathscr{N} \psi, \tag{23}$$

from which the variational principle follows. It is a principle of the conventional type in which λ is expressed as the ratio of two positive definite forms. However, it is clear that, according to equation (2),

$$\mathscr{L}^* = (\mathscr{N}\mathscr{M})^* = \mathscr{M}^*\mathscr{N}^* = \mathscr{M}\mathscr{N}, \tag{24}$$

and this is not equal to $\mathscr{L}(=\mathscr{N}\mathscr{M})$ since $\mathscr{M},\mathscr{N}$ do not commute. Thus, the characteristic value problem is not self-adjoint but, provided the auxiliary variable ψ is introduced, the variational principle nevertheless involves the ratio of two positive definite forms.

In the following sections, we consider three particular problems which are not self-adjoint; namely, the Taylor problem of the stability of viscous flow between rotating cylinders, and the onset of convection in a rotating fluid sphere heated within, and a modified form of the dynamo problem.

III. Stability of Couette Flow

From the mathematical viewpoint, this example is particularly suited for demonstrating the non-uniqueness of the adjoint system itself. For simplicity, we will limit the discussion to the case in which the separation between the rotating cylinders is small compared to their radii. The method can, however, be extended fairly easily to the more general case. We consider, then, the characteristic value problem posed by the differential equation

$$(D^2 - a^2)^3 v = -a^2 \,\mathrm{Ta}\,(1 + \alpha z)v, \tag{25}$$

whose solutions are subjected to the boundary conditions

$$v = (D^2 - a^2)v = D(D^2 - a^2)v = 0, \qquad \text{at} \qquad z = 0, 1, \tag{26}$$

where Ta, the Taylor number, is the characteristic value parameter, $D = d/dz$, and $a, \alpha (0 \geqslant \alpha \geqslant -3.0)$ are real constants [2, 3]. Chandrasekhar has shown that the best numerical results can be obtained by splitting equation (25) into two separate equations; a fourth order one (in an auxiliary variable ϕ) which is solved exactly, and a second order one which is solved by other means. We write, therefore,

$$\mathscr{L}: \begin{cases} (D^2 - a^2)^2 \phi = (1 + \alpha z)v, & (27) \\ (D^2 - a^2)v = -a^2 \,\mathrm{Ta}\,\phi, & (28) \end{cases}$$

where

$$\mathscr{B}: \quad \phi = D\phi = v = 0, \qquad \text{at} \qquad z = 0, 1. \tag{29}$$

The adjoint problem can be found quite systematically by the method explained in § II using the orthogonal base of functions

$$f_n = f_n^* = \sin n\pi z, \qquad n = 1, 2, \ldots \tag{30}$$

It is then easily shown that, in this matrix representation, the problem reduces to an infinite set of linear equations equivalent to a set already derived by Chandrasekhar [3, eq. (30)]. However, much the simplest way of defining an adjoint problem is by direct integration by parts in the manner described in § II (eq. 2 *et seq.*). We will show, in fact, that the adjoint problem may be expressed as

$$\mathscr{L}^*: \begin{cases} (D^2 - a^2)^2 \phi^* = v^*, & (27^*) \\ (D^2 - a^2) v^* = -a^2 \mathrm{Ta}^* (1 + \alpha z)\phi^*, & (28^*) \end{cases}$$

where

$$\mathscr{B}^* = \mathscr{B}: \ \phi^* = D\phi^* = v^* = 0, \qquad \text{at} \qquad z = 0, 1. \tag{29*}$$

To prove this, consider the integral

$$I = \int_0^1 [(Dv)(Dv^*) + a^2 v v^*]\, dz \tag{31}$$

evaluated in two different ways. Integrating by parts using equations (27*), (28), (29*), we find

$$I = -\int_0^1 v^*(D^2 - a^2)v\, dz = a^2 \mathrm{Ta} \int_0^1 \phi (D^2 - a^2)^2 \phi^*\, dz. \tag{32}$$

Integrating by parts twice, using equations (29), we obtain

$$I = a^2 \mathrm{Ta} \int_0^1 [(D^2 - a^2)\phi][(D^2 - a^2)\phi^*]\, dz. \tag{33}$$

Alternatively, integrating I by parts using equations (27), (28*), and (29) gives

$$I = -\int_0^1 v(D^2 - a^2)v^*\, dz = a^2 \mathrm{Ta}^* \int_0^1 \phi^* (D^2 - a^2)^2 \phi\, dz. \tag{32*}$$

Integrating by parts twice more, using equations (29*), we obtain

$$I = a^2 \mathrm{Ta}^* \int_0^1 [(D^2 - a^2)\phi][(D^2 - a^2)\phi^*]\, dz. \tag{33*}$$

Comparing equations (33) and (33*), we see that I vanishes unless Ta and Ta* are equal, and also that

$$\mathrm{Ta} = \frac{\int_0^1 [(Dv)(Dv^*) + a^2 v v^*]\, dz}{a^2 \int_0^1 [(D^2 - a^2)\phi][(D^2 - a^2)\phi^*]\, dz}. \tag{34}$$

That this forms the basis of a variational principle follows from an argument too close to that given in § II to be worth repeating here.

It should be noticed that the set (27*) to (29*) is not the only representation of the adjoint problem in terms of a differential operator. The set given corresponds to the *same* boundary conditions (26) as those of the original problem but a *different* differential equation, namely

$$(D^2 - a^2)^2 \frac{1}{(1 + \alpha z)} (D^2 - a^2) v^* = - a^2 \,\mathrm{Ta}\, v^*. \tag{35}$$

However, it is clear from equations (27*) to (29*) that we can also represent the adjoint problem by the *same* differential equation (25) but with *different* boundary conditions, namely

$$v^* = Dv^* = (D^2 - a^2)^2 v^* = 0, \qquad \text{at} \qquad z = 0, 1, \tag{36}$$

which, are, in a sense conditions complementary to (26). It is obviously also possible to represent the adjoint problem in yet another way in which neither differential equation nor boundary conditions are the same, namely

$$(D^2 - a^2) \frac{1}{(1 + \alpha z)} (D^2 - a^2)^2 v^* = - a^2 \,\mathrm{Ta}\, v^*, \tag{37}$$

where

$$(D^2 - a^2) v^* = (D^2 - a^2)^2 v^* = D(D^2 - a^2)^2 v^* = 0, \qquad \text{at} \qquad z = 0, 1. \tag{38}$$

IV. Convection in a Rotating Sphere

From the mathematical viewpoint, this example is particularly suited for demonstrating that, with but one representation of the adjoint problem, it is possible to find several distinct variational principles differing in the choice of the auxiliary variables. In suitably chosen units,

the stationary states of marginal stability of convection in a rotating sphere are governed by the equations

$$\mathscr{L}: \begin{cases} \nabla^2\Theta + \mathbf{u}\cdot\mathbf{r} = 0, & (39) \\ \mathrm{Ra}\,\Theta\mathbf{r} = \mathrm{curl}^2\,\mathbf{u} + \mathrm{Ta}^{1/2}\,\mathbf{l}_z \times \mathbf{u} + \mathrm{grad}\,\tilde{\omega}, & (40) \\ \mathrm{div}\,\mathbf{u} = 0, & (41) \end{cases}$$

and the boundary conditions

$$\mathscr{B}: \begin{cases} \Theta = 0, \quad \text{at} \quad r = 1, & (42) \\ \mathbf{u} = 0, \quad \text{at} \quad r = 1. & (43) \end{cases}$$

Here Θ denotes the temperature fluctuation; $\tilde{\omega}$, the pressure fluctuation; $\mathbf{u}$, the fluid velocity; $\mathbf{r}$, the coordinate vector from the centre of the sphere; and $\mathbf{l}_z$, a unit vector along the axis of rotation. The conditions (43) are appropriate to a rigid boundary at the surface of the sphere; the case of a free boundary poses no extra problems but is slightly more cumbersome algebraically. Ra and Ta denote the Rayleigh number and Taylor number, respectively. Full details of the derivation of these equations are given by Chandrasekhar [4]. If Ta is non-zero, the characteristic value problem is not self-adjoint. We will show that a representation of the adjoint system is

$$\mathscr{L}^*: \begin{cases} \nabla^2\Theta^* + \mathbf{u}^*\cdot\mathbf{r} = 0, & (39^*) \\ \mathrm{Ra}^*\,\Theta^*\mathbf{r} = \mathrm{curl}^2\,\mathbf{u}^* - \mathrm{Ta}^{1/2}\,\mathbf{l}_z \times \mathbf{u}^* + \mathrm{grad}\,\tilde{\omega}^*, & (40^*) \\ \mathrm{div}\;\mathbf{u}^* = 0, & (41^*) \end{cases}$$

together with the same boundary conditions

$$\mathscr{B}^* = \mathscr{B}: \begin{cases} \Theta^* = 0, \quad \text{at} \quad r = 1, & (42^*) \\ \mathbf{u}^* = 0, \quad \text{at} \quad r = 1. & (43^*) \end{cases}$$

That Ra and Ra* are equal is physically obvious, since the equations (39*) to (43*) correspond to the original problem with an opposite sense of rotation, i.e., taking spherical polar coordinates (r, θ, ϕ) with $\mathbf{l}_z$ as axis, if

$$\Theta = \Theta(r, \theta, \phi), \qquad \mathbf{u} = [u_r(r, \theta, \phi), u_\theta(r, \theta, \phi), u_\phi(r, \theta, \phi)], \qquad (44)$$

satisfies equations (39) to (43), then

$$\Theta^* = \Theta(r, \theta, -\phi), \mathbf{u}^* = [u_r(r, \theta, -\phi), u_\theta(r, \theta, -\phi), -u_\phi(r, \theta, -\phi)], \quad (44^*)$$

satisfies equations (39*) to (43*), provided we set Ra = Ra*. To prove that equations (39*) to (43*) do, indeed, define the adjoint problem, consider the integral

$$I = \int (\operatorname{grad} \Theta) \cdot (\operatorname{grad} \Theta^*)\, d\tau \tag{45}$$

evaluated in two different ways. By the divergence theorem and equations (39) and (42*), we have

$$I = -\int \Theta^* \nabla^2 \Theta\, d\tau = \int \mathbf{u} \cdot \Theta^* \mathbf{r}\, d\tau. \tag{46}$$

Using now equations (40*), (41*), (43) in conjunction with the divergence theorem, we find

$$I = \frac{1}{\mathrm{Ra}} \left\{ \int (\operatorname{curl} \mathbf{u}) \cdot (\operatorname{curl} \mathbf{u}^*)\, d\tau + \mathrm{Ta}^{1/2} \int \mathbf{1}_z \cdot (\mathbf{u} \times \mathbf{u}^*)\, d\tau \right\}. \tag{47}$$

Alternatively, by the divergence theorem and equations (39*) and (42), we have

$$I = -\int \Theta \nabla^2 \Theta^*\, d\tau = \int \mathbf{u}^* \cdot \Theta \mathbf{r}\, d\tau. \tag{46*}$$

Using now equations (40), (41), (43*) in conjunction with the divergence theorem, we find

$$I = \frac{1}{\mathrm{Ra}^*} \left\{ \int (\operatorname{curl} \mathbf{u}) \cdot (\operatorname{curl} \mathbf{u}^*)\, d\tau - \mathrm{Ta}^{1/2} \int \mathbf{1}_z \cdot (\mathbf{u}^* \times \mathbf{u})\, d\tau \right\}. \tag{47*}$$

Comparing equations (47) and (47*), we see that I vanishes unless Ra and Ra* are equal, and also that

$$\mathrm{Ra} = \frac{\int (\operatorname{curl} \mathbf{u}) \cdot (\operatorname{curl} \mathbf{u}^*)\, d\tau + \mathrm{Ta}^{1/2} \int \mathbf{1}_z \cdot (\mathbf{u} \times \mathbf{u}^*)\, d\tau}{\int (\operatorname{grad} \Theta) \cdot (\operatorname{grad} \Theta^*)\, d\tau} \tag{48}$$

If Ta is zero, the problem is self-adjoint, and equation (48) leads to the variational principle discovered by Chandrasekhar [1] for that case. The proof that the expression (48) forms the basis of a variational method in the more general case contains a point of divergence from the proof of § II, and will, therefore, be considered below. The attitude we take is that Θ and Θ^* are auxiliary variables determined from $\mathbf{u}$ and $\mathbf{u}^*$ by the subsidiary equations (39) and (39*) and the conditions (42) and (42*).

Thus, when we make small independent solenoidal variations $\delta\mathbf{u}$ and $\delta\mathbf{u}^*$ to $\mathbf{u}$ and $\mathbf{u}^*$, we will not suppose that $\delta\Theta$ and $\delta\Theta^*$ are arbitrary, but we will consider that they are determined from $\delta\mathbf{u}$ and $\delta\mathbf{u}^*$ by equations (39), (39*), (42), (42*).

According to equation (48), the first order change $\delta\mathrm{Ra}$ in Ra produced by the small variations $\delta\mathbf{u}$ and $\delta\mathbf{u}^*$ is

$$\delta\,\mathrm{Ra} = \frac{1}{I}\left[\int\{[(\operatorname{curl}\mathbf{u})\cdot(\operatorname{curl}\delta\mathbf{u}^*) + \mathrm{Ta}^{1/2}\,\mathbf{1}_z\cdot(\mathbf{u}\times\delta\mathbf{u}^*) - \mathrm{Ra}\,(\operatorname{grad}\Theta)\cdot(\operatorname{grad}\delta\Theta^*)] + [(\operatorname{curl}\mathbf{u}^*)\cdot(\operatorname{curl}\delta\mathbf{u}) - \mathrm{Ta}^{1/2}\,\mathbf{1}_z\cdot(\mathbf{u}^*\times\delta\mathbf{u}) - \mathrm{Ra}\,(\operatorname{grad}\Theta^*)\cdot(\operatorname{grad}\delta\Theta)]\}\,d\tau\right]. \tag{49}$$

On applying the divergence theorem, using $\mathscr{B}$ and $\mathscr{B}^*$ and equations (39), (39*) to relate $\delta\Theta$ and $\delta\Theta^*$ to $\delta\mathbf{u}$ and $\delta\mathbf{u}^*$, we find

$$\delta\,\mathrm{Ra} = \frac{1}{I}\left[\int\{\delta\mathbf{u}^*\cdot[\operatorname{curl}^2\mathbf{u} + \mathrm{Ta}^{1/2}\,(\mathbf{1}_z\times\mathbf{u}) - \mathrm{Ra}\,\Theta\mathbf{r}] + \delta\mathbf{u}\cdot[\operatorname{curl}^2\mathbf{u}^* - \mathrm{Ta}^{1/2}\,(\mathbf{1}_z\times\mathbf{u}^*) - \mathrm{Ra}\,\Theta^*\mathbf{r}]\}\,d\tau\right]. \tag{50}$$

Hence, if $\mathbf{u}$ and $\mathbf{u}^*$ satisfy equations (40) and (40*), then Ra is zero for all independent *solenoidal* variations $\delta\mathbf{u}$ and $\delta\mathbf{u}^*$. Conversely, if $\delta\mathrm{Ra}$ vanishes for all independent *solenoidal* variations $\delta\mathbf{u}$ and $\delta\mathbf{u}^*$ satisfying the boundary conditions (43) and (43*), then there exist single-valued functions $\tilde{\omega}$ and $\tilde{\omega}^*$ such that

$$\operatorname{curl}^2\mathbf{u} + \mathrm{Ta}^{1/2}\,\mathbf{1}_z\times\mathbf{u}^* - \mathrm{Ra}\,\Theta = -\operatorname{grad}\tilde{\omega}, \tag{40}$$

$$\operatorname{curl}^2\mathbf{u}^* - \mathrm{Ta}^{1/2}\,\mathbf{1}_z\times\mathbf{u}^* - \mathrm{Ra}\,\Theta^* = -\operatorname{grad}\tilde{\omega}^*. \tag{40*}$$

Note that we must use the facts that $\delta\mathbf{u}$ and $\delta\mathbf{u}^*$ are solenoidal. This reflects the fact that the characteristic vectors $\mathbf{u}(\mathbf{u}^*)$ satisfying $\mathscr{L}$ and $\mathscr{B}$ ($\mathscr{L}^*$ and $\mathscr{B}^*$) do not form a base by which *any* vector satisfying equation (43) (equation 43*) can be expanded, but only a base by which any such *solenoidal* vector can be expanded.

The formulation of a variational principle is not unique. Let us divide $\mathbf{u}$ into its toroidal and poloidal parts

$$\mathbf{u} = \mathbf{u}_T + \mathbf{u}_S = \operatorname{curl} T\mathbf{r} + \operatorname{curl}^2 S\mathbf{r} \tag{51}$$

It may be shown, in a straightforward manner, that an arbitrary vector field $\mathbf{A}$ may be written in the form

$$\mathbf{A} = \operatorname{grad} U' + V'\mathbf{r} + \operatorname{curl} W'\mathbf{r}, \tag{52}$$

where

$$L^2U' = \left(1 + x_i \frac{\partial}{\partial x_i}\right) x_j A_j - r^2 \frac{\partial A_i}{\partial x_i}, \tag{53}$$

$$L^2V' = \frac{\partial}{\partial x_j}\left(1 + x_i \frac{\partial}{\partial x_i}\right) A_j - \nabla^2 x_i A_i,$$

$$L^2W' = \mathbf{r} \cdot \operatorname{curl} \mathbf{A},$$

and L^2 is a second order differential operator which commutes with ∇^2, and which can be expressed in any of the forms

$$L^2 = \left(x_i \frac{\partial}{\partial x_i} + x_i \frac{\partial}{\partial x_i} x_j \frac{\partial}{\partial x_j} - r^2 \nabla^2\right) = r^2 \left(\frac{\partial^2}{\partial r^2} + \frac{2}{r}\frac{\partial}{\partial r} - \nabla^2\right) \tag{54}$$

$$= -\left[\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta \frac{\partial}{\partial\theta}\right) + \frac{1}{\sin^2\theta}\frac{\partial^2}{\partial\phi^2}\right].$$

It therefore follows, writing $\mathbf{1}_r \times \mathbf{u}$ for $\mathbf{A}$, that

$$\mathbf{1}_r \times \mathbf{u} = \operatorname{grad} U + V\mathbf{r} + \operatorname{curl} W\mathbf{r}, \tag{55}$$

where

$$L^2U = r\left[\sin\theta \frac{\partial}{\partial\theta} + \cos\theta L^2\right] T - \frac{\partial}{\partial\phi}\left[\frac{\partial}{\partial r}(rS) + L^2S\right], \tag{56}$$

$$L^2V = Q^3T + \frac{\partial}{\partial\phi}\nabla^2 S,$$

$$L^2W = Q^3S - \frac{\partial T}{\partial\phi}$$

and where Q^3 is a third order differential operator:

$$Q^3 = \frac{1}{r\sin\theta}\left\{\left[\frac{\partial}{\partial\theta}\left(\sin\theta\cos\theta\frac{\partial}{\partial\theta}\right) + \cot\theta\frac{\partial^2}{\partial\phi^2}\right]\left(r\frac{\partial}{\partial r} + 1\right) - \right. \tag{57}$$

$$\left. \left[\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\right) + \frac{1}{\sin\theta}\frac{\partial^2}{\partial\phi^2}\right]\left(\sin\theta\frac{\partial}{\partial\theta}\right)\right\}$$

$$= -\frac{1}{r}\left\{\left(\sin\theta\frac{\partial}{\partial\theta} + \cos\theta L^2\right)\left(r\frac{\partial}{\partial r} + 1\right) - L^2\left(\sin\theta\frac{\partial}{\partial\theta}\right)\right\}.$$

It is clear, by integrating three times by parts, that

$$\int AQ^3B\,d\tau = -\int BQ^3A\,d\tau, \tag{58}$$

provided the integrated parts vanish (as they do for the functions A, B to which equation (58) is applied below).

Applying equations (51), (55) and (56) to equations (39) and (40), we find

$$\nabla^2\Theta = -L^2S, \tag{59}$$

$$\nabla^2\left[\nabla^2L^2S + \mathrm{Ta}^{1/2}\frac{\partial S}{\partial\phi}\right] + \mathrm{Ta}^{1/2}Q^3T = \mathrm{Ra}\,L^2\Theta, \tag{60}$$

$$\left[\nabla^2L^2T + \mathrm{Ta}^{1/2}\frac{\partial T}{\partial\phi}\right] - \mathrm{Ta}^{1/2}Q^3S = 0, \tag{61}$$

together with another equation which relates $\tilde{\omega}$, T, S and which is of no significance here. The boundary conditions on Θ, S, T are

$$\Theta = S = T = \partial S/\partial r = 0, \qquad \text{at} \qquad r = 1. \tag{62}$$

As before the adjoint problem can be represented by the same boundary conditions and by equations derived from (59) to (62) by reversing the sign of $\mathrm{Ta}^{1/2}$. This implies, again, that one solution to the adjoint system is

$$\Theta^*(r,\theta,\phi) = \Theta(r,\theta,-\phi); \qquad S^*(r,\theta,\phi) = S(r,\theta,-\phi); \tag{63}$$

$$T^*(r,\theta,\phi) = -T(r,\theta,-\phi).$$

Before proceeding further, notice that, from the equation adjoint to (59), the boundary conditions (62), and applications of the divergence theorem, that

$$\int S^*L^2\Theta\,d\tau = \int (\mathrm{grad}\,\Theta)\cdot(\mathrm{grad}\,\Theta^*)\,d\tau \tag{64}$$

By applications of the divergence theorem and use of equations (60) and (64), we find

$$\int (\nabla^2S^*)(L^2\nabla^2S)\,d\tau = \int S^*\nabla^4L^2S\,d\tau \tag{65}$$

$$= \mathrm{Ra}\int S^*L^2\Theta\,d\tau - \mathrm{Ta}^{1/2}\int S^*\frac{\partial}{\partial\phi}\nabla^2S\,d\tau - \mathrm{Ta}^{1/2}\int S^*Q^3T\,d\tau$$

$$= \mathrm{Ra}\int (\mathrm{grad}\,\Theta)\cdot(\mathrm{grad}\,\Theta^*)\,d\tau - \mathrm{Ta}^{1/2}\int S^*\frac{\partial}{\partial\phi}\nabla^2S\,d\tau - \mathrm{Ta}^{1/2}\int S^*Q^3T\,d\tau,$$

a result which may also be written

$$\int (\operatorname{curl} \mathbf{u}_S) \cdot (\operatorname{curl} \mathbf{u}_S^*)\, d\tau = \mathrm{Ra} \int (\operatorname{grad} \Theta) \cdot (\operatorname{grad} \Theta^*)\, d\tau - \mathrm{Ta}^{1/2} \int \mathbf{1}_z \cdot (\mathbf{u} \times \mathbf{u}_S^*)\, d\tau. \tag{66}$$

Similarly, by the divergence theorem and equation (61*), we find

$$\int TL^2 \nabla^2 T^*\, d\tau = \mathrm{Ta}^{1/2} \int T \frac{\partial T^*}{\partial \phi}\, d\tau - \mathrm{Ta}^{1/2} \int TQ^3 S^*\, d\tau, \tag{67}$$

a result which may also be written

$$\int (\operatorname{curl} \mathbf{u}_T) \cdot (\operatorname{curl} \mathbf{u}_T^*)\, d\tau = -\mathrm{Ta}^{1/2} \int \mathbf{1}_z \cdot (\mathbf{u} \times \mathbf{u}_T^*)\, d\tau. \tag{68}$$

By *adding* equations (66) and (68), we recover the variational principle (48) which, as we have already remarked, is not self-adjoint and gives a stationary value for Ra for all independent variations δS, δS^*, δT, δT^* when S, T satisfies equations (60) and (61). By subtracting equations (45) and (46), we find

$$\mathrm{Ra} \int (\operatorname{grad} \Theta) \cdot (\operatorname{grad} \Theta^*)\, d\tau = \int [\operatorname{curl} (\mathbf{u}_S + \mathbf{u}_T)] \cdot [\operatorname{curl} (\mathbf{u}_S^* - \mathbf{u}_T^*)]\, d\tau + \mathrm{Ta}^{1/2} \int \mathbf{1}_z \cdot [(\mathbf{u}_S + \mathbf{u}_T) \times (\mathbf{u}_S^* - \mathbf{u}_T^*)]\, d\tau. \tag{69}$$

It can be shown that this also forms the basis for a variational principle which is not self-adjoint in general. Expression (69) gives a stationary value for Ra for all independent variations δS and δS^* when equations (60) and (60*) are satisfied, *provided* T and T^* are regarded as auxiliary variables *defined* from S and S^* by equations (61) and (61*) (and the boundary conditions that T and T^* vanish at the surface of the sphere).

For solutions independent of ϕ, equation (63) shows that

$$\mathbf{u}_S + \mathbf{u}_T = \mathbf{u}_S^* - \mathbf{u}_T^* = \mathbf{u}, \qquad \Theta^* = \Theta, \tag{70}$$

and hence equation (69) gives rise to a variational principle which is self-adjoint (cf. *5*, *6*):

$$\mathrm{Ra} = \frac{\int (\operatorname{curl} \mathbf{u})^2\, d\tau}{\int (\operatorname{grad} \Theta)^2\, d\tau}. \tag{71}$$

This has the familiar physical interpretation that, in a state of marginal stability, there is a balance between the rate at which viscosity degrades hydrodynamic energy into heat and the rate at which buoyancy forces do work. However, we reiterate that equation (69) can form the basis of a variational principle, only when we interpret equation (61) as a restraint imposed by the rotation. This was unnecessary for the principle (48).

Even in the case where solutions are sought which are not independent of ϕ, it is possible to derive a variational principle which is self-adjoint by the application of alternative constraints. Write

$$\left.\begin{aligned} \Theta &= \Theta_1 \cos m\phi + \Theta_2 \sin m\phi \\ S &= S_1 \cos m\phi + S_2 \sin m\phi \\ T &= T_1 \cos m\phi + T_2 \sin m\phi \end{aligned}\right\} \tag{72}$$

Then, by equations (59) to (61), we have

$$\nabla^2\Theta_1 = -L^2 S_1, \tag{73}$$

$$\nabla^2\Theta_2 = -L^2 S_2, \tag{74}$$

$$\nabla^2[\nabla^2 L^2 S_1 + m\,\mathrm{Ta}^{1/2} S_2] = \mathrm{Ra}\,L^2\Theta_1 - \mathrm{Ta}^{1/2} Q^3 T_1, \tag{75}$$

$$\nabla^2[\nabla^2 L^2 S_2 - m\,\mathrm{Ta}^{1/2} S_1] = \mathrm{Ra}\,L^2\Theta_2 - \mathrm{Ta}^{1/2} Q^3 T_2, \tag{76}$$

$$\nabla^2 L^2 T_1 + m\,\mathrm{Ta}^{1/2} T_2 = \mathrm{Ta}^{1/2} Q^3 S_1, \tag{77}$$

$$\nabla^2 L^2 T_2 - m\,\mathrm{Ta}^{1/2} T_1 = \mathrm{Ta}^{1/2} Q^3 S_2, \tag{78}$$

where ∇^2, L^2, Q^3 are now to be understood to have their old meanings with $\partial^2/\partial\phi^2$ replaced by $-m^2$. It follows from equations (73) to (78) that

$$\mathrm{Ra}\int [(\mathrm{grad}\,\Theta_1)^2 + (\mathrm{grad}\,\Theta_2)^2]\,d\tau = \int [S_1\nabla^4 L^2 S_1 + S_2\nabla^4 L^2 S_2 - T_1\nabla^2 L^2 T_1 - T_2\nabla^2 L^2 T_2]\,d\tau \tag{79}$$

and this is identical with equation (71). It can easily be verified that this defines a variational principle which is self-adjoint. Expression (79) gives an extreme value of Ra for all independent variations $\delta\Theta_2$, δS_1 and δT_2 when equations (74), (75) and (78) are satisfied *provided* Θ_1, S_2 and T_1 are regarded as auxiliary variables *defined* from Θ_2, S_1 and T_2 by equations (73), (76) and (77) and the appropriate boundary conditions. (The same is true, *mutatis mutandis*, if Θ_2, S_1 and T_2 are regarded as subsidiary variables defined from Θ_1, S_2 and T_1.)

Of course, as it stands the determination of Θ_1, S_2 and T_1 from equations (73), (76) and (77) involves a prior knowledge of Ra. However, this is simply rectified by a change of variable which alters the auxiliary problem into

$$\nabla^2\Theta_1 = -L^2S_1 \tag{80}$$

$$\nabla^2(\nabla^2L^2S_2 - mS_1) = L^2\Theta_2 - Q^3T_2 \tag{81}$$

$$\nabla^2L^2T_1 + mT_2 = Q^3S_1 \tag{82}$$

$$\Theta_1 = S_2 = T_1 = \partial S_2/\partial r = 0, \qquad \text{at} \qquad r = 1, \tag{83}$$

the characteristic value problem into

$$\nabla^2\Theta_2 = -\operatorname{Ra} L^2S_2 \tag{84}$$

$$\nabla^2(\nabla^2L^2S_1 + m\operatorname{Ta} S_2) = \operatorname{Ra} L^2\Theta_1 - \operatorname{Ta} Q^3T_1 \tag{85}$$

$$\nabla^2L^2T_2 - m\operatorname{Ta} T_1 = \operatorname{Ta} Q^3S_2 \tag{86}$$

$$\Theta_2 = S_1 = T_2 = \partial S_1/\partial r = 0, \qquad \text{at} \qquad r = 1, \tag{87}$$

and the basis of the variational principle into

$$\operatorname{Ra}\int(\operatorname{grad}\Theta_1)^2\,d\tau + \frac{\operatorname{Ta}}{\operatorname{Ra}}\int(\operatorname{grad}\Theta_2)^2\,d\tau \tag{88}$$

$$= \int(S_1\nabla^4L^2S_1 - T_2\nabla^2L^2T_2)\,d\tau + \operatorname{Ta}\int(S_2\nabla^4L^2S_2 - T_1\nabla^2L^2T_1)\,d\tau.$$

V. The Dynamo Problem for a Conducting Fluid of Infinite Extent

The characteristic value problem governing the existence of a steady self-maintained dynamo action in a fluid of infinite extent is

$$\mathscr{L}: \begin{cases} \operatorname{curl}\mathbf{B} = \operatorname{Rm}(-\operatorname{grad}\Phi + \mathbf{u}\times\mathbf{B}), & (89)\\ \operatorname{div}\mathbf{B} = 0, & (90) \end{cases}$$

and, at great distances (R),

$$\mathscr{B}: \quad \mathbf{B} = O(R^{-2}),\ R\to\infty, \tag{91}$$

(*not* $\mathbf{0}(R^{-3})$, since the fluid is unbounded). Here $\mathbf{B}$ is the magnetic field, Φ is the electrostatic potential measured in units of ULB (U and L are a characteristic velocity and length associated with the system), and $\mathbf{u}$ is the fluid velocity measured in units of U. We assume that $\mathbf{u}$ vanishes

as $R \to \infty$. The characteristic values Rm is sometimes called a "magnetic Reynolds number", and is defined by

$$\mathrm{Rm} = \mathrm{UL}/\lambda,$$

where λ is the "magnetic diffusivity" which, in electromagnetic units, is equal to $1/4\pi\sigma\mu$ (σ and μ are the conductivity and permeability of the fluid).

The adjoint problem is

$$\mathscr{L}^*: \begin{cases} \operatorname{curl} \mathbf{B}^* = \mathrm{Rm}^* (-\operatorname{grad} \Phi - \mathbf{u} \times \mathbf{B}^*), & (89^*) \\ \operatorname{div} \mathbf{B}^* = 0, & (90^*) \end{cases}$$

$$\mathscr{B}^*: \qquad \mathbf{B}^* = O(R^{-2}),\ R \to \infty, \qquad (91^*)$$

This can easily be proved. By the divergence theorem and equations (91), (91*), we have

$$\int \mathbf{B}^* \cdot \operatorname{curl} \mathbf{B}\, d\tau = \int \mathbf{B} \cdot \operatorname{curl} \mathbf{B}^*\, d\tau. \qquad (92)$$

But, by equations (89), and (90) and another application of the divergence theorem, the left-hand side of equation (92) is equal to

$$\mathrm{Rm} \int \mathbf{u} \cdot (\mathbf{B} \times \mathbf{B}^*)\, d\tau,$$

and, similarly, the right-hand side is equal to

$$\mathrm{Rm}^* \int \mathbf{u} \cdot (\mathbf{B} \times \mathbf{B}^*)\, d\tau$$

Thus, all the integrals concerned vanish unless Rm and Rm are equal, and then we have

$$\mathrm{Rm} = \frac{\int \mathbf{B}^* \cdot \operatorname{curl} \mathbf{B}\, d\tau}{\int \mathbf{u} \cdot (\mathbf{B} \times \mathbf{B}^*)\, d\tau}. \qquad (93)$$

The variational principle follows as before.

It is interesting to notice that, by proving that Rm and Rm* are equal, we have shown that, if dynamo action is possible in an infinite fluid with a certain velocity field $\mathbf{u}$, then it is also possible with the velocity field $-\mathbf{u}$. This is certainly not to be expected for a fluid of finite extent. It is not, therefore, surprising to find that, for a fluid dynamo of finite extent, equations (89*) to (91*) do not define a characteristic value problem that is adjoint to that of equations (89) to (91). In fact, there

seems to be no obvious way of constructing the adjoint problem in this case because of the discontinuity in curl **B** which is generally present at the surface of the fluid.

ACKNOWLEDGMENT

It is a pleasure to record my gratitude to Prof. S. Chandrasekhar for his helpful interest and advice.

REFERENCES

1. Chandrasekhar, S. *Phil. Mag.* (7th series), **43**, 1317–1329 (1952).
2. Chandrasekhar, S. *Am. Math. Monthly*, **61**, 32–45 (1954).
3. Chandrasekhar, S. *Mathematika*, **1**, 5–13 (1954).
4. Chandrasekhar, S. *Phil. Mag.* (8th series), **2**, 845–858 (1957).
5. Chandrasekhar, S. *Phil. Mag.* (8th series), **2**, 1282–1284 (1957).
6. Chandrasekhar, S. Max Planck Festschrift, 1958.

Reprinted without change of pagination from the
Journal of Fluid Mechanics, *volume* 17, *part* 1, *pp.* 1–20, 1963

On the motion of a liquid in a spheroidal cavity of a precessing rigid body

By K. STEWARTSON
Mathematics Research Center, University of Wisconsin, Madison, Wisconsin, U.S.A.†
AND P. H. ROBERTS
Yerkes Observatory, Williams Bay, Wisconsin, U.S.A.

(Received 17 December 1962 and in revised form 19 April 1963)

The flow set up in an oblate cavity of a precessing rigid body is examined under the assumptions that the ellipticity of the spheroidal boundary of the fluid is large compared with Ω/ω and that the boundary-layer thickness is small compared with the deviations of the boundary from sphericity ($\boldsymbol{\omega}$ is the angular velocity of the rigid body about the axis of symmetry, $\boldsymbol{\Omega}$ is the angular velocity with which this axis precesses).

The motion of the fluid is found by considering an initial-value problem in which the axis of rotation of the spheroid is impulsively moved at a time $t = 0$; before that time this axis is supposed to be fixed in space, the fluid and envelope turning about it as a solid body. The solution is divided into a steady motion and transients, and, by evaluating the effects of the viscous boundary layer, the transients are shown to decay with time. The steady motion which remains consists of a primary rigid-body rotation with the envelope, superimposed on which is a circulation with constant vorticity in planes perpendicular to $\boldsymbol{\omega} \times (\boldsymbol{\omega} \times \boldsymbol{\Omega})$, the streamlines being similar and similarly situated ellipses.

The possible effects of the luni-solar precession on the fluid motions in the Earth's core are discussed.

1. Introduction

Some years ago, Bondi & Lyttleton (1953) examined the motion of a liquid in a spherical cavity of a rigid body which was rotating with an angular velocity $\boldsymbol{\omega}$ about an axis through the centre of the sphere when this axis, in turn, was precessing with angular velocity $\boldsymbol{\Omega}$ ($\Omega \ll \omega$) about an axis fixed in space. The aim of their investigation was to throw some light on the effect of the precession of the Earth on the motion of the liquid core. The analytical approach used by Bondi & Lyttleton was to suppose that the fluid is Newtonian and practically inviscid. Then, in the first approximation, viscosity was neglected, and it was hoped to correct for viscosity by a thin boundary layer near the surface of the cavity. Unfortunately, on attempting to solve the steady inviscid equations, they arrived at a definite contradiction which strongly suggested to them that no steady-state motion of a permanent character is possible for the fluid.

† Present address: Department of Mathematics, The University of Durham, Durham.

The spherical cavity presents a special difficulty in the theory of rotating fluids because without viscosity no motion can be communicated to the fluid by a rotation of the boundary. The angular velocity acquired by the fluid, even if the sphere is not precessing, is due entirely to the generation of vorticity in the boundary layer at the surface of the cavity and one might expect it to take a time $O(a^2/\nu)$, where a is the radius of the cavity and ν the kinematic viscosity, for the motion of the fluid and sphere to rotate together as a rigid body. Let us suppose this time has passed, that the precessing motion is set up at $t = 0$, and that subsequently the fluid may be regarded as inviscid. Then while the solid body precesses the fluid rotates about a fixed axis, being unaffected by the motion of its envelope. The motion of the fluid is actually steady but relative to the envelope it appears to be unsteady, its axis of rotation precessing with angular velocity $-\mathbf{\Omega}$ and, when Ωt is small, this means that the velocity perturbations are proportional to Ωt. Thus the contradiction obtained by Bondi & Lyttleton does not necessarily mean that the motion of the fluid is not ultimately steady. For a real fluid, one might expect that the motion described above would occur if $\Omega a^2/\nu \gg 1$, and if, on the other hand, $\Omega a^2/\nu \ll 1$, the motion of the fluid and envelope would be virtually a rigid body precession.

This discussion strongly supports their view that the determination of the motion is likely to prove a problem of great difficulty, and consequently we thought it desirable to begin our attack by studying the related problem of the oblate spheroidal envelope which includes the sphere as a special case but proves easier because the motion of the boundary of the cavity must communicate itself in part to the fluid. This problem is not without interest in the geophysical context with which Bondi & Lyttleton were concerned because, although one can say that the core of the Earth is nearly a sphere in that

$$(a^2 - b^2)/a^2 \ll 1,$$

where a and b are the semi-major and minor axes of the core, one cannot be sure of the values of

$$R_1 = (a^2 - b^2)\,\omega/\Omega a^2, \quad R_2 = \Omega a^2/\nu,$$

which, with $R_3 = \omega a^2/\nu$ ($\gg 1$), control the motion of the fluid in the core relative to the rigid-body rotation. If $R_1 = 0$, the motion is critically dependent on R_2 while if R_1 is large the motion is independent of R_2 except, possibly, if $R_1 R_2 \lesssim 1$. Our investigation here is based on the assumptions that $R_1 \gg 1$, $R_1 R_2 \gg 1$; it is hoped to explore other limiting situations in a later paper. These assumptions are not unreasonable in the geophysical context since, for the surface of the Earth, $(a^2 - b^2)/2a^2 \simeq 1/297$, and the value of this parameter for the core is not likely to be many orders of magnitude less. Further, $\omega/\Omega \simeq 10^7$, and ν is almost certainly much less than $10^8\,\mathrm{cm}^2/\mathrm{sec}$. We shall consider these geophysical questions further in § 7.

Even for an oblate spheroidal cavity containing almost inviscid fluid the determination of the flow has some unusual features. Thus the governing equations for an inviscid fluid are hyperbolic while the boundary condition is a relation between the function and its normal derivative. Such a boundary condition is appropriate to an elliptic differential equation and there are no corresponding

existence and uniqueness theorems for hyperbolic equations. In fact one can construct examples of non-existence and non-uniqueness as follows.

Suppose that the governing equation is

$$\frac{\partial^2\phi}{\partial x^2}-\frac{\partial^2\phi}{\partial y^2}=0, \tag{1.1}$$

and that (*a*) $\phi = 0$ on the circle $x^2+y^2 = 1$, or (*b*) $\phi = 0$ on three sides of the square $|x| \leqslant 1$, $|y| \leqslant 1$ and $\phi = 1$ on the fourth side. One solution of (*a*) is

$$\phi = A(x^2+y^2-1)$$

inside the circle, where A is arbitrary. This solution can be added to any other, so that there is no unique solution of (1.1) in this case. Again, every solution of (1.1) which vanishes on three sides of a square must vanish on the fourth side so that the solution of problem (*b*) cannot exist.

Fortunately in the present problem there is a simple physical explanation of these phenomena. The non-uniqueness is associated with the existence of free oscillatory motions of the fluid in which the boundary remains fixed and the non-existence is associated with the resonance which can occur when the boundary is made to oscillate with the same period as that of one of the free oscillations of the fluid.

The difficulty about the inappropriate type of boundary condition is removed by considering motions which start from a position of relative rest. We suppose that the fluid and envelope are initially rotating as if solid with angular velocity $\boldsymbol{\omega}$ and that at time $t = 0$ the envelope is set precessing with angular velocity $\boldsymbol{\Omega}$. The equations governing the induced motion of the fluid are unsteady but the time dependence is removed on applying a Laplace transformation with parameter s and the equation is reduced to a single one of Laplace's form apart from a scaling factor which is a function of s. A general solution can be written down and, on inversion, the history of the motion can be traced. From an examination of the solution at large times we can relate the non-uniqueness to free oscillations set up by the initial motion and the non-existence to resonance.

For a spheroid the actual inviscid solution is straightforward being effectively given by Poincaré (1910); the difficulty found by Bondi & Lyttleton for $a = b$ is seen to be a resonance and explicable on the lines stated in physical terms at the beginning of this section.

From a geophysical standpoint an argument based on an initial motion is incomplete without considering the effect of dissipation which on the geological time scale might be expected to have a serious effect on the solution. In the inviscid solution predicted for a spheroid the velocity components are linear functions of position so that apparently there is no dissipation in the interior, it being confined to the boundary layer which must arise through violation of the no-slip condition. This is investigated in § 4. The inviscid motion outside consists of two parts: (i) dependent on the azimuth angle θ and due to the steady motion of the boundary, (ii) independent of θ and due to the angular velocity which the boundary acquired initially but failed to communicate immediately to the fluid. The boundary layer associated with (i) was studied by Bondi & Lyttleton but we repeat their investigation here since we are including time as a

parameter. It develops a singularity on certain circles of latitude as s approaches any value on the imaginary axis between $-3i$ and i. The boundary layer associated with (ii) has also been studied earlier by Proudman (1956) and Stewartson (1957), and it develops singularities as s approaches any value on the imaginary axis between $\pm 2i$.

Bondi & Lyttleton (1953) noticed that the normal velocity on the critical circle become infinite in the limit and suggested that the flow near it was unstable leading to turbulence here and elsewhere. Although one cannot be sure without carrying out an actual experiment, more recent investigations do not support this view. The singularity on a critical line is a singularity in the boundary-layer sense, i.e. if v_n is the normal velocity just outside the boundary layer, the equations imply that $v_n = O(\nu^{\frac{1}{4}})$ almost everywhere and on approaching the critical line $\nu^{-\frac{1}{2}} v_n \to \infty$. This does not necessarily mean that $v_n \to \infty$ and indeed, in a related problem (Roberts & Stewartson 1963), a more detailed investigation of the flow properties near the critical line implied $v_n = O(\nu^{\frac{2}{3}})$ which is still small in the limit $\nu \to 0$. Further, a detailed solution of a problem related to (ii) by Stewartson (1957) elucidates the role of the critical circles as the origin of shear layers which penetrate into the fluid, transporting fluid from one part of the boundary to another via the interior and also adjusting the angular velocity of the main body of fluid when required. Free shear layers are certainly notoriously prone to instability; however, after Proudman (1956) had postulated their existence in a discussion of the flow between two concentric spheres, experiments carried out by Fultz & Moore (1962) to test his conclusions did not reveal any serious instability. In fact, in the case when the outer sphere was rotated more rapidly than the inner sphere, a shear layer of the predicted structure was observed. It was noted however that, when the inner sphere was rotated the faster, the shear layer was much broader and more diffusive due, it appeared, to the formation of a street of line vortices. But, even in this case, the main characteristics of the flow far from the shear layer were not greatly affected.

In §§ 5 and 6 the (tertiary) modification to the secondary flow induced by these boundary layers is discussed with particular reference to the oscillations produced by the initial motion of the envelope. It is shown that, of the motions dependent on θ, only the steady motion survives as $t \to \infty$. The effect of the boundary layer independent of θ is to lead to a breakdown in the tertiary flow as $t \to \infty$, and it is argued that this must mean that, relative to the boundary, the motion independent of θ must die out as $t \to \infty$ due to the communication of vorticity to the fluid via the boundary layer.

Since the steady solution dependent on θ is solely determined by the residue of the pole of the Laplace transform of the velocity at $s = 0$, it is independent of the initial condition assumed. Consequently the motion which finally develops as a result of the joint action of the moving boundary and dissipation is the same whatever the initial condition assumed, and the non-uniqueness found when the steady problem is considered is seen to be illusory, i.e. any other solutions of the inviscid problem would undergo slow changes through the action of dissipation, ultimately taking on the form described above.

2. Equations of unsteady motion

We consider a mass of incompressible fluid which occupies the whole of a rigid envelope whose internal boundary, S_0, is a surface of revolution with axis of symmetry L_B. Initially the fluid and the envelope are rotating about the axis with angular velocity $\boldsymbol{\omega}$ which is an absolute constant. At time $t = 0$, the axis L_B is set rotating with a small uniform angular velocity $\boldsymbol{\Omega}$ about an axis L_S fixed in space which intersects L_B at a point O, and which is inclined at an angle α to it. For definiteness, we shall suppose that the perturbed motion is started impulsively so that $\boldsymbol{\Omega}$ is also an absolute constant. We wish to find the subsequent motion of the fluid.

Consider a reference frame, $\mathscr{F}$, rotating with angular velocity $\boldsymbol{\Omega}$, relative to which L_B and L_S are stationary, and let the velocity of the fluid relative to $\mathscr{F}$ be

$$\mathbf{w} = \mathbf{u} + \boldsymbol{\omega} \times \mathbf{r}. \tag{2.1}$$

At $t = 0-$, $\mathbf{u} = 0$, and at $t = 0+$, $\mathbf{u} \neq 0$ in virtue of the impulsive motion of S_0, but it will be determined by the initial motion of S_0. The actual velocity, $\mathbf{v}$, of the fluid, relative to axes fixed in space and instantaneously coinciding with $\mathscr{F}$, is

$$\mathbf{v} = \mathbf{u} + \boldsymbol{\omega} \times \mathbf{r} + \boldsymbol{\Omega} \times \mathbf{r}. \tag{2.2}$$

We know that, at $t = 0+$, $\mathbf{u} + \boldsymbol{\Omega} \times \mathbf{r}$ is irrotational and uniquely determined by the motion of S_0, and hence is also known at $t = 0+$. The equations of motion relative to $\mathscr{F}$ are

$$\partial \mathbf{v}/\partial t + \boldsymbol{\Omega} \times \mathbf{v} + (\mathbf{w} . \operatorname{grad}) \mathbf{v} = -\operatorname{grad}(W + p/\rho) + \nu \nabla^2 \mathbf{v}, \tag{2.3}$$

$$\operatorname{div} \mathbf{v} = 0, \tag{2.4}$$

where p is the pressure, ρ the density, ν the kinematic viscosity, and W the potential per unit mass of the external forces, supposed conservative, which act on the fluid. We now express (2.3) in terms of $\mathbf{u}$, obtaining

$$\partial \mathbf{u}/\partial t + 2(\boldsymbol{\omega} + \boldsymbol{\Omega}) \times \mathbf{u} - (\mathbf{u} + \boldsymbol{\omega} \times \mathbf{r}) \times \operatorname{curl} \mathbf{u}$$
$$= -\operatorname{grad}[W + (p/\rho) - \tfrac{1}{2}\{(\boldsymbol{\omega} + \boldsymbol{\Omega}) \times \mathbf{r}\}^2 + \mathbf{u} . (\boldsymbol{\omega} \times \mathbf{r}) + \tfrac{1}{2}\mathbf{u}^2] + \nu \nabla^2 \mathbf{u} + \mathbf{r} \times (\boldsymbol{\Omega} \times \boldsymbol{\omega}),$$

whence, on neglecting squares and products of $\mathbf{u}$ and $\boldsymbol{\Omega}$, we find

$$\partial \mathbf{u}/\partial t + 2\boldsymbol{\omega} \times \mathbf{u} - (\boldsymbol{\omega} \times \mathbf{r}) \times \operatorname{curl} \mathbf{u} = -\omega \operatorname{grad} V' + \nu \nabla^2 \mathbf{u} - 2(\boldsymbol{\Omega} . \mathbf{r})\, \boldsymbol{\omega}, \tag{2.5}$$

where

$$\omega V' = W + (p/\rho) + \tfrac{1}{2}\mathbf{v}^2 - (\boldsymbol{\omega} \times \mathbf{r})^2 + (\boldsymbol{\omega} . \mathbf{r})(\boldsymbol{\Omega} . \mathbf{r}) - 2(\boldsymbol{\Omega} . \boldsymbol{\omega})\, \mathbf{r}^2. \tag{2.6}$$

This equation differs from that given by Bondi & Lyttleton (1953), viz.

$$[\partial \mathbf{u}/\partial t] + 2\boldsymbol{\omega} \times \mathbf{u} - (\boldsymbol{\omega} \times \mathbf{r}) \times \operatorname{curl} \mathbf{u} = -\omega \operatorname{grad} V + \nu \nabla^2 \mathbf{u} - (\boldsymbol{\Omega} . \mathbf{r})\, \boldsymbol{\omega}, \tag{2.7}$$

where

$$\omega V = W + (p/\rho) + \tfrac{1}{2}\mathbf{v}^2 - (\boldsymbol{\omega} \times \mathbf{r})^2 + 2(\boldsymbol{\omega} . \mathbf{r})(\boldsymbol{\Omega} . \mathbf{r}) - \tfrac{5}{2}(\boldsymbol{\Omega} . \boldsymbol{\omega})\, \mathbf{r}^2.$$

The first term in (2.7) in brackets was omitted by Bondi & Lyttleton because they assumed that the motion is steady. Although both V and the forcing term in (2.7) are incorrect, in the present problem only the absence of the factor 2 in the forcing term is of significance. We shall show (§ 3) that this factor is crucial

for obtaining a consistent solution of the problem when S_0 is a sphere. The equation of continuity (2.4) is equivalent to

$$\operatorname{div} \mathbf{u} = 0. \tag{2.8}$$

Now choose a system of cylindrical polar co-ordinates (r, θ, z) in $\mathscr{F}$ with L_B as the z-axis, r measuring the distance of a typical point $P(\mathbf{r})$ from L_B, and θ being the angle between the plane defined by P and L_B and the plane defined by L_S and L_B. Then

$$(\boldsymbol{\Omega} . \mathbf{r})\, \boldsymbol{\omega} = (0, 0, -\Omega\omega z \cos\alpha - \Omega\omega r \sin\alpha \cos\theta), \tag{2.9}$$

where $\Omega = -|\boldsymbol{\Omega}|$ is the (retrograde) angular velocity of precession.

To begin with, neglect viscosity, and denote the velocity by $\mathbf{u}_0$ with components $(u_{0r}, u_{0\theta}, u_{0z})$ respectively along the directions of r, θ and z increasing; (2.5) and (2.8) are then equivalent to

$$\left.\begin{aligned}
&\frac{\partial u_{0r}}{\partial t} + \omega \frac{\partial u_{0r}}{\partial \theta} - 2\omega u_{0\theta} = -\omega \frac{\partial V''}{\partial r}, \\
&\frac{\partial u_{0\theta}}{\partial t} + \omega \frac{\partial u_{0\theta}}{\partial \theta} + 2\omega u_{0r} = -\frac{\omega}{r} \frac{\partial V''}{\partial \theta}, \\
&\frac{\partial u_{0z}}{\partial t} + \omega \frac{\partial u_{0z}}{\partial \theta} = -\omega \frac{\partial V''}{\partial z} + 2\omega\Omega r \sin\alpha \cos\theta, \\
&\frac{1}{r} \frac{\partial}{\partial r}(r u_{0r}) + \frac{1}{r} \frac{\partial u_{0\theta}}{\partial \theta} + \frac{\partial u_{0z}}{\partial z} = 0,
\end{aligned}\right\} \tag{2.10}$$

where

$$V'' = V' - \Omega z^2 \cos\alpha - r u_{0\theta}.$$

The boundary conditions require that the normal component of the fluid velocity is zero on S_0, and also that, at $t = 0+$, $\mathbf{u} + \boldsymbol{\Omega} \times \mathbf{r}$ is irrotational with velocity potential ϕ_i (say).

In order to solve (2.10) it is convenient to take the Laplace transform of the dependent variables with respect to time denoting the result by an asterisk, e.g.

$$u_{0r}^*(r, \theta, z, s) = \int_0^\infty e^{-\omega s t} u_{0r}(r, \theta, z, t)\, dt. \tag{2.11}$$

The governing equations then reduce to

$$\left.\begin{aligned}
&\left(\frac{\partial}{\partial \theta} + s\right) u_{0r}^* - 2u_{0\theta}^* = -\frac{\partial V''^*}{\partial r} + \frac{1}{\omega}(u_{0r})_i, \\
&\left(\frac{\partial}{\partial \theta} + s\right) u_{0\theta}^* + 2u_{0r}^* = -\frac{1}{r} \frac{\partial V''^*}{\partial \theta} + \frac{1}{\omega}(u_{0\theta})_i, \\
&\left(\frac{\partial}{\partial \theta} + s\right) u_{0z}^* = -\frac{\partial V''^*}{\partial z} + \frac{1}{\omega}(u_{0z})_i + \frac{2\Omega \sin\alpha}{\omega s} r \cos\theta,
\end{aligned}\right\} \tag{2.12}$$

where the suffix i denotes initial conditions. Since, at $t = 0+$, $\mathbf{u} + \boldsymbol{\Omega} \times \mathbf{r}$ is derivable from the potential ϕ_i, equations (2.12) may be simplified, on writing

$$V^* = V''^* - \omega^{-1}\phi_i + \Omega\omega^{-1} r z \sin\alpha \sin\theta, \tag{2.13}$$

to

$$\left.\begin{aligned}\left(\frac{\partial}{\partial\theta}+s\right)u^*_{0r}-2u^*_{0\theta}&=-\frac{\partial V^*}{\partial r},\\ \left(\frac{\partial}{\partial\theta}+s\right)u^*_{0\theta}+2u^*_{0r}&=-\frac{1}{r}\frac{\partial V^*}{\partial\theta}+\frac{\Omega}{\omega}r\cos\alpha,\\ \left(\frac{\partial}{\partial\theta}+s\right)u^*_{0z}&=-\frac{\partial V^*}{\partial z}+\frac{2\Omega}{\omega s}r\sin\alpha(\cos\theta+s\sin\theta),\end{aligned}\right\}\tag{2.14}$$

and the equation of continuity to

$$\frac{1}{r}\frac{\partial}{\partial r}(ru^*_{0r})+\frac{1}{r}\frac{\partial u^*_{0\theta}}{\partial\theta}+\frac{\partial u^*_{0z}}{\partial z}=0.\tag{2.15}$$

A further simplification is obtained by writing

$$\left.\begin{aligned}u^*_{0r}&=\mathscr{R}(\bar{u}_{0r}e^{i\theta}),\quad u^*_{0\theta}=(\Omega r\cos\alpha/\omega s)+\mathscr{R}(\bar{u}_{0\theta}e^{i\theta}),\\ u^*_{0z}&=\mathscr{R}(\bar{u}_{0z}e^{i\theta}),\quad V^*=(\Omega r^2\cos\alpha/\omega s)+\mathscr{R}(\bar{V}e^{i\theta}),\end{aligned}\right\}\tag{2.16}$$

where $\bar{u}_{0r}$, $\bar{u}_{0\theta}$, $\bar{u}_{0z}$, and $\bar{V}$ are independent of θ. On substituting into (2.14) and (2.15), we obtain

$$\left.\begin{aligned}\bar{u}_{0r}&=-\frac{1}{(s+i)^2+4}\left[(s+i)\frac{\partial\bar{V}}{\partial r}+\frac{2i\bar{V}}{r}\right],\\ \bar{u}_{0\theta}&=-\frac{i}{(s+i)^2+4}\left[2i\frac{\partial\bar{V}}{\partial r}+(s+i)\frac{\bar{V}}{r}\right],\\ \bar{u}_{0z}&=-\frac{2i\Omega\sin\alpha}{\omega s}r-\frac{1}{s+i}\frac{\partial\bar{V}}{\partial z},\end{aligned}\right\}\tag{2.17}$$

where $\bar{V}$ satisfies

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial\bar{V}}{\partial r}\right)-\frac{\bar{V}}{r^2}+\frac{[(s+i)^2+4]}{(s+i)^2}\frac{\partial^2\bar{V}}{\partial z^2}=0.\tag{2.18}$$

If S_0 is given by $f(r,z)=1$, the boundary condition to be satisfied is

$$\bar{u}_{0r}\frac{\partial f}{\partial r}+\bar{u}_{0z}\frac{\partial f}{\partial z}=0\quad\text{at}\quad f=1.\tag{2.19}$$

3. The secondary inviscid flow for a spheroid

The problem posed by equations (2.17) to (2.19) can be reduced to a solution of Laplace's equation on writing $[(s+i)^2+4]/(s+i)^2=\xi^2$, and replacing z by ξz. We can expect, therefore, that it can always be solved for the boundary condition (2.19), which is equivalent to a linear relation between $\bar{V}$, its normal derivative, and possibly also its tangential derivative. Having obtained the solution, the ultimate flow may be found by letting $s\to 0$, and taking note of the poles and branch points in the solution in the half plane $\mathscr{R}(s)\geqslant 0$.

In this paper we are particularly concerned with the properties of the solution when the envelope S_0 is the oblate spheroid

$$\frac{r^2}{a^2}+\frac{z^2}{b^2}=1\quad(a\geqslant b),\tag{3.1}$$

of which Bondi & Lyttleton considered the special case $a = b$. For the boundary (3.1), (2.19) becomes

$$\frac{r}{a^2}\bar{u}_{0r}+\frac{z}{b^2}\bar{u}_{0z} = 0, \tag{3.2}$$

and the appropriate solution of (2.18) is found by writing

$$\bar{V} = Arz, \tag{3.3}$$

where A is independent of r and z. Substituting into (2.17) we have

$$\bar{u}_{0r} = -\frac{Az}{(s-i)}, \quad \bar{u}_{0\theta} = -\frac{iAz}{(s-i)}, \quad \bar{u}_{0z} = -\frac{iA}{(s+i)}-\frac{2i\Omega r\sin\alpha}{\omega s}. \tag{3.4}$$

Hence, from (3.2),

$$A = -\frac{2i\Omega a^2\sin\alpha(s^2+1)}{\omega s[a^2(s-i)+b^2(s+i)]}. \tag{3.5}$$

Inverting the solution with respect to s, and using (2.16), we have

$$u_{0r} = \frac{2\Omega z a^2\sin\alpha}{a^2-b^2}\left[\sin\theta-\frac{2a^2}{a^2+b^2}\sin(\theta+\omega kt)\right], \tag{3.6}$$

$$u_{0\theta} = \frac{2\Omega z a^2\sin\alpha}{a^2-b^2}\left[\cos\theta-\frac{2a^2}{a^2+b^2}\cos(\theta+\omega kt)\right]+\Omega r\cos\alpha, \tag{3.7}$$

$$u_{0z} = -\frac{2\Omega r b^2\sin\alpha}{a^2-b^2}\left[\sin\theta-\frac{2a^2}{a^2+b^2}\sin(\theta+\omega kt)\right], \tag{3.8}$$

where $$k = (a^2-b^2)/(a^2+b^2).$$

According to (3.6) to (3.8), the secondary motion of the fluid consists of two parts: one is essentially due to the initial motion of the boundaries; the other is an ultimately steady motion. The first part can itself be divided into two components. One is a rigid body rotation of angular velocity $\Omega\cos\alpha$ about L_B which the rotation of the boundary *failed* to communicate to the fluid. The other is a free oscillation of the fluid. The steady part of the motion consists of the rigid body rotation $\boldsymbol{\omega}$ (cf. equation (2.1)), and a circulation (given by the terms in $\sin\theta$, $\cos\theta$ and $\sin\theta$ in (3.6), (3.7) and (3.8), respectively) in planes perpendicular to L_S the streamlines being similar and similarly situated ellipses. The vorticity of this circulation is constant and equal to $-2\Omega\sin\alpha(a^2+b^2)/(a^2-b^2)$. We shall show below that, if the fluid has a small viscosity, the effect of viscous dissipation in the boundary layer is to damp out all motions except the rigid body rotation and the circulation in planes perpendicular to L_S. Since the choice of initial conditions affects the remaining terms only, it follows that the ultimate motion of the fluid is unique and that the effect of a small viscosity is to convert an improperly posed mathematical problem into a properly posed one.

In the case of a sphere $a = b$, the relevant periods of free oscillation of the fluid are infinite, so that a resonance develops and

$$\left.\begin{aligned} u_{0r} &= -\Omega z\sin\alpha(\sin\theta+\omega t\cos\theta),\\ u_{0\theta} &= \Omega z\sin\alpha(\cos\theta+\omega t\sin\theta)+\Omega r\cos\alpha,\\ u_{0z} &= \Omega r\sin\alpha(\sin\theta+\omega t\cos\theta),\end{aligned}\right\} \tag{3.9}$$

or, in vector notation,

$$\mathbf{u} = -\boldsymbol{\Omega} \times \mathbf{r} - t(\boldsymbol{\Omega} \times \boldsymbol{\omega}) \times \mathbf{r}. \tag{3.10}$$

The physical interpretation of this result is that, as one expects, the motion of a spherical boundary exerts no influence on the motion of the fluid and hence, relative to the frame $\mathscr{F}$, the axis of rotation of the fluid rotates with angular velocity $-\boldsymbol{\Omega}$ about L_S. Consequently, after time t ($\Omega t \ll 1$), the fluid rotates about an axis making an angle

$$|\boldsymbol{\Omega} t \times \boldsymbol{\omega}|/|\boldsymbol{\omega}|,$$

with L_B, and this is indicated by (3.10). It is noted that, on substituting (3.10) into (2.7), the equations of motion are satisfied identically, and so any change in the motion of the fluid must be initiated through a boundary layer arising from the non-satisfaction of the requirement that $\mathbf{u} = 0$ on the boundary S_0. Had the equation (2.7), required by Bondi & Lyttleton, been used, t would be replaced by $\frac{1}{2}t$ in (3.10), and the new formula for $\mathbf{u}$ would not be consistent with this physical argument.

4. The boundary layer

The inviscid solution obtained in the previous section satisfies not only the inviscid equations but also the viscous equations (2.5). Viscosity therefore manifests itself only through the fact that this solution does not satisfy the viscous boundary condition $\mathbf{u} = 0$ on the spheroidal envelope S_0, and leads to a boundary layer whose thickness we can anticipate to be $O(\nu^{\frac{1}{2}})$ if ν is small where the adjustments in the tangential components of velocity are made. We shall calculate this boundary layer in the present section and its effect on the inviscid flow in the interior will be discussed in the following sections.

Let us write

$$\mathbf{u} = \mathbf{u}_0 + \mathbf{u}_1, \tag{4.1}$$

where $\mathbf{u}$ is the actual velocity of the fluid at any point, $\mathbf{u}_0$ is the velocity if $\nu = 0$, and $\mathbf{u}_1$ is the correction due to viscosity. Then, to order $\nu^{\frac{1}{2}}$, $\mathbf{u}_1$ is the contribution from the boundary layer to $\mathbf{u}$. Introduce spheroidal polar co-ordinates $(\lambda_1, \mu_1, \theta)$, where λ_1 and μ_1 are defined by

$$r = (\lambda_1^2 + c^2)^{\frac{1}{2}}(1-\mu_1^2)^{\frac{1}{2}}, \quad z = \lambda_1 \mu_1, \tag{4.2}$$

$c^2 = a^2 - b^2$ and the spheroid S_0 is given by $\lambda_1 = b$. Then, using (3.6) to (3.8), we see that the boundary condition $\mathbf{u} = 0$ requires that, at $\lambda_1 = b$,

$$\left.\begin{aligned} u_{1\lambda_1} &= 0, \\ u_{1\mu_1} &= \frac{2\Omega a b \sin\alpha}{a^2-b^2}\left[\sin\theta - \frac{2a^2}{a^2+b^2}\sin(\theta+\omega k t)\right](b^2+\mu_1^2 c^2)^{\frac{1}{2}}, \\ u_{1\theta} &= -\frac{2\Omega a^2 b \sin\alpha}{a^2-b^2}\left[\cos\theta - \frac{2a^2}{a^2+b^2}\cos(\theta+\omega k t)\right]\mu_1 - \Omega a \cos\alpha(1-\mu_1^2)^{\frac{1}{2}}. \end{aligned}\right\} \tag{4.3}$$

Beyond the boundary layer, when $(b-\lambda_1)\nu^{\frac{1}{2}}$ is large and positive, $\mathbf{u}_1 = o(1)$; we shall then be particularly interested in $u_{1\lambda}$ which is $O(\nu^{\frac{1}{2}})$ and, unlike the other components of velocity, is not exponentially small. Consequently it engenders

a tertiary flow though the interior of S_0. From the boundary condition (4.3) the flow in the boundary layer may be divided into two parts one of which ($\mathbf{u}_2$) is dependent on θ, and one ($\mathbf{u}_3$) which is independent of θ. We consider the part dependent on θ first.

Write the Laplace transform of $\mathbf{u}_2$ with respect to ωt as

$$\mathscr{R}(\bar{\mathbf{u}}_2 e^{i\theta}), \tag{4.4}$$

and in the boundary layer make the conventional assumption that the operator $\partial/\partial\lambda_1 = O(\nu^{-\frac{1}{2}})$ while $\partial/\partial\mu_1$ and $\partial/\partial\theta$ are $O(1)$. Then as is usual with Ekman flows, in the governing equations θ and μ_1 may be taken constant and λ_1 may be replaced by b except when it appears in $\partial/\partial\lambda_1$. The governing equations reduce to

$$(s+i)\,\bar{u}_{2\mu_1} + \frac{2\mu_1 a}{(b^2+c^2\mu_1^2)^{\frac{1}{2}}}\,\bar{u}_{2\theta} = \frac{a^2\nu}{\omega(b^2+c^2\mu_1^2)}\,\frac{\partial^2\bar{u}_{2\mu_1}}{\partial\lambda_1^2}, \tag{4.5}$$

$$(s+i)\,\bar{u}_{2\theta} - \frac{2\mu_1 a}{(b^2+c^2\mu_1^2)^{\frac{1}{2}}}\,\bar{u}_{2\mu_1} = \frac{a^2\nu}{\omega(b^2+c^2\mu_1^2)}\,\frac{\partial^2\bar{u}_{2\theta}}{\partial\lambda_1^2}, \tag{4.6}$$

$$a(b^2+c^2\mu_1^2)^{\frac{1}{2}}\frac{\partial\bar{u}_{2\lambda_1}}{\partial\lambda_1} + \frac{\partial}{\partial\mu_1}\left[(1-\mu_1^2)^{\frac{1}{2}}(b^2+c^2\mu_1^2)^{\frac{1}{2}}\,\bar{u}_{2\mu_1}\right] + \frac{i\bar{u}_{2\theta}(b^2+c^2\mu_1^2)}{a(1-\mu_1^2)^{\frac{1}{2}}} = 0, \tag{4.7}$$

and the boundary conditions to

$$\bar{u}_{2\lambda_1} = 0, \quad \bar{u}_{2\mu_1} = \frac{2i\Omega ab\sin\alpha(b^2+c^2\mu_1^2)^{\frac{1}{2}}(s+i)}{\omega s[s(a^2+b^2)-i(a^2-b^2)]}, \quad \bar{u}_{2\theta} = \frac{2\Omega a^2 b\sin\alpha(s+i)\,\mu_1}{\omega s[s(a^2+b^2)-i(a^2-b^2)]},$$

at $\lambda_1 = b$. Also $\bar{u}_{2\mu_1}$ and $\bar{u}_{2\theta} \to 0$ as $(b-\lambda_1)\,\nu^{\frac{1}{2}} \to \infty$.

Integrating (4.5), (4.6) we have

$$\left.\begin{aligned} u_{2\mu_1} &= \frac{\Omega abi\sin\alpha(s+i)}{\omega s[s(a^2+b^2)-i(a^2-b^2)]} \quad \cdot \\ &\quad \times\{[a\mu_1+\sqrt{(b^2+c^2\mu_1^2)}]\,e^{\delta_2(\lambda_1-b)} - [a\mu_1-\sqrt{(b^2+c^2\mu_1^2)}]\,e^{\delta_1(\lambda_1-b)}\}, \\ u_{2\theta} &= \frac{\Omega ab\sin\alpha(s+i)}{\omega s[s(a^2+b^2)-i(a^2-b^2)]} \\ &\quad \times\{[a\mu_1+\sqrt{(b^2+c^2\mu_1^2)}]\,e^{\delta_2(\lambda_1-b)} + [a\mu_1-\sqrt{(b^2+c^2\mu_1^2)}]\,e^{\delta_1(\lambda_1-b)}\}, \end{aligned}\right\} \tag{4.8}$$

where

$$\left.\begin{aligned} \delta_2^2 &= \frac{\omega}{\nu a^2}(b^2+c^2\mu_1^2)\left[s+i-\frac{2i\mu_1 a}{\sqrt{(b^2+c^2\mu_1^2)}}\right], \\ \delta_1^2 &= \frac{\omega}{\nu a^2}(b^2+c^2\mu_1^2)\left[s+i+\frac{2i\mu_1 a}{\sqrt{(b^2+c^2\mu_1^2)}}\right], \end{aligned}\right\} \tag{4.9}$$

and the signs of the square root of the expression for δ_1 and δ_2 are decided by requiring that, when s is large, real and positive, δ_1 and δ_2 are both positive. The value of $u_{2\lambda_1}$ now follows by the direct integration of (4.7) so that all properties of the boundary layer are now formally known. It is noted that the assumptions on which this boundary-layer theory is built are consistent provided only that δ_1 and δ_2 do not vanish. Exceptional cases arise only if s is purely imaginary and $|s| < 3$, when the assumptions break down on certain circles $\mu_2 = \text{const}$. Conceivably this may be serious at large times since singularities occur in the s-plane on the imaginary axis; on inverting we could get contributions to $\mathbf{u}_2$ of order

$t^n e^{is\sigma t}$ ($n > 0$). However, it is believed that this does not happen. In a closely related problem (Roberts & Stewartson 1963) the flow has been studied in the neighbourhood of these lines, and it has been shown that the main effect is to change the thickness of the layer from $O(\nu^{\frac{1}{4}})$ to $O(\nu^{\frac{1}{5}})$. Adapting the argument to the present problem similar results are obtained, and it is found that, in the neighbourhood of the singular lines, $u_{2\lambda_1} = O(\nu^{\frac{1}{5}})$, instead of $O(\nu^{\frac{1}{4}})$ as elsewhere in the boundary layer. Further, on computing the contribution from the neighbourhood of these lines to the flow in the deep interior of the spheroid, we find it to be $O(\nu^{\frac{1}{5}})$ as against $O(\nu^{\frac{1}{4}})$ from the rest of the boundary layer.

Of particular interest is the value of $\bar{u}_{2\lambda_1}$ just outside the boundary layer, i.e. when $(b-\lambda_1)\,\nu^{\frac{1}{5}}$ is large, but $(b-\lambda_1)$ is small. This is obtained by integrating (4.7) with respect to λ_1 from $-\infty$ to b and using (4.9). Denoting the value of $\bar{u}_{2\lambda_1}$ by $\bar{u}_{4\lambda_1}(b, \mu_1, \theta)$ we have

$$\begin{aligned} a(b^2+c^2\mu_1^2)^{\frac{1}{2}}\,\bar{u}_{4\lambda_1}(b,\mu_1,\theta) &= \frac{\Omega abi\sin\alpha(s+i)}{\omega s[s(a^2+b^2)-i(a^2-b^2)]}\left[\frac{\partial}{\partial\mu_1}\left\{(1-\mu_1^2)^{\frac{1}{2}}\,(b^2+c^2\mu_1^2)^{\frac{1}{2}}\right.\right. \\ &\quad\times\left[\frac{a\mu_1+\sqrt{(b^2+c^2\mu_1^2)}}{\delta_2}-\frac{a\mu_1-\sqrt{(b^2+c^2\mu_1^2)}}{\delta_1}\right]\Big\}+\frac{(b^2+c^2\mu_1^2)}{a(1-\mu_1^2)^{\frac{1}{2}}} \\ &\quad\times\left.\left\{\frac{a\mu_1+(b^2+c^2\mu_1^2)^{\frac{1}{2}}}{\delta_2}+\frac{a\mu_1-(b^2+c^2\mu_1^2)^{\frac{1}{2}}}{\delta_1}\right\}\right]. \end{aligned} \tag{4.10}$$

Now let us consider the second part ($\mathbf{u}_3$) of the boundary layer due to that part of (4.3) independent of θ. Let the Laplace transform of $\mathbf{u}_3$ with respect to ωt be $\mathbf{u}_3^*$. Then at $\lambda_1 = b$

$$u^*_{3\lambda_1} = 0, \quad u^*_{3\mu_1} = 0, \quad u^*_{3\theta} = -\Omega a\cos\alpha(1-\mu_1^2)^{\frac{1}{2}}/\omega s, \tag{4.11}$$

and by an analogous argument to that given above the equations governing the flow in the boundary layer reduce to

$$su^*_{3\mu_1}+\frac{2\mu_1 a}{(b^2+c^2\mu_1^2)^{\frac{1}{2}}}u^*_{3\theta} = \frac{a^2\nu}{\omega(b^2+c^2\mu_1^2)}\frac{\partial^2 u^*_{3\mu_1}}{\partial\lambda_1^2}, \tag{4.12}$$

$$su^*_{3\theta}-\frac{2\mu_1 a}{(b^2+c^2\mu_1^2)^{\frac{1}{2}}}u^*_{3\mu_1} = \frac{a^2\nu}{\omega(b^2+c^2\mu_1^2)}\frac{\partial^2 u^*_{3\theta}}{\partial\lambda_1^2}, \tag{4.13}$$

$$a(b^2+c^2\mu_1^2)^{\frac{1}{2}}\frac{\partial u^*_{3\lambda_1}}{\partial\lambda_1}+\frac{\partial}{\partial\mu_1}\left[(1-\mu_1^2)^{\frac{1}{2}}\,(b^2+c^2\mu_1^2)^{\frac{1}{2}}\,u^*_{3\mu_1}\right] = 0. \tag{4.14}$$

Integrating (4.12) and (4.13) we have

$$\left.\begin{aligned} u^*_{3\mu_1} &= -\frac{\Omega ai\cos\alpha(1-\mu_1^2)^{\frac{1}{2}}}{2\omega s}\left[e^{\delta_3(\lambda_1-b)}-e^{\delta_4(\lambda_1-b)}\right], \\ u^*_{3\theta} &= -\frac{\Omega a\cos\alpha(1-\mu_1^2)^{\frac{1}{2}}}{2\omega s}\left[e^{\delta_3(\lambda_1-b)}+e^{\delta_4(\lambda_1-b)}\right], \end{aligned}\right\} \tag{4.15}$$

where

$$\left.\begin{aligned} \delta_3^2 &= \frac{\omega}{\nu a^2}(b^2+c^2\mu_1^2)\left(s-\frac{2i\mu_1 a}{(b^2+c^2\mu_1^2)^{\frac{1}{2}}}\right), \\ \delta_4^2 &= \frac{\omega}{\nu a^2}(b^2+c^2\mu_1^2)\left(s+\frac{2i\mu_1 a}{(b^2+c^2\mu_1^2)^{\frac{1}{2}}}\right), \end{aligned}\right\} \tag{4.16}$$

and δ_3 and δ_4 are real and positive when s is real, positive and large. The consistency of this solution is subject to the same qualifications as (4.10). Of particular interest here too is the value of $u^*_{3\lambda_1}$ just outside the boundary layer, and denoting it by $u^*_{5\lambda_1}(b, \mu_1, \theta)$ we have

$$a(b^2+c^2\mu_1^2)^{\frac{1}{2}}\, u^*_{5\lambda_1} = -\frac{\Omega a i \cos\alpha}{2\omega s}\frac{\partial}{\partial\mu_1}\left[(b^2+c^2\mu_1^2)^{\frac{1}{2}}(1-\mu_1^2)^{\frac{1}{2}}\left(\frac{1}{\delta_3}-\frac{1}{\delta_4}\right)\right]. \quad (4.17)$$

5. The tertiary inviscid flow (i)

The boundary-layer solution obtained in § 4 shows that there must be a normal velocity at the edge of the layer (i.e. as $\lambda_1 \to b-$ on the inviscid scale) of order $\nu^{\frac{1}{2}}$ and given by the sum of (4.10) and (4.17). In turn this must induce a tertiary inviscid flow throughout the interior of the spheroid, i.e. a flow governed by (2.14) and (2.15) excluding the forcing terms, and satisfying (2.19) with a right-hand side effectively equal to the sum of (4.10) and (4.17) (instead of being zero). Since (4.10) is dependent on θ while (4.17) is not, it is convenient to treat their contributions separately and in this section we shall consider the consequences of (4.10). In § 6 below we shall consider the consequences of (4.17).

Denoting the corresponding velocity in the spheroid by $\mathbf{u}_4$ we have, from (2.17), that $\mathbf{u}_4$ can be expressed in terms of a scalar $\overline{V}_4$ which satisfies (2.18), the forcing term in (2.17) being again set equal to zero. Hence, after stretching the z co-ordinate, $\overline{V}_4$ satisfies Laplace's equation and the appropriate solution can be formally written down. Let

$$r = \frac{2}{[(s+i)^2+4]^{\frac{1}{2}}}(1-\mu_4^2)^{\frac{1}{2}}(\lambda_4^2+\gamma_4^2)^{\frac{1}{2}}, \quad z = \frac{2}{(s+i)}\mu_4\lambda_4, \quad (5.1)$$

where

$$\gamma_4^2 = a^2+\tfrac{1}{4}c^2(s+i)^2. \quad (5.2)$$

In terms of μ_4 and λ_4 the spheroid is given by

$$\lambda_4 = \tfrac{1}{2}(s+i)\,b \quad (5.3)$$

and on the spheroid $\mu_4 = \mu_1 = \mu$. The most general acceptable form for $\overline{V}_4$ is

$$\overline{V}_4 = (1-\mu_4^2)^{\frac{1}{2}}(\lambda_4^2+\gamma_4^2)^{\frac{1}{2}}\sum_{n=1}^{\infty} A_n P_n'(i\lambda_4/\gamma_4)\,P_n'(\mu_4), \quad (5.4)$$

the P_n being Legendre polynomials and the A_n constants to be found. Further, the boundary condition associated with (4.10), viz.

$$\frac{r}{a^2}u_{4r}+\frac{z}{b^2}u_{4z} = \left(\frac{r^2}{a^4}+\frac{z^2}{b^4}\right)^{\frac{1}{2}} u_{4\lambda_1}(b,\mu,\theta),$$

becomes, in terms of $\overline{V}_4$,

$$\frac{a}{2}\frac{\partial\overline{V}_4}{\partial\lambda_4}+\frac{2ib}{a[(s+i)^2+4]}\overline{V}_4 = (b^2+c^2\mu^2)^{\frac{1}{2}}\,\overline{u}_{4\lambda_1} \quad (5.5)$$

at $\lambda_4 = \frac{1}{2}(s+i)\,b$. It follows from (4.10) that $\overline{u}_{4\lambda_1}$ is an odd function of μ and behaves like $(1-\mu^2)^{\frac{1}{2}}$ near $\mu = \pm 1$. Consequently n is even in (5.4), and the various A_n are determined from the equation

$$A_n\left\{\frac{a}{2}\frac{\partial}{\partial\lambda_4}\left[(\lambda_4^2+\gamma_4^2)^{\frac{1}{2}}P_n'\left(\frac{i\lambda_4}{\gamma_4}\right)\right]+\frac{2ib(\lambda_4^2+\gamma_4^2)^{\frac{1}{2}}}{a[(s+i)^2+4]}P_n'\left(\frac{i\lambda_4}{\gamma_4}\right)\right\} = B_n, \tag{5.6}$$

where $\lambda_4 = \frac{1}{2}(s+i)\,b$ and

$$\frac{2n(n+1)}{2n+1}B_n = \int_{-1}^{+1}(b^2+c^2\mu^2)^{\frac{1}{2}}\,\overline{u}_{4\lambda_1}P_n'(\mu)\,(1-\mu^2)^{\frac{1}{2}}\,d\mu. \tag{5.7}$$

From this point on, the determination of V_4 is in principle straightforward but involves the evaluation of complicated inversion and other integrals. However, since ν is assumed to be small the effect of the tertiary flow may be neglected unless the A_n have poles, *qua* functions of s, in the half plane $\mathscr{R}(s) > 0$, or sufficiently strong singularities on the imaginary axis of s. Our purpose in the rest of this section is to give arguments to exclude such possibilities.

First we show that A_n has no singularities in the half plane $\mathscr{R}(s) > 0$. A singularity can arise if either the coefficient of A_n in (5.6) vanishes or if B_n is singular, or if (2.18) breaks down. Taking these possibilities in order, the coefficient of A_n vanishes if

$$(s-i)\,bP_n'(\psi)+2i\gamma_4 n(n+1)\,P_n(\psi) = 0, \tag{5.8}$$

using the properties of Legendre polynomials and writing

$$\psi = i(s+i)\,b/[4a^2+c^2(s+i)^2]^{\frac{1}{2}}. \tag{5.9}$$

Writing $s = i\sigma$, (5.8) reduces to

$$F(\sigma) \equiv (1-\sigma)\,\psi P_n'(\psi)+(1+\sigma)\,n(n+1)\,P_n(\psi) = 0, \tag{5.10}$$

where

$$\psi = -(1+\sigma)\,b/[4a^2-c^2(1+\sigma)^2]^{\frac{1}{2}}; \tag{5.11}$$

we shall now show that the roots of (5.10) occur at real values of σ. We observe that physically this result is to be expected because it means that the periods of free oscillation of the fluid in the spheroidal envelope are real. In order to prove this result, we note that (5.10) is effectively a polynomial of degree $n+1$ in σ and hence it is enough to prove that $F(\sigma)$ has $n+1$ zeros on the real axis of σ. One such zero is clearly at $\sigma = -1$, where $\psi = 0$. Further, between $\sigma = -1$ and $\sigma = +1$, where $\psi = 1$, P_n has $\frac{1}{2}n$ zeros and consequently, from the interlacing property of the zeros of P_n and P_n', $F(\sigma)$ has $\frac{1}{2}n-1$ zeros between the first positive zero of P_n, *qua* function of ψ, and $\sigma = 1$. A similar remark applies to $-3 < \sigma < -1$. Again if $\frac{1}{2}n$ is even, $F(\sigma) > 0$ if $\sigma+1$ is small and positive, and is negative at the first zero of P_n for which $\sigma > -1$. Hence $F(\sigma)$ vanishes in this range too. A similar argument applies if $\frac{1}{2}n$ is odd. Counting up we see that there are exactly $n+1$ zeros of $F(\sigma)$ in $-3 < \sigma < 1$, all simple, and leading to finite oscillations of the fluid except when they coincide with poles of B_n.

Now consider the singularities of B_n. There are two simple poles at $s = 0$ and at $s = ik$, from (3.8) and (4.10). In addition δ_1 and δ_2 vanish within the range $-3i < s < i$ for any acceptable value of μ. In the evaluation of B_n we need the weighted integral of $\overline{u}_{4\lambda_1}$; consequently the associated singularities of B_n only

occur at $s = -3i$ and $s = i$, and will be considered at the same time as the singularities of the governing equation (2.18). The poles of B_n at $s = 0$ and $s = ik$ can lead to motions which are apparently large as $t \to \infty$ if one of the zeros of (5.8) occurs at the same values of s. The second possibility ($s = ik$) is unlikely if $n \neq 2$ since it implies that two free periods of oscillation of the fluid are equal, but we have not been able to rule this out. The first possibility ($s = 0$) is real, and there seems little doubt for any n we can choose a value of a/b to satisfy (5.8) by $s = 0$; for example, if $n = 4$ and

$$6a = (\sqrt{39} + \sqrt{15})\, b,$$

the left-hand side of (5.8) vanishes at $s = 0$. Using the argument above for $s = ik$, $k = (a^2 - b^2)/(a^2 + b^2)$, we can say that if $a = b$ it is most unlikely that (5.8) is satisfied by $s = 0$ for $n \neq 2$. In the case of particular interest therefore, when $a \approx b$, we conclude that the possibility of satisfying (5.8) at $s = 0$ is not likely to be serious except if $n = 2$.

Let us consider the case $n = 2$ in some detail. Here A_2 has a double pole at $s = ik$ because the coefficient of A_2 vanishes and because B_2 has a simple pole from (4.10). Using the formula for A in (3.5)

$$B_2 = \frac{5}{4}\int_{-1}^{1} \mu(1-\mu^2)^{\frac{1}{2}}\,(b^2 + c^2\mu^2)^{\frac{1}{2}}\,\bar{u}_{4\lambda_1}\, d\mu \tag{5.12}$$

$$= -\frac{5A(s-i)}{4b}\int_{-1}^{1} d\mu\, \frac{(b^2 + c^2\mu^2)^{\frac{1}{2}}}{a\delta_2}\,[a\mu + (b^2 + c^2\mu^2)^{\frac{1}{2}}]\,[\mu(b^2 + c^2\mu^2)^{\frac{1}{2}} - a(1 - 2\mu^2)]. \tag{5.13}$$

The corresponding value of A_2 is now easily worked out from (5.6) and, reverting to the co-ordinates (r, θ, z), the corresponding contribution to $\overline{V}_4$ is

$$\frac{3rz(s^2+1)\,B_2}{a^2(s-i) + b^2(s+i)}. \tag{5.14}$$

Let the coefficient of A in (5.13) be χ. The implication of (5.14) is that the secondary motion described by (3.3), and (3.5) gives rise to a complicated tertiary motion of which the second harmonic is of the same form as (3.3) but with a double pole at $s = ik$. Although formally this means that the tertiary flow increases without limit as $t \to \infty$ it is to be expected that the double pole arises through an error of order χ in the position of the simple pole of A in (3.5). To see this we note that the contribution from (5.14) to the tertiary flow is ultimately dominant and it gives rise via the boundary layer to a similar form to (5.14) in the quaternary flow but now with a triple pole at $s = ik$, and so on. Adding up all these contributions to $\overline{V}$ which are proportional to rz, *qua* functions of r and z, their total is

$$Arz \sum_{m=0}^{\infty} \left[\frac{3(s^2+1)\,\chi}{a^2(s-i) + b^2(s+i)}\right]^m = -\frac{2\Omega a^2 i \sin\alpha (s^2+1)\, rz}{\omega s[a^2(s-i) + b^2(s+i) - 3(s^2+1)\,\chi]}. \tag{5.15}$$

Thus the effect of the boundary layer is to shift the pole $s = ik$ to

$$s = ik + \frac{3(1-k^2)\,\chi}{a^2 + b^2} + \ldots \qquad \left(k = \frac{a^2 - b^2}{a^2 + b^2}\right),$$

the neglected terms being $o(\nu^{\frac{1}{2}})$ and most probably $O(\nu^{\frac{1}{2}})$; see Roberts & Stewartson (1963). Of particular interest is the position of the pole when

$$1 \gg \frac{a^2-b^2}{a^2+b^2} \gg \left(\frac{\nu}{\omega a^2}\right)^{\frac{1}{2}}; \tag{5.16}$$

the second condition means that the departures of the surface from the mean sphere, although small, are much larger than the boundary-layer thickness.

So far as χ is concerned it is sufficient to set $a = b$ whence

$$\chi = -\frac{5ai}{4}\int_{-1}^{1} \frac{(1+\mu)^2(1-2\mu)\,d\mu}{\delta_2}, \tag{5.17}$$

where $$\delta_2 = a^{-1}\left(\frac{\omega a^2}{\nu}\right)^{\frac{1}{2}}(s+i-2i\mu)^{\frac{1}{2}},$$

and since $s \approx 0$

$$\delta_2 = \begin{cases} a^{-1}\left(\dfrac{\omega a^2}{\nu}\right)^{\frac{1}{2}} e^{i\pi/4}(1-2\mu)^{\frac{1}{2}}, & \text{if} \quad \mu < \tfrac{1}{2}, \\ a^{-1}\left(\dfrac{\omega a^2}{\nu}\right)^{\frac{1}{2}} e^{-i\pi/4}(2\mu-1)^{\frac{1}{2}}, & \text{if} \quad \mu > \tfrac{1}{2}. \end{cases}$$

It follows that $$\chi = \tfrac{5}{4}ia^2\left(\frac{\nu}{2\omega a^2}\right)^{\frac{1}{2}}(0{\cdot}195 + 1{\cdot}976i), \tag{5.18}$$

whence the pole is moved to

$$s = i\frac{(a^2-b^2)}{(a^2+b^2)} - \left(\frac{\nu}{2\omega a^2}\right)^{\frac{1}{2}}(3{\cdot}720 - 0{\cdot}365i) + \ldots, \tag{5.19}$$

and, since it is in the half plane $\mathscr{R}(s) < 0$, the associated contribution to the flow dies out as $t \to \infty$. A similar result can be expected for all b/a. We note that the effect of the viscosity on the residue of the pole at $s = 0$ and therefore also on the ultimate motion is small if (5.16) is satisfied, the proportionate changes in the amplitude and orientation of the circulation being $O[\nu/\omega(a-b)^2]^{\frac{1}{2}}$.

It is of speculative interest in connexion with the case $R_1 = 0$ to note that if we set $a = b$ in (5.15) and use the formula for χ in (5.18), then as $t \to \infty$

$$V'' \to 0{\cdot}38\left(\frac{\omega a^2}{\nu}\right)^{\frac{1}{2}} \Omega rz \sin\alpha \cos(\theta - 1{\cdot}46) + \Omega r^2 \cos\alpha$$

(see (2.13), (2.16): the limit procedures have not been justified). This result suggests that for a sphere the secondary circulation is of the same kind as for a spheroid but the amplitude is $O[\Omega a^2(\omega a^2/\nu)^{\frac{1}{2}}]$ and the orientation is changed.

The last cases to be discussed are the behaviour of the solution in the neighbourhood of $s = \pm i$ and $-3i$, where the differential equation (2.18) takes on a singular form. Although δ_2 vanishes at $\mu a = (b^2+c^2\mu^2)^{\frac{1}{2}}$ when $s = i$ it follows from (4.10) and (5.7) that B_n is bounded as $s \to i$ for all n. Hence, substituting into (5.6) and noting that $\lambda_4 = i\gamma_4$ on S_0 if $s = i$, it follows that

$$A_n = O(s-i)^{\frac{1}{2}}, \tag{5.20}$$

near $s = i$. Further the nth harmonic in (5.4) can be written as

$$\tfrac{1}{4}[(s+i)^2+4]^{\frac{1}{2}}(s+i)\,rzA_n G_{n/2}\{\tfrac{1}{4}[(s+i)^2+4]r^2,\ \tfrac{1}{4}(s+i)^2z^2\}, \tag{5.21}$$

where $G_{n/2}(\alpha, \beta)$ is a homogeneous polynomial of degree $\frac{1}{2}n$ in α and β with constant coefficients. Hence the contribution to the tertiary flow is $O(s-i)$ as $s \to i$, and its inverse with respect to s tends to zero as $t \to \infty$. A similar remark applies to the neighbourhood of $s = -3i$. In the neighbourhood of $s = -i$, B_n is $O(s+i)$ from (4.10), and the coefficient of A_n in (5.6) is $O(1)$ since n is even. Consequently (5.21) is $O(s+i)^3$ near $s = -i$ and its contribution to the tertiary flow tends to zero as $t \to \infty$.

Summarizing we can expect that, of all the components of the secondary flow which depend on θ, only the steady component will remain as $t \to \infty$, the boundary layer serving to damp out the oscillatory terms.

6. The tertiary inviscid flow (ii)

In this section we consider the inviscid flow engendered by (4.17) which is independent of θ. We assume that the associated inviscid flow $\mathbf{u}_5$ is also independent of θ whence, on taking the Laplace transform with respect to ωt, the governing equations reduce to

$$\left.\begin{aligned} su^*_{5r} - 2u^*_{s\theta} &= -\frac{\partial V^*_5}{\partial r}, \quad su^*_{5\theta} + 2u^*_{5r} = 0, \\ su^*_{5z} &= -\frac{\partial V^*_5}{\partial z}, \quad \frac{\partial}{\partial r}(ru^*_{5r}) + \frac{\partial}{\partial z}(ru^*_{5z}) = 0; \end{aligned}\right\} \tag{6.1}$$

these equations also follow from (2.12) on assuming that $\mathbf{u}_5$ is independent of θ and neglecting the initial and forcing terms. Further, on S_0,

$$\frac{r}{a^2}u^*_{5r} + \frac{z}{b^2}u^*_{5z} = \left[\frac{r^2}{a^4} + \frac{z^2}{b^4}\right]^{\frac{1}{2}} u^*_{5\lambda_1}(b, \mu), \tag{6.2}$$

from (4.17). In terms of V^*_5,

$$u^*_{5r} = -\frac{s}{s^2+4}\frac{\partial V^*_5}{\partial r}, \quad u^*_{5z} = -\frac{1}{s}\frac{\partial V^*_5}{\partial z}, \tag{6.3}$$

so that

$$\frac{\partial}{\partial r}\left[r\frac{\partial V^*_5}{\partial r}\right] + \frac{(s^2+4)}{s^2}\frac{\partial^2 V^*_5}{\partial z^2} = 0. \tag{6.4}$$

Introduce new co-ordinates μ_5, λ_5 such that

$$r = \frac{2}{s^2+4}(\lambda_5^2 + \gamma_5^2)^{\frac{1}{2}}(1-\mu_5^2)^{\frac{1}{2}}, \quad z = \frac{2}{s}\mu_5\lambda_5, \tag{6.5}$$

where $\gamma_5 = a^2 + \frac{1}{4}s^2c^2$, S_0 is given by $\lambda_5 = \frac{1}{2}b_5$ and, on S_0, $\mu_5 = \mu_4 = \mu_1 = \mu$.

The boundary conditions now reduce to

$$-\frac{\partial V^*_5}{\partial \lambda_5} = a^{-1}(b^2 + c^2\mu^2)^{\frac{1}{2}} u^*_{5\lambda}, \tag{6.6}$$

on S_0. The most general appropriate solution for V^*_5 in S_0 is

$$V^*_5 = \sum_{n=0}^{\infty} C_n P_n\left(\frac{i\lambda_5}{\gamma_5}\right) P_n(\mu_5), \tag{6.7}$$

where C_n are constants and, since $u^*_{5\lambda}$ is an even function of μ, n also must be even. Hence

$$\frac{2}{(2n+1)} C_n P'_n\left(\frac{ib_5}{2\gamma_5}\right) = -\frac{\Omega i \cos\alpha}{2\omega s} \int_{-1}^{+1} P_n(\mu) \frac{\partial}{\partial\mu}\left[(b^2+c^2\mu^2)^{\frac{1}{2}}(1-\mu^2)\left(\frac{1}{\delta_3}-\frac{1}{\delta_4}\right)\right] d\mu. \tag{6.8}$$

The main interest here centres on the behaviour of the solution at large times, contributions to which arise from three sources. To begin with C_n will develop a simple pole whenever P'_n vanishes unless $s = 0$: the corresponding motion, according to (6.7), is a feeble oscillation and one can expect to show, using a similar argument to that which led to (5.19), that viscosity damps it out ultimately. Further the differential equation is singular at $s = \pm 2i$ where $ibs/2\gamma_5 = \pm 1$. Since $P'_n(\pm 1) \neq 0$, the leading terms from this cause must become zero, as $\to \infty$. Finally, one must consider the case $s = 0$. When s is very small

$$C_n = -\frac{(2n+1)\,\Omega\cos\alpha a^3}{n(n+1)\,b\omega s^2 P_n(0)} \left(\frac{\nu}{\omega a^2}\right)^{\frac{1}{2}} \int_{-1}^{+1} d\mu (1-\mu^2) P'_n(\mu) \frac{(b^2+c^2\mu^2)^{\frac{1}{2}}}{(-2i\mu a)^{\frac{1}{2}}}, \tag{6.9}$$

so that, on inverting (6.7), V^*_5 contains a term which is proportional to t when t is large. The implication is that the boundary layer exerts a decisive influence on the inviscid flow outside in this instance.

The reason for the difference between the solution in § 5 near $s = -i$ and the solution in § 6 near $s = 0$, although both governing equations are similar and singular at these points is two-fold. First (4.10) contains a factor $(s+i)$ while (4.17) contains a factor s^{-1} so that the behaviour of V^*_5 near $s = 0$ is bound to be more singular than the behaviour of V^*_4 near $s = -i$. There is, however, a second deeper reason. Near $s = 0$ the governing equation (6.1) reduces to

$$u^*_{5r} = 0, \quad \frac{\partial u^*_{5z}}{\partial z} = 0, \tag{6.10}$$

which means that u^*_{5z} is independent of z and therefore must be an even function of μ. Consequently V^*_5 is an odd function of μ, but the boundary condition (6.2) requires it to be an even function. This patent contradiction means that a serious breakdown in the solution must occur as $t \to \infty$. From studies of the steady state (e.g. Proudman 1956) we know that the boundary layer in such a case exerts a decisive influence on the flow outside, adjusting it until the condition that $u_{5\lambda}$ is an odd function of μ is satisfied, which is only possible if $u_{5\lambda} \to 0$ at $t \to \infty$. Hence the initial angular velocity $u_\theta = \Omega r \cos\alpha$ is also damped out by the boundary layer.

On the other hand, near $s = -i$, while the governing equations (2.17) and (2.18) also reduce in part to

$$\frac{\partial u^*_{4z}}{\partial z} = 0, \tag{6.11}$$

so that u^*_{4z} is an odd function of μ, this condition is actually satisfied by (4.10) so that no corresponding difficulties occur.

7. Brief summary: geophysical application

Our analysis has established that, no matter what the initial motion of the fluid in the cavity may be, if

$$R_1 = \frac{(a^2-b^2)\,\omega}{\Omega a^2}, \quad R_1 R_2 = \frac{\omega(a^2-b^2)}{\nu}, \quad R_3 = \frac{\omega a^2}{\nu}$$

are all large, the flow will ultimately (i.e. in a time of order a^2/ν) be the same. It may be described most easily in a frame of reference rotating about the axis of precession with the angular velocity Ω of precession. The primary flow is then merely a solid-body rotation of the fluid and envelope together. Superimposed upon this is a steady flow which, apart from a boundary layer whose thickness is nearly everywhere of order $(\nu/\omega)^{\frac{1}{2}}$ within which the fluid adjusts to the no-slip conditions at the interface, is characterized by closed elliptical streamlines and by uniform vorticity

$$\frac{2(a^2+b^2)}{\omega^2(a^2-b^2)}\,\boldsymbol{\omega}\times(\boldsymbol{\Omega}\times\boldsymbol{\omega}). \tag{7.1}$$

The streamlines lie in planes perpendicular to this, and are similar to the elliptical section of the boundary by a plane through the axis of symmetry. In fact, if y be a Cartesian co-ordinate drawn along $\boldsymbol{\omega}\times\boldsymbol{\Omega}$, z along $\boldsymbol{\omega}$ and x drawn to complete the triad, this flow is given by

$$\frac{2\Omega a^2 b^2 \sin\alpha}{(a^2-b^2)}\left(0, \frac{z}{b^2}, -\frac{y}{a^2}\right). \tag{7.2}$$

Its greatest magnitude (taken at the poles) is

$$U = 2\Omega a^2 b \sin\alpha/(a^2-b^2). \tag{7.3}$$

We will now amplify the brief remarks given in the Introduction on the geophysical relevance of the present theory to the hydrodynamics of the Earth's core. Let us first briefly estimate the parameters R_i. Many calculations have been made of the ellipticity ϵ of the equidensity surfaces within the Earth. They have been generally based on Clairaut's equation (cf. e.g. Chandrasekhar & Roberts 1963), and have led to the conclusion that ϵ decreases monotonically from the value of 1/297 for the Earth's surface to a value of approximately 1/470 at the centre of the Earth (e.g. Lambert & Darling 1951). In particular, a computation of the ellipticity of the core has given $\epsilon = 1/390$, i.e.

$$\frac{(a^2-b^2)}{a^2} \doteq 2\epsilon \doteq 0{\cdot}00513. \tag{7.4}$$

Now for the luni-solar precession

$$\alpha = 23{\cdot}452^\circ \quad \text{and} \quad 2\pi/\Omega = 25{,}725\,\text{years}, \tag{7.5}$$

thus

$$\omega/\Omega = 9{\cdot}40\times 10^6. \tag{7.6}$$

It follows that

$$R_1 = 4{\cdot}82\times 10^4. \tag{7.7}$$

The kinematic viscosity of the Earth's core is notoriously difficult to estimate (e.g. Hide & Roberts 1961), and values between $10^{-3}\,\mathrm{cm}^2/\mathrm{sec}$ and $10^{3}\,\mathrm{cm}^2/\mathrm{sec}$ have been offered, of which the latter is certainly an extreme upper limit. Even this value, however, implies that both $R_1 R_2$ and R_3 exceed unity; in all probability, therefore, they do so by large factors. Thus, the basic tenets of the analysis presented in the earlier sections are satisfied, and its particular relevance to theories of the geomagnetic field can be legitimately examined.

It is today generally accepted that the source of the geomagnetic field is to be found in a self-excited dynamo located deep within the Earth, the 'moving parts' being the hydrodynamic motions of the core. Naturally, if a driving mechanism were not available, Ohmic dissipation would cause the dynamo to run down in an electromagnetic decay time ($\sim$ 10,000 years). On the other hand, palaeomagnetic studies have established that the magnitude of the geomagnetic field has not varied greatly over geological time. Evidently, then, a capital question is to determine the nature of the driving mechanism which must be present. In an early study, Bullard (1949) singled out thermal convection as the most likely cause. He was, however, unable to exclude the possibility that the dynamo is precessionally driven. Indeed, the subsequent analysis by Bondi & Lyttleton (1953), though inconclusive, drew attention to the singularities of the boundary layer at latitudes 30° N. and 30° S. (cf. § 4 above) as a possible source of instabilities which would intermittently shed turbulent eddies. These, the authors felt, would through their inductive effects contribute to the secular variation of the geomagnetic field in these latitudes. They might also explain the observed small irregular fluctuations in the Earth's state of rotation. It is fair, therefore, to state that their ideas enhanced, if anything, the interest in the precessional effects. However, unsteady motion was invoked by Bondi & Lyttleton to meet a situation for which, they maintained, a steady solution did not exist. In our model, on the other hand, a steady flow does establish itself, and there seems to be no obvious reason why it should be unstable or be associated with the observed irregular fluctuations, either in the geomagnetic field or in rotation.

Bullard called attention to early studies of precessionally driven flows within a spheroidal cavity, and in particular ascribed to Poincaré (1910) a theorem that 'the ellipticity of the earth is sufficient to ensure that the material of the core moves with the rest of the earth like a rigid body in a small motion of precession, even in the absence of viscosity'. Again, he comments 'It has not been proved that the motion...is possible for a precession of finite amplitude (the angle at the vertex of the core swept out by the earth's axis is 47°)'. It seems to us, however, that Poincaré's solution, like ours, depends only on the assumption that $\mathbf{u}$ and $\boldsymbol{\Omega}$ are small; it does not require that α, too, is small. His solution, like ours, consists of a primary solid-body rotation with the mantle upon which a steady secondary flow of uniform vorticity is superimposed. (He did not, however, demonstrate that this solution is the one realized in practice in the limit $\nu t \to \infty$.) Bullard did not discuss the secondary flow. It should be observed, however, that its magnitude (cf. equation (7.3)) is of the order of 1 cm/sec. The horizontal velocities near the surface of the core, as inferred by studies of the westward drift, are also of this order of magnitude (or even rather smaller). Thus the

precessional flow should be quite large enough to make its inductive effects felt; assuming an electrical conductivity σ of 3×10^5 mho/m, the corresponding magnetic Reynolds number R_m is

$$R_m = \mu_0 \sigma a U \doteq 1{\cdot}3 \times 10^4 \gg 1 \tag{7.8}$$

(μ_0 is the permeability of free space, $4\pi \times 10^{-7}$ Henry/m). For dynamo action it is necessary that R_m should, indeed, be large. However, it is also necessary that the motions have a sufficiently low degree of symmetry, and it seems to us that, in this respect, the motion (7.2) would be incapable of amplifying and maintaining a stray magnetic field.

Finally one further point should be noted. The theory we have described has ignored Lorentz forces. This is appropriate when the solution is used to determine whether precessional flows can, by themelves, amplify a stray weak magnetic field. There is some difficulty, however, in applying it without modification to the actual geophysical situation in which strong magnetic fields are present. The precessional flows we have described may be thought of as being driven by a force per unit mass of approximately $\omega\Omega a$, i.e. about 10^{-7} cm^2/sec. The Lorentz forces per unit mass $B^2/\mu_0\rho a$ may well be larger; indeed, if we assume in the core a geomagnetic field B of only 50 G, their magnitude is 10^{-7} cm^2/sec. The toroidal field is, however, probably considerably in excess of 50 G. Further, since the lower mantle is appreciably conducting, there will be a magnetic coupling of core and mantle; similar order of magnitude estimates may be made, and indicate that this effect, is likely to dominate the viscous coupling.

This research has been supported in part (K. S.) by the U.S. Army under Contract No. DA-11-022-ORD-2059 and in part (P. H. R.) by the U.S. Air Force under Research Grant AF-AFOSR 62-136.

REFERENCES

Bondi, H. & Lyttleton, R. A. 1953 *Proc. Camb. Phil. Soc.* **49**, 498.

Bullard, E. C. 1949 *Proc. Roy. Soc.* A, **197**, 433.

Chandrasekhar, S. & Roberts, P. H. 1963. To be published in *Astrophys. J.*

Fultz, D. & Moore, D. W. 1962 Private communication.

Hide, R. & Roberts, P. H. 1961 *Physics and Chemistry of the Earth*, Vol. 4. London: Pergamon.

Lambert, W. D. & Darling, F. W. 1951 *Internal Constitution of the Earth* (ed. B. Gutenberg). New York: Dover.

Poincaré, H. 1910 *Bull. Astr.* **27**, 321.

Proudman, I. 1956 *J. Fluid Mech.* **1**, 505.

Roberts, P. H. & Stewartson, K. 1963 *Astrophys. J.* **137**, 777.

Stewartson, K. 1957 *J. Fluid Mech.* **3**, 17.

PRINTED IN GREAT BRITAIN AT THE UNIVERSITY PRINTING HOUSE, CAMBRIDGE
(BROOKE CRUTCHLEY, UNIVERSITY PRINTER)

Reprinted for private circulation from
THE ASTROPHYSICAL JOURNAL
Vol. 138, No. 3, October 1, 1963

THE ELLIPTICITY OF A SLOWLY ROTATING CONFIGURATION

S. CHANDRASEKHAR AND P. H. ROBERTS
University of Chicago
Received April 29, 1963

ABSTRACT

The second-order virial theorem is used to set upper and lower bounds for m/ϵ_R for a slowly rotating configuration, where m is the ratio of the centrifugal acceleration at the equator to the (average) gravitational acceleration on its surface and ϵ_R is the ellipticity of its slightly oblate figure of equilibrium. The bounds obtained are explicitly evaluated for the polytropes, for a model consisting of a core and a mantle of constant densities, and for a particular model for the earth.

I. INTRODUCTION

For a slowly rotating configuration, the ratio m, of the centrifugal acceleration at the equator to the average gravitational acceleration on its surface, and the ellipticity ϵ_R, of its slightly oblate figure of equilibrium, are in a relationship of cause and effect. It is, in fact, one of Newton's well-known theorems (cf. Todhunter 1873, Sec. 27) that, for a homogeneous configuration,

$$\frac{m}{\epsilon_R} = \frac{4}{5}. \tag{1}$$

Quite generally, m/ϵ_R is functionally dependent on the distribution of the density in the configuration; and the establishment of this dependence by Clairaut is one of the great achievements that followed in Newton's wake.

If one restricts one's self to configurations in which (in the non-rotating state) the density $\rho_0(r)$ at a point, does not exceed the mean density, $\bar{\rho}_0(r)$, interior to that point, then it is known that (cf. Chandrasekhar 1933, eq. [100])

$$1 \leq \tfrac{5}{4}\frac{m}{\epsilon_R} < 2.5\,. \tag{2}$$

And if the configuration is one in which the departures from homogeneity are not very great, then an approximate formula, derived by Darwin (1899) on the basis of a transformation of Clairaut's equation due to Radau, relates m/ϵ_R to the moment of inertia,

$$I = \int_V \rho_0(r)\, r^2 dx = 4\pi \int_0 \rho_0(r)\, r^4 dr\,, \tag{3}$$

of the configuration by

$$\frac{I}{MR^2} = 1 - \tfrac{2}{5}\left(\tfrac{5}{2}\frac{m}{\epsilon_R} - 1\right)^{1/2}, \tag{4}$$

where M is its mass and R is its (mean) radius. (In eq. [3], $\rho_0(r)$ is the zero-order spherically symmetric part of the density distribution; see eq. [10] below.)

In Table 1, the values of I/MR^2 for the polytropes are compared with those deduced with the aid of equation (4) and the known values of $5m/2\epsilon_R$. It would appear from this table that Darwin's formula (4) cannot be applied if the central density of the configuration exceeds, say, six times its mean density.

In this paper, we shall set certain upper and lower bounds to $5m/4\epsilon_R$ from an application to this problem of the second-order virial equations which have already proved their utility in other connections (see the various papers of Chandrasekhar, Lebovitz, and Roberts in the recent issues of the *Astrophysical Journal*).

II. AN INTEGRAL RELATION

Let the origin of the co-ordinate system be at the center of mass of the configuration; and let the z-axis coincide with the axis of rotation; and, finally, let Ω denote the uniform angular velocity of the rotation. We shall assume that the configuration has symmetry about the z-axis; this assumption is justified for slowly rotating configurations in which we are presently interested.

A fundamental relation provided by the virial theorem is

$$\Omega^2 I_{11} = \mathfrak{W}_{33} - \mathfrak{W}_{11}\,, \tag{5}$$

where

$$\mathfrak{W}_{ij} = \int_V \rho x_i \frac{\partial \mathfrak{V}}{\partial x_j}\, dx \qquad \text{and} \qquad I_{ij} = \int_V \rho x_i x_j dx \tag{6}$$

are the potential energy and the moment of inertia tensors. In equations (6), $\mathfrak{V}$ denotes the gravitational potential and ρ the density.

TABLE 1

APPLICATION OF DARWIN'S FORMULA TO POLYTROPES*

n	$\rho_c/\bar{\rho}$	$5m/2\epsilon_R$	I/MR^2	
			Exact	By Darwin's Formula
0.....	1.0	2.0000	0.60000	0.6000
1.0.....	3.29	3.2899	.39207	.3947
1.5.....	5.99	3.8865	.30690	.3204
2.0.....	11.40	4.3559	.23227	.2673
3.0.....	54.18	4.8596	.11304	.2142
3.5.....	152.9	4.9513	0.06832	0.2049

* The values of $5m/2\epsilon_R$ given in Chandrasekhar (1933, Table 3) have been revised in accordance with the results of the more precise integrations tabulated in Chandrasekhar and Lebovitz (1962, Table 7).

In the case of axisymmetry, $\mathfrak{W}_{33}$ and $\mathfrak{W}_{11}$ are the two distinct components of the potential energy tensor; and in spherical polar co-ordinates (r, ϑ, φ), the expressions for them become

$$\mathfrak{W}_{33} = \int_V \rho r \left[\mu^2 \frac{\partial \mathfrak{V}}{\partial r} + \frac{\mu(1-\mu^2)}{r} \frac{\partial \mathfrak{V}}{\partial \mu} \right] dx \tag{7}$$

and

$$\mathfrak{W}_{11} = \tfrac{1}{2} \int_V \rho r \left[(1-\mu^2) \frac{\partial \mathfrak{V}}{\partial r} - \frac{\mu(1-\mu^2)}{r} \frac{\partial \mathfrak{V}}{\partial \mu} \right] dx\,, \tag{8}$$

where $\mu = \cos\vartheta$. With the foregoing expressions for $\mathfrak{W}_{33}$ and $\mathfrak{W}_{11}$, equation (5) gives

$$\Omega^2 I_{11} = \int_V \rho r \left[P_2(\mu) \frac{\partial \mathfrak{V}}{\partial r} + \tfrac{3}{2} \frac{\mu(1-\mu^2)}{r} \frac{\partial \mathfrak{V}}{\partial \mu} \right] dx\,. \tag{9}$$

Equation (9) is exact. We shall now apply it to a slowly rotating configuration in which the departures from spherical symmetry, considered small, are governed by the Legendre function, $P_2(\mu)$. We shall write, then,

$$\rho = \rho_0(r) + \rho_2(r) P_2(\mu) \tag{10}$$

and

$$\mathfrak{V} = \mathfrak{V}_0(r) + \mathfrak{V}_2(r) P_2(\mu)\,, \tag{11}$$

where

$$|\rho_2(r)| \ll |\rho_0(r)| \quad \text{and} \quad |\mathfrak{V}_2(r)| \ll |\mathfrak{V}_0(r)| \,. \tag{12}$$

Under these same circumstances, we may also suppose that the boundary of the configuration is given by

$$R(\mu) = R[1 - \tfrac{2}{3}\epsilon_R P_2(\mu)] \,, \tag{13}$$

where ϵ_R defines the ellipticity of the figure of equilibrium.

In the framework of the foregoing assumptions, we may clearly replace I_{11} in equation (9) by its zero-order value,

$$\tfrac{1}{3}I = \tfrac{1}{3}\int_V \rho_0(r)\, r^2 dx = \tfrac{4}{3}\pi \int_0^R \rho_0(r)\, r^4 dr \,. \tag{14}$$

Now substituting for ρ and $\mathfrak{V}$ in accordance with equations (10) and (11) in equation (9), we find (on ignoring all quantities of the second order and after performing the integrations over the angles)

$$\tfrac{1}{3}\Omega^2 I = \tfrac{4}{5}\pi \int_0^R \rho_0(r)\, r^3 \left(\frac{d\mathfrak{V}_2}{dr} + \frac{3}{r}\mathfrak{V}_2\right) dr + \tfrac{4}{5}\pi \int_0^R \rho_2(r)\, r^3 \frac{d\mathfrak{V}_0}{dr}\, dr \,. \tag{15}$$

Since the gravitational potential is derived in a linear fashion from the distribution of density, it is clear that $\mathfrak{V}_0$ and $\mathfrak{V}_2$ should be expressible in terms of ρ_0 and ρ_2, respectively. Thus, by expanding $|x - x'|^{-1}$ in spherical harmonics in the usual manner, we readily find, from Poisson's integral,

$$\mathfrak{V}_0(r) + \mathfrak{V}_2(r)P_2(\mu) = G\int_V \frac{\rho_0(r') + \rho_2(r')P_2(\mu')}{|x - x'|}\, dx' \,, \tag{16}$$

that

$$\mathfrak{V}_0(r) = 4\pi G\left[\frac{1}{r}\int_0^r \rho_0(s)\, s^2 ds + \int_r^R \rho_0(s)\, s\, ds\right] \tag{17}$$

and

$$\mathfrak{V}_2(r) = \tfrac{4}{5}\pi G\left[\frac{1}{r^3}\int_0^r \rho_2(s)\, s^4 ds + r^2 \int_r^R \frac{\rho_2(s)}{s}\, ds\right] . \tag{18}$$

Relations which follow from equations (17) and (18) are[1]

$$\frac{d\mathfrak{V}_0}{dr} = -\frac{4\pi G}{r^2}\int_0^r \rho_0(s)\, s^2 ds = -\frac{GM(r)}{r^2} \tag{19}$$

and

$$\frac{1}{r^3}\frac{d}{dr}(r^3\mathfrak{V}_2) = \frac{d\mathfrak{V}_2}{dr} + \frac{3}{r}\mathfrak{V}_2 = 4\pi G r \int_r \frac{\rho_2(s)}{s}\, ds \,, \tag{20}$$

where, in equation (19), $M(r)$ is the mass interior to r in the zeroth approximation.

Inserting relations (19) and (20) in equation (15), we obtain

$$\Omega^2 I = \frac{48\pi^2 G}{5}\left[\int_0^R dr\, \rho_0(r)\, r^4 \int_r^R ds\, \frac{\rho_2(s)}{s} - \int_0^R dr\, \rho_2(r)\, r \int_0^r ds\, \rho_0(s)\, s^2\right] . \tag{21}$$

[1] These relations are readily seen to be the first integrals of the equations,

$$\frac{d^2\mathfrak{V}_0}{dr^2} + \frac{2}{r}\frac{d\mathfrak{V}_0}{dr} = -4\pi G\rho_0$$

and

$$\frac{d^2\mathfrak{V}_2}{dr^2} + \frac{2}{r}\frac{d\mathfrak{V}_2}{dr} - \frac{6}{r^2}\mathfrak{V}_2 = -4\pi G\rho_2 \,,$$

governing $\mathfrak{V}_0$ and $\mathfrak{V}_2$.

Inverting the order of the integrations in the second term on the right-hand side, we can write

$$\Omega^2 I = \frac{48\pi^2 G}{5}\int_0^R dr\,\rho_0(r)\,r^2\left[r^2\int_r^R ds\,\frac{\rho_2(s)}{s} - \int_r^R ds\,\rho_2(s)\,s\right], \tag{22}$$

or

$$\Omega^2 I = \frac{12\pi G}{5}\int_0^R dr\,\frac{dM(r)}{dr}\left[r^2\int_r^R ds\,\frac{\rho_2(s)}{s} - \int_r^R ds\,\rho_2(s)\,s\right]. \tag{23}$$

An integration by parts now leads to the result

$$\Omega^2 I = -\frac{24\pi G}{5}\int_0^R dr M(r)\,r\int_r^R ds\,\frac{\rho_2(s)}{s}. \tag{24}$$

In virtue of equation (20), we can also write

$$\Omega^2 I = -\tfrac{6}{5}\int_0^R \frac{M(r)}{r^3}\,\frac{d}{dr}\,(r^3\mathfrak{V}_2)\,dr\,; \tag{25}$$

and this is the desired integral relation.

III. AN INTEGRAL EQUATION GOVERNING THE ELLIPTICITY OF THE SURFACES OF EQUAL GEOPOTENTIAL

The definition of the geopotential is

$$\Psi = \mathfrak{V} + \tfrac{1}{2}\Omega^2 r^2(1-\mu^2)\,, \tag{26}$$

so that

$$\text{grad } p = \rho\,\text{grad}\,\Psi\,. \tag{27}$$

Let the surfaces of equal geopotential be

$$r[1 - \tfrac{2}{3}\epsilon(r)P_2(\mu)]\,; \tag{28}$$

$\epsilon(r)$ defines, then, the varying ellipticity of these surfaces. At the boundary of the configuration, $\epsilon(r)$ takes the value ϵ_R.

For the assumed form of $\mathfrak{V}$,

$$\Psi = \mathfrak{V}_0(r) + \mathfrak{V}_2(r)P_2(\mu) + \tfrac{1}{3}\Omega^2 r^2[1 - P_2(\mu)]\,. \tag{29}$$

The condition that Ψ so defined is constant over the surfaces (28) requires (in our present approximation) that

$$\mathfrak{V}_2(r) = \tfrac{2}{3}\,r\epsilon(r)\,\frac{d\mathfrak{V}_0}{dr} + \tfrac{1}{3}\Omega^2 r^2\,; \tag{30}$$

or (cf. eq. [19])

$$\mathfrak{V}_2(r) = -\tfrac{2}{3}G\epsilon(r)\,\frac{M(r)}{r} + \tfrac{1}{3}\Omega^2 r^2\,. \tag{31}$$

From this last equation it follows that

$$\frac{d}{dr}\,(r^3\mathfrak{V}_2) = -\tfrac{2}{3}G\frac{d}{dr}[\epsilon(r)M(r)\,r^2] + \tfrac{5}{3}\Omega^2 r^4\,. \tag{32}$$

Returning to equation (25), we may, in view of equation (32), write it in the form

$$\Omega^2 I = -2\Omega^2\int_0^R M(r)\,r\,dr + \tfrac{4}{5}G\int_0^R \frac{M(r)}{r^3}\,\frac{d}{dr}\,[\epsilon(r)M(r)\,r^2]\,dr\,. \tag{33}$$

On the other hand,

$$2\int_0^R M(r)\,r\,dr = MR^2 - 4\pi\int_0^R \rho_0(r)\,r^4\,dr = MR^2 - I\,. \tag{34}$$

Equation (33), therefore, becomes

$$\Omega^2 MR^2 = \tfrac{4}{5}G\int_0^R \frac{M(r)}{r^3}\frac{d}{dr}[\epsilon(r)M(r)\,r^2]\,dr\,. \tag{35}$$

Now, by definition,

$$m = \frac{\Omega^2 R}{GM/R^2} = \frac{\Omega^2 R^3}{GM}\,. \tag{36}$$

In terms of m, we can rewrite equation (35) in the form

$$\tfrac{5}{4}m = \frac{R}{M^2}\int_0^R \frac{M(r)}{r^3}\frac{d}{dr}[\epsilon(r)M(r)\,r^2]\,dr\,; \tag{37}$$

and an integration by parts gives

$$\tfrac{5}{4}m = \epsilon_R + \frac{R}{M^2}\int_0^R \epsilon(r)M(r)\,r^2\frac{d}{dr}\left[-\frac{M(r)}{r^3}\right]dr\,. \tag{38}$$

Making use of the relation

$$-\frac{d}{dr}\frac{M(r)}{r^3} = -\tfrac{4}{3}\pi\frac{d\bar{\rho}_0(r)}{dr} = \frac{4\pi}{r}[\bar{\rho}_0(r) - \rho_0(r)]\,, \tag{39}$$

we can also write

$$\tfrac{5}{4}m = \epsilon_R + 4\pi\frac{R}{M^2}\int_0^R \epsilon(r)M(r)\,r\,[\bar{\rho}_0(r) - \rho_0(r)]\,dr\,. \tag{40}$$

The integrals which occur in equations (37), (38), and (40) are, therefore, all positive definite, so long as

$$\bar{\rho}_0(r) \geq \rho_0(r)\,. \tag{41}$$

IV. UPPER AND LOWER BOUNDS FOR m/ϵ_R

It is known (cf. Jeffreys 1959, p. 149) that, for configurations in which condition (41) is met, $\epsilon(r)$ is a monotonic increasing function of r and that

$$\epsilon_R \geq \epsilon(r) \geq \epsilon_R\left(\frac{r}{R}\right)^3. \tag{42}$$

Therefore, by inserting these bounds for $\epsilon(r)$ in equation (37), (38), or (40), we shall obtain corresponding bounds for m/ϵ_R. Thus, by replacing $\epsilon(r)$ by ϵ_R in equation (38), we obtain

$$\tfrac{5}{4}\frac{m}{\epsilon_R} \leq 1 - \frac{R}{M^2}\int_0^R M(r)\,r^2\frac{d}{dr}\frac{M(r)}{r^3}\,dr\,. \tag{43}$$

Now, by successive integrations by parts, we find

$$\begin{aligned}\int_0^R M(r)\,r^2\frac{d}{dr}\frac{M(r)}{r^3}\,dr &= \int_0^R \frac{M(r)\,dM(r)}{r} + 3\int_0^R M^2(r)\frac{d}{dr}\left(\frac{1}{r}\right)dr \\ &= 3\frac{M^2}{R} - 5\int_0^R \frac{M(r)\,dM(r)}{r} = 3\frac{M^2}{R} - 5\frac{|\mathfrak{W}|}{G}\,,\end{aligned} \tag{44}$$

where $\mathfrak{W}$ ($=\mathfrak{W}_{ii}$) is the gravitational potential energy of the configuration. We thus obtain the inequality,

$$\tfrac{5}{4}\frac{m}{\epsilon_R} \leq 5\,\frac{|\mathfrak{W}|}{GM^2/R} - 2\,, \tag{45}$$

which gives the desired upper bound.

Similarly, by replacing $\epsilon(r)$ by $\epsilon_R(r/R)^3$ in equation (37), we obtain

$$\tfrac{5}{4}\frac{m}{\epsilon_R} \geq \frac{1}{M^2R^2}\int_0^R \frac{M(r)}{r^3}\,\frac{d}{dr}\,[M(r)\,r^5]\,dr\,. \tag{46}$$

The integral on the right-hand side of this inequality is

$$\tfrac{5}{2}\int_0^R M^2(r)\,d(r^2) + \int_0^R r^2M(r)\,dM(r) = \tfrac{5}{2}\,M^2R^2 - 4\int_0^R r^2M(r)\,dM(r)\,. \tag{47}$$

We thus obtain the inequality,

$$\tfrac{5}{4}\frac{m}{\epsilon_R} \geq 2.5 - \frac{4}{M^2R^2}\int_0^R r^2M(r)\,dM(r)\,, \tag{48}$$

which gives the desired lower bound.

V. SOME ILLUSTRATIONS

We shall now consider some applications of the inequalities found in the preceding section.

a) The Polytropes

For the gravitational potential energy of a polytrope, we have Emden's well-known formula,

$$\mathfrak{W} = -\frac{3}{5-n}\,\frac{GM^2}{R}\,; \tag{49}$$

and the inequality (45) giving the upper bound of m/ϵ_R becomes, in this case,

$$\tfrac{5}{4}\frac{m}{\epsilon_R} \leq \frac{5+2n}{5-n}\,. \tag{50}$$

It should be noted that inequality (50) does not provide a meaningful bound for $n > \frac{5}{3}$: for we know that, in all cases, $5\,m/4\epsilon_R < 2.5$ (cf. eq. [2]); and only for $n > \frac{5}{3}$ is the right-hand side of (50) less than 2.5.

Turning to inequality (48) giving the lower bound of m/ϵ_R and expressing the various quantities in terms of Emden's variables and units (cf. Chandrasekhar and Lebovitz 1962, eq. [8]), we have

$$\tfrac{5}{4}\frac{m}{\epsilon_R} \geq 2.5 + \frac{4}{\xi_1{}^6(\theta_1')^2}\int_0^{\xi_1}\xi^6\theta^n\,\frac{d\theta}{d\xi}\,d\xi\,. \tag{51}$$

After an integration by parts, the inequality (51) becomes

$$\tfrac{5}{4}\frac{m}{\epsilon_R} \geq 2.5 - \frac{24}{(n+1)\,\xi_1{}^6(\theta_1')^2}\int_0^{\xi_1}\theta^{n+1}\xi^5 d\xi\,, \tag{52}$$

where it may be noted that, according to a reduction formula due to Milne (1929),

$$\int_0^{\xi_1}\theta^{n+1}\xi^5 d\xi = \frac{n+1}{2(n+7)}\left[\xi_1{}^6(\theta_1')^2 - 12\int_0^{\xi_1}\xi^3\theta^2 d\xi\right]. \tag{53}$$

In Table 2, the upper and the lower bounds of $5m/4\epsilon_R$ given by the inequalities (50) and (52) are listed, together with their known values and also the values deduced from Darwin's formula (4).

b) A Model Consisting of a Core and a Mantle of Different Densities

A model often used as an idealization of the density distribution which occurs inside the earth consists of a core of a certain constant density and a mantle of a different constant density. We shall consider the implications of the inequalities (45) and (48) for this model.

TABLE 2

UPPER AND LOWER BOUNDS OF $5m/4\epsilon_R$ FOR POLYTROPES

n	Lower Bound	Upper Bound	Exact Value	By Darwin's Formula
0.....	1.000	1.000	1.0000	1.000
1.0.....	1.456	1.750	1.6449	1.655
1.5.....	1.661	2.286	1.9432	2.001
2.0.....	1.850	.....	2.1779	2.312
3.0.....	2.117	.....	2.4298	.....
3.5.....	2.293	.....	2.4757	.....
5.0.....	2.500	.....	2.5000	.....

Let the core occupy a fraction q of the radius R and let α be the ratio of the densities of the core and the mantle. The integrals which occur in the inequalities (45) and (48) are readily evaluated for this model, and we find

$$\frac{5}{4}\frac{m}{\epsilon_R} \leq 3\,\frac{\alpha^2 q^5 + (1-q^5) - 2.5(1-\alpha)\,q^3(1-q^2)}{(\alpha q^3 + 1 - q^3)^2} - 2 \tag{54}$$

and

$$\frac{5}{4}\frac{m}{\epsilon_R} \geq 2.5 - 1.5\,\frac{\alpha^2 q^8 + (1-q^8) - 1.6(1-\alpha)\,q^3(1-q^5)}{(\alpha q^3 + 1 - q^3)^2}. \tag{55}$$

For the earth

$$m = 0.003499 \qquad \text{and} \qquad \epsilon_R = 1/297; \tag{56}$$

and it is further estimated that

$$q = 0.545\,. \tag{57}$$

These values for m, ϵ_R, and q, when inserted in formulae (54) and (55), lead to two quadratic inequalities for α, which, when solved, give

$$4.68 \geq \alpha \geq 2.62\,. \tag{58}$$

The value of α generally used, when the present model for the earth is assumed, is 2.9. The lower bound for α determined by formula (54) is, thus, very close to its "true value."

c) A Particular Model for the Earth

As a last illustration, we shall consider a particular model for the earth which Bullard (1946) has used for a direct integration of Clairaut's equation. Using the values of $\rho(r)$ and $\bar{\rho}(r)$ tabulated in Bullard's paper, we find (by numerical integration) that

$$\frac{|\mathfrak{W}|}{GM^2/R} = 0.6669 \qquad \text{and} \qquad \frac{1}{M^2R^2}\int_0^R r^2 M(r)\,dM(r) = 0.3326\,; \tag{59}$$

and these values, when inserted into inequalities (45) and (48), give

$$1.170 \leq \tfrac{5}{4}\frac{m}{\epsilon_R} \leq 1.335. \tag{60}$$

The actual value of $5m/4\epsilon_R$ for the earth is 1.299; it thus differs from the upper bound given by formula (60) by less than 3 per cent.

The research reported in this paper (as it pertains to S. Chandrasekhar) has in part been supported by the Office of Naval Research under contract Nonr-2121(24) with the University of Chicago; and (as it pertains to P. H. Roberts) has been supported in part by a National Science Foundation Grant No. N.S.F.-GP-975.

APPENDIX

Clairaut's Equation

It will be noticed that nowhere in the text was there an occasion to define the ellipticity of the surfaces of constant density. However, if the equation of state is barotropic, then it follows from equation (27) that the surfaces of equal geopotential are also surfaces of equal pressure and density. The requirement (in these circumstances) that the density specified in equation (10) is constant over the same surfaces (28) the geopotential is constant, yields the additional relation,

$$\rho_2(r) = \tfrac{2}{3}\epsilon(r)\, r \frac{d\rho_0}{dr}. \tag{A, 1}$$

This relation when inserted in equation (20) gives

$$\frac{d}{dr}(r^3\mathfrak{V}_2) = \tfrac{8}{3}\pi G r^4 \int_r^R \epsilon(s)\frac{d\rho_0(s)}{ds}\, ds; \tag{A, 2}$$

and this equation in combination with equation (32) gives

$$4\pi G\int_r^R \epsilon(s)\frac{d\rho_0(s)}{ds}\, ds = -\frac{G}{r^4}\frac{d}{dr}[\epsilon(r)M(r)r^2] + \tfrac{5}{2}\Omega^2. \tag{A, 3}$$

Differentiating this equation with respect to r, we obtain

$$4\pi\epsilon(r)\frac{d\rho_0}{dr} = \frac{d}{dr}\left\{\frac{1}{r^4}\frac{d}{dr}[\epsilon(r)M(r)r^2]\right\}. \tag{A, 4}$$

Equation (A, 4), as can be verified by expansion, is the same as Clairaut's equation; its present derivation emphasizes (what is sometimes obscured) that its validity strictly depends on the equation of state being barotropic.

REFERENCES

Bullard, E. C. 1946, *Geophys. Suppl. M.N.*, **5**, 186.
Chandrasekhar, S. 1933, *M.N.*, **93**, 539.
Chandrasekhar, S., and Lebovitz, N. R. 1962, *Ap. J.*, **136**, 1082.
Darwin, G. H. 1899, *M.N.*, **60**, 92.
Jeffreys, H. 1959, *The Earth* (Cambridge: Cambridge University Press).
Milne, E. A. 1929, *M.N.*, **89**, 739.
Todhunter, I. 1873, *History of the Mathematical Theories of Attraction and the Figure of the Earth* (original ed.; London: Constable & Co., 1873; reprint ed.; New York: Dover Publications, Inc., 1962).

Singularities of Hartmann layers

By P. H. Roberts

University of Newcastle upon Tyne

(*Communicated by K. Stewartson, F.R.S.—Received* 13 *January* 1967)

This paper is concerned with the parallel flow of conducting fluid along an insulating pipe of uniform cross-section perpendicular to which a uniform magnetic field, $\mathbf{B}_0$, is applied. The cross-section is supposed to have tangents parallel to $\mathbf{B}_0$ only at isolated points of its perimeter. The density, kinematic viscosity and electrical conductivity of the fluid are denoted by ρ, ν and σ, respectively. It is known (Shercliff 1962) that, in the limit $B_0 \to \infty$, the flow may be divided into three parts: (i) a Hartmann boundary layer, thickness

$$\sim (\rho\nu/\sigma)^{\frac{1}{2}}(B_0 \cos\theta)^{-1},$$

at every point of the wall except those at which $\cos\theta = 0$, where θ is the angle between $\mathbf{B}_0$ and the normal to the wall at the point concerned, (ii) a 'mainstream', far from the walls, which is controlled by the Hartmann layers and in which a quasi-hydrostatic balance subsists between the Lorentz force and the applied pressure gradient driving the motion, and (iii) a complicated boundary-layer singularity, at each point of the wall at which $\cos\theta = 0$, which is controlled by the flow in regions (i) and (ii). The solutions for regions (i) and (ii) can be obtained easily by Shercliff's methods.

It is the purpose of this paper to elucidate region (iii). Here the boundary-layer thickness is $O(M^{-\frac{1}{2}})$ and extends round the periphery of the wall for a distance which is $O(M^{-\frac{1}{2}})$, where M is a Hartmann number, $B_0\mathscr{L}(\sigma/\rho\nu)^{\frac{1}{2}}$, based on a typical dimension, $\mathscr{L}$, of the pipe. The corresponding contribution to U, the mean flow down the duct, is of order $M^{-\frac{3}{2}}$. In fact, for a circular duct of radius $a(=\mathscr{L})$, the main case discussed in this paper, it contributes the final term to the following expression:

$$U = \frac{64}{3\pi} U_0\left[\frac{1}{M} - \frac{3\pi}{2M^2} + \frac{3{\cdot}273}{M^{\frac{3}{2}}}\right].$$

Here U_0 is the mean flow in the absence of field.

1. Introduction: formal solution

The problem of determining the flow of conducting fluid down an insulating circular duct across which a uniform field, $\mathbf{B}_0$, lies may be reduced to that of solving

$$\frac{\partial^2 v}{\partial y^2} + \frac{\partial^2 v}{\partial z^2} + M\frac{\partial b}{\partial z} = -1, \tag{1}$$

$$\frac{\partial^2 b}{\partial y^2} + \frac{\partial^2 b}{\partial z^2} + M\frac{\partial v}{\partial z} = 0, \tag{2}$$

subject to

$$v = b = 0 \quad \text{on} \quad r = 1, \tag{3}$$

(see for example, Hunt & Stewartson 1965). Here y and z are Cartesian coordinates perpendicular to the flow direction Ox, and Oz is parallel to $\mathbf{B}_0$. The origin O lies at the centre of the tube, $r = \sqrt{(y^2+z^2)}$ is one polar coordinate, and θ is the other (measured from Oz, see figure 1). The variables v and b denote a suitably normalized velocity of flow and induced magnetic field, both in the x direction. The Hartmann number is

$$M = aB_0(\sigma/\rho\nu)^{\frac{1}{2}}, \tag{4}$$

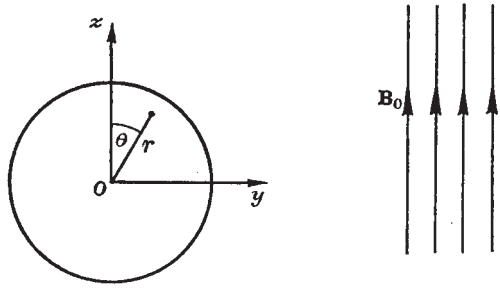


FIGURE 1.

where a is the radius of the pipe, ν is the viscosity of the fluid and σ is its electrical conductivity (m.k.s. units). We shall frequently replace M by 2α:

$$\alpha = \tfrac{1}{2}M. \tag{5}$$

From (1) to (3), we conclude that

$$\left(\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}\right) X + 2\alpha \frac{\partial X}{\partial z} = -1, \tag{6}$$

and
$$X = 0 \quad \text{on} \quad r = 1, \tag{7}$$

where
$$X = v + b. \tag{8}$$

It is likewise easy to write an equation for $v-b$, but it is unnecessary to do so. Because of the symmetry of the duct with respect to the y axis, it is easily shown that

$$\begin{aligned} v(y,z) &= v(-y,z) = \tfrac{1}{2}[X(y,z)+X(-y,z)], \\ b(y,z) &= -b(-y,z) = \tfrac{1}{2}[X(y,z)-X(-y,z)]. \end{aligned} \tag{9}$$

Thus, if X can be found v and b may be extracted from it by taking even and odd parts, respectively.

It is a comparatively simple matter to write down a formal solution to (6) and (7). If we introduce a variable Y by

$$X = -\frac{z}{2\alpha} + \mathrm{e}^{-\alpha z} Y, \tag{10}$$

we obtain, from (6)
$$\left(\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}\right) Y = \alpha^2 Y,$$

or
$$\left(\frac{\partial^2}{\partial r^2}+\frac{1}{r}\frac{\partial}{\partial r}+\frac{1}{r^2}\frac{\partial^2}{\partial \theta^2}\right) Y = \alpha^2 Y, \tag{11}$$

where, by (7) and (10)

$$Y = \frac{1}{2\alpha}\cos\theta\, \mathrm{e}^{\alpha\cos\theta} \quad (\text{on } r = 1). \tag{12}$$

The solutions of (11) which are non-singular in the pipe are of the form

$$Y = I_n(\alpha r) \begin{matrix}\cos \\ \sin\end{matrix} n\theta, \tag{13}$$

where I denotes the modified Bessel function of the first kind.

This suggests that we rewrite (12) in the form (Watson 1944, §2·1)

$$Y = \frac{1}{2\alpha} \sum_{n=0}^{\infty} \epsilon_n I_n'(\alpha) \cos n\theta \quad (\text{on } r = 1), \tag{14}$$

where

$$\epsilon_n = \begin{cases} 1 & \text{if } n = 0, \\ 2 & \text{if } n \neq 0. \end{cases} \tag{15}$$

From (13), we now obtain

$$Y = \frac{1}{2\alpha} \sum_{n=0}^{\infty} \epsilon_n \frac{I_n'(\alpha)}{I_n(\alpha)} I_n(\alpha r) \cos n\theta, \tag{16}$$

$$X = -\frac{1}{2\alpha} r \cos\theta + \frac{1}{2\alpha} e^{-\alpha r \cos\theta} \sum_{n=0}^{\infty} \epsilon_n \frac{I_n'(\alpha)}{I_n(\alpha)} I_n(\alpha r) \cos n\theta. \tag{17}$$

Although (17) provides the solution of (6) and (7), it is in a form which is computationally useless for large values of α, since in these cases the series on the right of (17) is very poorly convergent. It is therefore desirable to obtain the asymptotic expression of X in powers of $1/\alpha$, and, for this purpose, boundary-layer methods are called for.

In the 'mainstream' far from the walls, we may assume that

$$\frac{\partial}{\partial y}, \quad \frac{\partial}{\partial z} = O(1). \tag{18}$$

Then, to leading order, (6) gives

$$\frac{\partial X}{\partial z} = -\frac{1}{2\alpha},$$

i.e.

$$X = -\frac{z}{2\alpha} + \frac{f(y)}{2\alpha}. \tag{19}$$

We can obtain no information about f until the boundary layer has been constructed.

To determine the form of X in the boundary layer, we write

$$X = -\frac{z}{2\alpha} + \frac{f(y)}{2\alpha} + X', \tag{20}$$

where, in order to agree with (19),

$$X' \to 0 \quad \text{as} \quad \alpha(1-r) \to \infty. \tag{21}$$

The fact, used here, that α^{-1} is a measure of the boundary-layer thickness, cannot be properly appreciated until later. Substituting (20) into (6) we obtain, to leading order,

$$\left(\frac{\partial^2}{\partial r^2} + \frac{1}{r}\frac{\partial}{\partial r} + \frac{1}{r^2}\frac{\partial^2}{\partial \theta^2}\right) X' + 2\alpha \left(\cos\theta \frac{\partial}{\partial r} - \frac{\sin\theta}{r}\frac{\partial}{\partial \theta}\right) X' = 0. \tag{22}$$

Let us now examine the form of this equation near a point, θ, on the wall, where $\theta \neq \pm \frac{1}{2}\pi$. We may anticipate that the rate of change of X through the boundary layer will be large compared with that along the boundary layer, i.e.

$$\frac{\partial}{\partial r} \gg \frac{\partial}{\partial \theta} \quad (\alpha \to \infty). \tag{23}$$

On this assumption, which may be checked later for self consistency, (22) reduces, in leading order, to

$$\frac{\partial^2 X'}{\partial r^2} + 2\alpha \cos\theta \frac{\partial X'}{\partial r} = 0, \tag{24}$$

whence

$$X' = g(\theta) + h(\theta)\, e^{2\alpha(1-r)\cos\theta}. \tag{25}$$

There are now two cases to consider:

(i) $-\frac{1}{2}\pi < \theta < \frac{1}{2}\pi$. The exponent appearing in (25) is positive and increases with distance from the walls. The only way in which (25) can agree with (21) is if

$$g = h = 0,$$

i.e.

$$X' = 0 \quad (-\tfrac{1}{2}\pi < \theta < \tfrac{1}{2}\pi). \tag{26}$$

We may now determine f by (7), (20) and (26)

$$f(y) = \sqrt{(1-y^2)}. \tag{27}$$

Thus, in the mainstream,

$$X = [\sqrt{(1-y^2)} - z]/2\alpha \tag{28}$$

to leading order and, in the boundary layer,

$$X = 0. \tag{29}$$

(ii) $\frac{1}{2}\pi < |\theta| < \pi$. To make (25) and (21) agree, we now take $g = 0$, giving

$$X' = h(\theta)\, e^{2\alpha(1-r)\cos\theta} \quad (\tfrac{1}{2}\pi < |\theta| < \pi), \tag{30}$$

and (7), (20), (27) and (30) give

$$h(\theta) = (1/\alpha)\cos\theta, \tag{31}$$

i.e.

$$X = -(1/\alpha)\cos\theta[1 - e^{2\alpha(1-r)\cos\theta}] \quad (\tfrac{1}{2}\pi < |\theta| < \pi). \tag{32}$$

to leading order.

In the subsequent development, it is necessary to take the expansion (28) to the next order. It may be shown that

$$X = \frac{1}{2\alpha}[\sqrt{(1-y^2)} - z]\left[1 - \frac{1}{2\alpha(1-y^2)^{\frac{3}{2}}} + O(\alpha^{-2})\right], \tag{33}$$

(cf. Gold 1962, eq. 27; Shercliff 1962, p. 517). This result holds also, in $|\theta| < \frac{1}{2}\pi$, in the boundary layer.

2. The boundary layer at $\theta = \frac{1}{2}\pi$

The boundary-layer analysis just given breaks down at $\theta = \frac{1}{2}\pi$. To study this region, we introduce two boundary-layer coordinates (ξ, η) to replace (r, θ):

$$r = 1 - \frac{\xi}{(2\alpha^2)^{\frac{1}{3}}}, \quad \theta = \tfrac{1}{2}\pi + \left(\frac{2}{\alpha}\right)^{\frac{1}{3}}\eta. \tag{34}$$

It is clear that (23) still holds, but the justification for the powers, $M^{-\frac{2}{3}}$ and $M^{-\frac{1}{3}}$, postulated in (34) is not apparent until later. It may be noted that, by (34),

$$\left.\begin{aligned} y = r\sin\theta &= 1 - (1/2\alpha^2)^{\frac{1}{3}}(\xi + \eta^2) + O(\alpha^{-\frac{4}{3}}), \\ z = r\cos\theta &= -(2/\alpha)^{\frac{1}{3}}\eta + O(\alpha^{-1}), \\ (1-y^2)^{\frac{1}{2}} &= (2/\alpha)^{\frac{1}{3}}(\xi + \eta^2)^{\frac{1}{2}} + O(\alpha^{-1}), \\ \alpha(1-r)\cos\theta &= -\xi\eta + O(\alpha^{-\frac{2}{3}}). \end{aligned}\right\} \tag{35}$$

On writing

$$X = -\frac{r}{2\alpha}\cos\theta + \frac{X_1}{(2\alpha^2)^{\frac{2}{3}}}, \tag{36}$$

equations (6), (34) and (36) give, to leading order,

$$\frac{\partial^2 X_1}{\partial \xi^2} + 2\eta \frac{\partial X_1}{\partial \xi} = \frac{\partial X_1}{\partial \eta}. \tag{37}$$

If we introduce a new variable Y_1, by

$$X_1 = \mathrm{e}^{-\xi\eta - \frac{1}{3}\eta^3} Y_1, \tag{38}$$

equation (37) becomes

$$\frac{\partial^2 Y_1}{\partial \xi^2} + \xi Y_1 = \frac{\partial Y_1}{\partial \eta}, \tag{39}$$

where, by (7), (34), (36) and (38),

$$Y_1 = -\eta \mathrm{e}^{\frac{1}{3}\eta^3} \quad \text{on} \quad \xi = 0. \tag{40}$$

In addition to obeying (40), Y_1, must match to the mainstream and boundary layers constructed in §1. This requires (cf. (33), (29), (32), (35), (36) and (38))

$$Y_1 \sim \mathrm{e}^{\xi\eta + \frac{1}{3}\eta^3}\left[\xi^{\frac{1}{2}} + \frac{\eta^2}{2\xi^{\frac{1}{2}}} - \frac{1}{4\xi} + O(\xi^{-\frac{3}{2}})\right] \quad (\xi \to \infty); \tag{41}$$

$$Y_1 \sim -\mathrm{e}^{\xi\eta + \frac{1}{3}\eta^3}\left[\eta + \frac{\xi}{2\eta} + O(\eta^{-2})\right] \quad (\eta \to -\infty); \tag{42}$$

$$Y_1 \sim -2\mathrm{e}^{-\xi\eta + \frac{1}{3}\eta^3}[\eta + O(\eta^{-1})] + \mathrm{e}^{\xi\eta + \frac{1}{3}\eta^3}\left[\eta + \frac{\xi}{2\eta} + O(\eta^{-2})\right] \quad (\eta \to +\infty). \tag{43}$$

We introduce the three cube roots of unity: $1, \omega, \omega^2$, where

$$\omega = \mathrm{e}^{\frac{2}{3}\pi \mathrm{i}}, \quad \omega^2 = \mathrm{e}^{-\frac{2}{3}\pi \mathrm{i}}, \tag{44}$$

and observe that (39) has solutions of the form

$$\mathrm{e}^{\eta\lambda x}\mathrm{Ai}(\lambda x - \xi), \quad \mathrm{e}^{\eta\lambda x}\mathrm{Ai}(\omega\lambda x - \omega\xi), \quad \mathrm{e}^{\eta\lambda x}\mathrm{Ai}\,(\omega^2\lambda x - \omega^2\xi), \tag{45}$$

for any λ, where $\mathrm{Ai}\,(z)$ denotes the Airy function

$$\mathrm{Ai}\,(z) = \frac{1}{\pi}\left(\frac{z}{3}\right)^{\frac{1}{2}} K_{\frac{1}{3}}\,(\tfrac{2}{3}z^{\frac{3}{2}}), \tag{46}$$

and K denotes the modified Bessel function of the second kind. Only two of the triad (45) are independent, since

$$\mathrm{Ai}(x) + \omega\,\mathrm{Ai}\,(\omega x) + \omega^2\,\mathrm{Ai}\,(\omega^2 x) = 0 \quad (\text{all } x). \tag{47}$$

It is therefore sometimes advantageous to introduce the second Airy function:

$$\begin{aligned}\mathrm{Bi}\,(z) &= \mathrm{i}\omega^2\,\mathrm{Ai}\,(\omega^2 z) - \mathrm{i}\omega\,\mathrm{Ai}\,(\omega z)\\ &= \frac{1}{\pi}\left(\frac{z}{3}\right)^{\frac{1}{2}}[2\pi I_{\frac{1}{3}}(\tfrac{2}{3}z^{\frac{3}{2}}) + 3^{\frac{1}{2}}K_{\frac{1}{3}}(\tfrac{2}{3}z^{\frac{3}{2}})],\end{aligned} \tag{48}$$

and to treat $\mathrm{Ai}\,(z)$ and $\mathrm{Bi}\,(z)$ as the independent solutions of Airy's equation.

We now observe that

$$\begin{aligned}
&\int_0^\infty \mathrm{Ai}'(x)\,[\mathrm{e}^{\eta x}+\omega^2\mathrm{e}^{\eta\omega x}+\omega\mathrm{e}^{\eta\omega^2 x}]\,\mathrm{d}x\\
&\quad=-\frac{1}{\pi\sqrt{3}}\int_0^\infty K_{\frac{2}{3}}(\tfrac{2}{3}x^{\frac{3}{2}})\,[\mathrm{e}^{\eta x}+\omega^2\mathrm{e}^{\eta\omega x}+\omega\mathrm{e}^{\eta\omega^2 x}]\,x^{\frac{1}{2}}\,\mathrm{d}x\\
&\quad=-\frac{1}{\pi\sqrt{3}}\sum_{m=0}^{\infty}\frac{\eta^m}{m!}\,(1+\omega^{m+2}+\omega^{2m+1})\int_0^\infty K_{\frac{2}{3}}(\tfrac{2}{3}x^{\frac{3}{2}})\,x^{m+1}\,\mathrm{d}x\\
&\quad=-\frac{\sqrt{3}}{\pi}\sum_{m=0}^{\infty}\frac{\eta^{3m+1}}{(3m+1)!}\left(\frac{3}{2}\right)^{2m+1}\int_0^\infty K_{\frac{2}{3}}(y)\,y^{2m+1}\,\mathrm{d}y\\
&\quad=-\frac{\sqrt{3}}{\pi}\sum_{m=0}^{\infty}\frac{2\pi\eta^{3m+1}}{3^{3m+\frac{3}{2}}m!\,\Gamma(m+\frac{2}{3})\,\Gamma(m+\frac{4}{3})}\left(\frac{3}{2}\right)^{2m+1}2^{2m}\,\Gamma(m+\tfrac{2}{3})\,\Gamma(m+\tfrac{4}{3})\\
&\quad=-\eta\sum_{m=0}^{\infty}\frac{1}{m!}\,(\tfrac{1}{3}\eta^3)^m=-\eta\,\mathrm{e}^{\frac{1}{3}\eta^3},
\end{aligned}\tag{49}$$

where we have used a property of $K_{\frac{2}{3}}$ listed in Watson (1944, p. 388). Thus, we may replace (40) by

$$Y_1=\int_0^\infty \mathrm{Ai}'(x)\,[\mathrm{e}^{\eta x}+\omega^2\mathrm{e}^{\eta\omega x}+\omega\mathrm{e}^{\eta\omega^2 x}]\,\mathrm{d}x\quad\text{on}\quad\xi=0.\tag{50}$$

This strongly suggests that we should take $\lambda=1,\omega,\omega^2$ successively in (45) and seek a solution which is a linear combination of integrals of the type

$$\int_0^\infty \mathrm{Ai}'(x)\,\frac{\mathrm{Ai}(x-\xi)}{\mathrm{Ai}(x)}\,\mathrm{e}^{\eta x}\,dx,\ldots.\tag{51}$$

The correct linear combination, as we will now see, is

$$\begin{aligned}
Y_1=\omega\int_0^\infty\frac{\mathrm{Ai}'(x)}{\mathrm{Ai}(x)}\,\mathrm{Ai}(x-\omega\xi)\,\mathrm{e}^{\eta\omega^2 x}\,\mathrm{d}x+\omega^2\int_0^\infty\frac{\mathrm{Ai}'(x)}{\mathrm{Ai}(x)}\,\mathrm{Ai}(x-\omega^2\xi)\,\mathrm{e}^{\eta\omega x}\,\mathrm{d}x\\
-\int_0^\infty\left[\omega^2\frac{\mathrm{Ai}'(\omega x)}{\mathrm{Ai}(\omega x)}\,\mathrm{Ai}(\omega x-\omega\xi)+\omega\frac{\mathrm{Ai}'(\omega^2 x)}{\mathrm{Ai}(\omega^2 x)}\,\mathrm{Ai}(\omega^2 x-\omega^2\xi)\right]\mathrm{e}^{\eta x}\,\mathrm{d}x.
\end{aligned}\tag{52}$$

It is clear, by (47), that (52) obeys (50). What is not so clear is the convergence of the last integral of (52) at large x for $\xi\neq 0$ (see (55a) below). To examine this question, we use the relations

$$\left.\begin{aligned}
\mathrm{Ai}(\omega x)&=-\tfrac{1}{2}\omega^2[\mathrm{Ai}(x)-\mathrm{iBi}(x)],\\
\mathrm{Ai}'(\omega x)&=-\tfrac{1}{2}\omega[\mathrm{Ai}'(x)-\mathrm{iBi}'(x)],
\end{aligned}\right\}\tag{53}$$

and the conjugate complex relations for $\mathrm{Ai}(\omega^2 x)$ and $\mathrm{Ai}'(\omega^2 x)$. Equation (52) may then be written in the form

$$\begin{aligned}
Y_1&=\int_0^\infty\frac{\mathrm{Ai}(x)\,\mathrm{Ai}'(x)+\mathrm{Bi}(x)\,\mathrm{Bi}'(x)}{\mathrm{Ai}^2(x)+\mathrm{Bi}^2(x)}\,\mathrm{Ai}(x-\xi)\,\mathrm{e}^{\eta x}\,\mathrm{d}x-\frac{1}{\pi}\int_0^\infty\frac{\mathrm{Bi}(x-\xi)}{\mathrm{Ai}^2(x)+\mathrm{Bi}^2(x)}\,\mathrm{e}^{\eta x}\,\mathrm{d}x\\
&\quad+\omega\int_0^\infty\frac{\mathrm{Ai}'(x)}{\mathrm{Ai}(x)}\,\mathrm{Ai}(x-\omega\xi)\,\mathrm{e}^{\eta\omega^2 x}\,\mathrm{d}x+\omega^2\int_0^\infty\frac{\mathrm{Ai}'(x)}{\mathrm{Ai}(x)}\,\mathrm{Ai}(x-\omega^2\xi)\,\mathrm{e}^{\eta\omega x}\,\mathrm{d}x\\
&=Y_a+Y_b+Y_c+Y_d,\quad\text{say.}
\end{aligned}\tag{54}$$

Recalling now that

$$\mathrm{Ai}(x) \sim \frac{1}{2\pi^{\frac{1}{2}}x^{\frac{1}{4}}}\,\mathrm{e}^{-\frac{2}{3}x^{\frac{3}{2}}}\left[1-\frac{5}{48x^{\frac{3}{2}}}+O(x^{-3})\right] \quad (|x|\to\infty, |\arg x|<\pi), \tag{55a}$$

$$\mathrm{Bi}(x) \sim \frac{1}{\pi^{\frac{1}{2}}x^{\frac{1}{4}}}\,\mathrm{e}^{\frac{2}{3}x^{\frac{3}{2}}}\left[1+\frac{5}{48x^{\frac{3}{2}}}+O(x^{-3})\right] \quad (|x|\to\infty, |\arg x|<\tfrac{1}{3}\pi), \tag{55b}$$

we see that the integrands of Y_a and Y_b are respectively

$$\frac{x^{\frac{1}{2}}}{2\pi^{\frac{1}{2}}(x-\xi)^{\frac{1}{4}}}\,\mathrm{e}^{-\frac{2}{3}(x-\xi)^{\frac{3}{2}}+\eta x}, \quad -\frac{x^{\frac{1}{2}}}{\pi^{\frac{1}{2}}(x-\xi)^{\frac{1}{4}}}\,\mathrm{e}^{\frac{2}{3}(x-\xi)^{\frac{3}{2}}-\frac{2}{3}x^{\frac{3}{2}}+\eta x}, \tag{56a,b}$$

asymptotically, as $x\to+\infty$. The integrals, are, therefore, convergent.

The combination (52) is by no means the only one that will satisfy (50). Before we may conclude that it is a correct solution of the problem, we must, of course, prove that it agrees with (41) to (43). Consider first (43), i.e. the behaviour of Y_1 as $\eta\to+\infty$ for fixed ξ. The dominant contribution to Y_1 arises from Y_a and Y_b, since these have large positive exponents. From (56a), we see that Y_a has a saddle point at

$$x=\eta^2+\xi+O(\eta^{-1}).$$

The contribution made to Y_1 is

$$Y_a \sim \eta\,\mathrm{e}^{\xi\eta+\frac{1}{3}\eta^3}+O(\eta^{-1}) \quad (\eta\to+\infty). \tag{57}$$

From (56b), we see that Y_b has a saddle point at

$$x=\eta^2-\xi+O(\eta^{-1}).$$

The contribution made to Y_1 is

$$Y_b \sim -2\eta\,\mathrm{e}^{-\xi\eta+\frac{1}{3}\eta^3}+O(\eta^{-1}) \quad (\eta\to+\infty). \tag{58}$$

The contributions (57) and (58) agree with (43).

Consider next (42), i.e. the behaviour of Y_1 as $\eta\to-\infty$ for fixed ξ. The exponents of η in the integrands of Y_c and Y_d now have positive real parts, and are certainly not negligible. Consider first Y_c. Since the poles of the integrand are real and negative, we may distort the contour in the complex x plane from its present position (the positive real axis) to that shown in figure 2, in which P denotes a point ($t=t_P$) to be chosen presently, and OP is the line $x=\omega t\,(t>0)$. The contribution $Y_c^{(1)}$ to Y_c from OP is

$$Y_c^{(1)}=\omega^2\int_0^{t_P}\frac{\mathrm{Ai}'(\omega t)}{\mathrm{Ai}(\omega t)}\,\mathrm{Ai}(\omega t-\omega\xi)\,\mathrm{e}^{\eta t}\,\mathrm{d}t. \tag{59}$$

For large t, the integrand of (59) is asymptotically equal to

$$\frac{\mathrm{i}t^{\frac{1}{2}}}{2\pi^{\frac{1}{2}}(t-\xi)^{\frac{1}{4}}}\,\mathrm{e}^{\frac{2}{3}(t-\xi)^{\frac{3}{2}}+\eta t}, \tag{60}$$

which reveals a saddle point at $\eta^2+\xi+O(1/\eta)$. We choose P to be this saddle point; i.e. to an adequate approximation we take

$$t_P=\eta^2+\xi, \tag{61}$$

and choose PQ (see figure 2) to be, at P, along the line of steepest descent from the saddle point:

$$x = \omega(\xi+\eta^2) + e^{\frac{1}{3}\pi i} s \quad (0 < s < \infty). \tag{62}$$

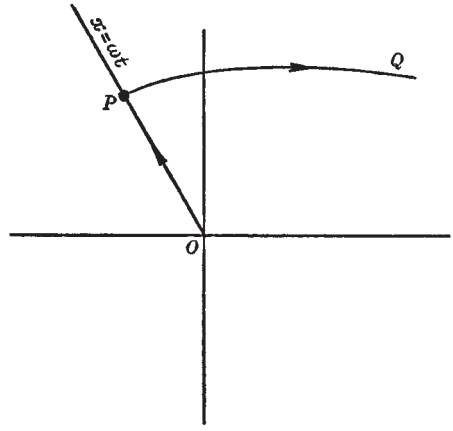


FIGURE 2. The complex x plane.

The modulus of the integrand of Y_c now decreases monotonically as the path OPQ is traversed. The greatest contribution to the integral arises from OP, but this, as we will presently see, cancels with like contributions from the other integrals. It is therefore of interest to evaluate the contribution $Y_c^{(2)}$, to Y_c from PQ. To this end, we substitute (62) in the definition of Y_c contained in (54), giving

$$\begin{aligned} Y_c^{(2)} &\sim \frac{(-\eta)^{\frac{1}{2}}}{2\pi^{\frac{1}{2}}} e^{\xi\eta+\frac{1}{3}\eta^3} \int_0^\infty e^{s^2/4\eta}\, ds + O(\eta^{-1}) \\ &= -\tfrac{1}{2}\eta\, e^{\xi\eta+\frac{1}{3}\eta^3} + O(\eta^{-1}) \quad (\eta \to -\infty). \end{aligned} \tag{63}$$

It is easily shown that the complex conjugate, $Y_d^{(2)}$, makes the same contribution to Y_1, i.e.

$$Y^{(2)} \equiv Y_c^{(2)} + Y_d^{(2)} \sim -\eta\, e^{\xi\eta+\frac{1}{3}\eta^3} + O(\eta^{-1}) \quad (\eta \to -\infty). \tag{64}$$

Consider next $Y_c^{(1)}$ and $Y_d^{(1)}$. By (59) and the corresponding expression for $Y_d^{(1)}$, we have

$$Y_c^{(1)} + Y_d^{(1)} = \int_0^{\xi+\eta^3} \left[\omega^2 \frac{\mathrm{Ai}'(\omega x)}{\mathrm{Ai}(\omega x)} \mathrm{Ai}(\omega x - \omega\xi) + \omega \frac{\mathrm{Ai}'(\omega^2 x)}{\mathrm{Ai}(\omega^2 x)} \mathrm{Ai}(\omega^2 x - \omega^2 \xi)\right] e^{\eta x}\, dx,$$

whence, taking the same steps that led from (52) to (54),

$$\begin{aligned} Y_1 = Y^{(2)} &+ \int_{\xi+\eta^3}^\infty \frac{\mathrm{Ai}(x)\,\mathrm{Ai}'(x) + \mathrm{Bi}(x)\,\mathrm{Bi}'(x)}{\mathrm{Ai}^2(x) + \mathrm{Bi}^2(x)} \mathrm{Ai}(x-\xi)\, e^{\eta x}\, dx \\ &+ \frac{1}{\pi} \int_{\xi+\eta^3}^\infty \frac{\mathrm{Bi}(x-\xi)}{\mathrm{Ai}^2(x) + \mathrm{Bi}^2(x)} e^{\eta x}\, dx. \end{aligned} \tag{65}$$

By (56), the contributions made by the two integrals on the right of (65) are, for large $|\eta|$, respectively of order

$$e^{\xi\eta+\frac{2}{3}\eta^3}, \quad e^{\xi\eta+\frac{1}{3}\eta^3-\frac{4}{3}(\xi+\eta^2)^{\frac{3}{2}}},$$

both of which are negligible compared with (64). Thus we finally obtain $Y_1 \sim Y^{(2)}$, i.e.

$$Y_1 \sim -\eta e^{\xi\eta+\frac{1}{3}\eta^3} + O(\eta^{-1}) \quad (\eta \to -\infty). \tag{66}$$

This agrees with (42).

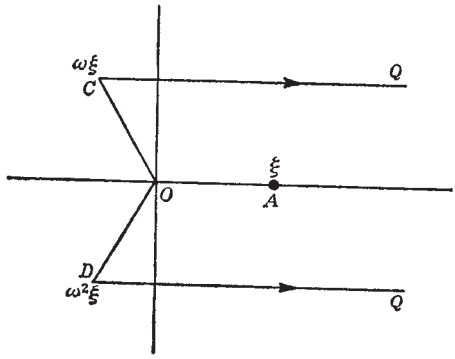


FIGURE 3. The complex x plane.

Finally consider (41), i.e. the behaviour of Y_1 as $\xi \to \infty$ for fixed η. We distort the contour for Y_c to OCQ where C is the point $\omega\xi$ and that for Y_d to ODQ where D is the point $\omega^2\xi$ (see figure 3). As in the case of $\eta \to -\infty$ considered above, the contribution to Y_1 from the rays OC and OD cancel with that from Y_a and Y_b over OA, where A is the point ξ. The remaining arcs give

$$\begin{aligned} e^{-\xi\eta}Y_1 = {} & \omega\int_0^\infty \frac{\mathrm{Ai}'(\omega\xi+s)}{\mathrm{Ai}(\omega\xi+s)}\mathrm{Ai}(s)\,e^{\eta\omega^2 s}\,ds + \omega^2\int_0^\infty \frac{\mathrm{Ai}'(\omega^2\xi+s)}{\mathrm{Ai}(\omega^2\xi+s)}\mathrm{Ai}(s)\,e^{\eta\omega s}\,ds \\ & -\int_0^\infty \left[\omega^2\frac{\mathrm{Ai}'(\omega\xi+\omega s)}{\mathrm{Ai}(\omega\xi+\omega s)}\mathrm{Ai}(\omega s) + \omega\frac{\mathrm{Ai}'(\omega^2\xi+\omega^2 s)}{\mathrm{Ai}(\omega^2\xi+\omega^2 s)}\mathrm{Ai}(\omega^2 s)\right] e^{\eta s}\,ds. \end{aligned} \tag{67}$$

By (55a), we have

$$\frac{\mathrm{Ai}'(x)}{\mathrm{Ai}(x)} \sim -x^{\frac{1}{2}} - \frac{1}{4x} + O(x^{-\frac{5}{2}}) \quad (|x| \to \infty, |\arg x| < \pi). \tag{68}$$

Expanding the integrand in (67), using (47) and (68), we obtain

$$\begin{aligned} e^{-\xi\eta}Y_1 \sim {} & \left(\xi^{\frac{1}{2}} - \frac{1}{4\xi}\right)\int_0^\infty \mathrm{Ai}(s)\left[e^{\eta s} + e^{\eta\omega s} + e^{\eta\omega^2 s}\right] ds \\ & + \frac{1}{2\xi^{\frac{1}{2}}}\int_0^\infty s\,\mathrm{Ai}(s)\left[e^{\eta s} + \omega e^{\eta\omega s} + \omega^2 e^{\eta\omega^2 s}\right] ds + O(\xi^{-\frac{3}{2}}) \\ = {} & e^{\frac{1}{3}\eta^3}\left[\xi^{\frac{1}{2}} + \frac{\eta^2}{2\xi^{\frac{1}{2}}} - \frac{1}{4\xi} + O(\xi^{-\frac{3}{2}})\right] \quad (\xi \to \infty). \end{aligned} \tag{69}$$

In reducing the expression to the final line, we have twice used the expansion method by which (49) was derived. This result agrees with (41).

4. The flux deficiency created by the singularities

The net flux of fluid down the pipe is

$$Q = \int_0^1 \mathrm{d}r \int_0^{2\pi} \mathrm{d}\theta(rv) = 2\int_0^1 \mathrm{d}r \int_0^{\frac{1}{2}\pi} r\,\mathrm{d}\theta\,[X(r,\theta)+X(r,\pi-\theta)], \tag{70}$$

where we have used (9). Were it not for the boundary layers, Q could be obtained by replacing X in (70) by its mainstream value, $X_{\mathrm{m.s.}}$. In fact, integration of (33) yields a contribution to Q of

$$\begin{aligned} Q_{\mathrm{m.s.}} &= 2\mathscr{F}\int_0^1 \mathrm{d}y \int_0^{(1-y^2)^{\frac{1}{2}}} \frac{\mathrm{d}z}{\alpha}\left[(1-y^2)^{\frac{1}{2}} - \frac{1}{2\alpha}(1-y^2)^{-1} + O(\alpha^{-2})\right] \\ &= \frac{4}{3\alpha} - \frac{\pi}{2\alpha^2} + O(\alpha^{-3}). \end{aligned} \tag{71}$$

Here $\mathscr{F}$ denotes the convergent part of the integral which follows it. In this connexion, we should note that, beyond the terms shown explicitly, expression (33) diverges when the integration (71) is performed, owing to the presence of a term in $(1-y^2)^{-\frac{3}{2}}$ in the expression for $X_{\mathrm{m.s.}}$. Of course, the integral (70) does not diverge, since the solution near the boundary-layer singularities $\theta = \pm\frac{1}{2}\pi$ is, unlike (33), bounded.

Equation (71) gives the leading term of Q correctly. It does not, however, give the higher order terms. The wall of the duct brings the fluid, with which it is in contact, to rest, and Q is, in fact, smaller than $Q_{\mathrm{m.s.}}$. The qualitative effects of the boundary layers are easy to estimate in order of magnitude. The fluid velocity, v_h, in the Hartmann layers differs from $v_{\mathrm{m.s.}}$ by $O(\alpha^{-1})$. Since, however, the thickness of the Hartmann layer is $O(\alpha^{-1})$ and its θ extent is $O(1)$, the correction it induces in Q is

$$O(\alpha^{-1}) \times O(\alpha^{-1}) \times O(1) = O(\alpha^{-2}).$$

(Thus the second term of (71) does not give the second term of Q correctly.) Similarly, near the singularities $\theta = \pm\frac{1}{2}\pi$, the flow velocity differs from $v_{\mathrm{m.s.}}$ and from v_h by $O(\alpha^{-\frac{4}{3}})$. Since the thickness (in the r-direction) of the singularity is $O(\alpha^{-\frac{2}{3}})$, and since its lateral extent (in the θ direction) is $O(\alpha^{-\frac{1}{3}})$, the correction to Q made by the neighbourhood of the singularity is

$$O(\alpha^{-\frac{4}{3}}) \times O(\alpha^{-\frac{2}{3}}) \times O(\alpha^{-\frac{1}{3}}) = O(\alpha^{-\frac{7}{3}}).$$

It is the objective of this section to evaluate this term explicitly.

We commence by studying Q_h, the term which must be added to $Q_{\mathrm{m.s.}}$ to allow for the Hartmann layers. This is not merely a matter of integrating v through these layers; we have already (over)allowed for the flow in them by integrating $v_{\mathrm{m.s.}}$ through them. Thus, in place of (70), we must take (cf. (19) and (20))

$$Q_h = \frac{2}{\alpha}\int_0^\infty \mathrm{d}[\alpha(1-r)] \int_0^{\frac{1}{2}\pi} \mathrm{d}\theta\,[X'(r,\theta)+X'(r,\pi-\theta)], \tag{72}$$

where we have extended the integration in r over the entire boundary-layer variable $\alpha(1-r)$. Using (29) to (31) we now obtain

$$\begin{aligned} Q_h &= -\frac{2}{\alpha^2}\mathscr{F}\int_0^\infty \mathrm{d}[\alpha(1-r)] \int_0^{\frac{1}{2}\pi} \mathrm{d}\theta \cos\theta\, \mathrm{e}^{-2\alpha(1-r)\cos\theta}\,[1+O(\alpha^{-1})] \\ &= -\frac{\pi}{2\alpha^2} + O(\alpha^{-3}), \end{aligned} \tag{73}$$

which to leading order is, in fact, equal to the second term of (71).

Next, we turn to the evaluation of Q_s, the term which must be added to $Q_{\text{m.s.}}+Q_h$ to allow for the boundary-layer singularities at $\theta = \pm\frac{1}{2}\pi$. Again, this is not merely a matter of integrating v through these layers; we must first subtract the flow already included in the $Q_{\text{m.s.}}$ and Q_h integrations. We obtain (cf. (31), (33) and (35))

$$\begin{aligned} Q_s &= \frac{2}{\alpha}\int_0^\infty \mathrm{d}\xi \int_0^\infty \mathrm{d}\eta \left[\left\{X(\xi,\eta) - \frac{1}{(2\alpha^2)^{\frac{2}{3}}}\left[(\xi+\eta^2)^{\frac{1}{2}} - \frac{1}{4(\xi+\eta^2)}\right]\right\}\right. \\ &\qquad \left. + \left\{X(\xi,-\eta) - \frac{1}{(2\alpha^2)^{\frac{2}{3}}}\left[(\xi+\eta^2)^{\frac{1}{2}} - \frac{1}{4(\xi+\eta^2)} - 2\eta\,\mathrm{e}^{-2\xi\eta}\right]\right\}\right] \\ &= \frac{2^{\frac{1}{3}}}{\alpha^{\frac{7}{3}}}\int_0^\infty \mathrm{d}\xi \int_0^\infty \mathrm{d}\eta \left[Y_1(\xi,\eta)\,\mathrm{e}^{-\xi\eta-\frac{1}{3}\eta^3} + Y_1(\xi,-\eta)\,\mathrm{e}^{\xi\eta+\frac{1}{3}\eta^3}\right. \\ &\qquad \left. - 2(\xi+\eta^2)^{\frac{1}{2}} + \frac{1}{2(\xi+\eta^2)} + 2\eta\,\mathrm{e}^{-2\xi\eta}\right]. \end{aligned} \tag{74}$$

If we use Y_1 in the form (67), the ξ-integration is immediate:

$$\begin{aligned} &\int_0^\Xi Y_1(\xi,\eta)\,\mathrm{e}^{-\xi\eta}\,\mathrm{d}\xi \\ &= \int_0^\infty \mathrm{d}s\,\mathrm{Ai}\,(s)\left[\mathrm{e}^{\eta\omega^2 s}\ln\left\{2\pi^{\frac{1}{2}}\,\mathrm{e}^{\frac{1}{6}\pi\mathrm{i}}\,\mathrm{Ai}\,(\omega\xi+s)\right\} + \mathrm{e}^{\eta\omega s}\ln\left\{2\pi^{\frac{1}{2}}\mathrm{e}^{-\frac{1}{6}\pi\mathrm{i}}\,\mathrm{Ai}(\omega^2\xi+s)\right\}\right]_0^\Xi \\ &\quad - \int_0^\infty \mathrm{d}s\,\mathrm{e}^{\eta s}\left[\omega\mathrm{Ai}(\omega s)\ln\left\{2\pi^{\frac{1}{2}}\,\mathrm{e}^{\frac{1}{6}\pi\mathrm{i}}\mathrm{Ai}(\omega\xi+\omega s)\right\} + \omega^2\mathrm{Ai}(\omega^2 s)\ln\left\{2\pi^{\frac{1}{2}}\mathrm{e}^{-\frac{1}{6}\pi\mathrm{i}}\mathrm{Ai}(\omega^2\xi+\omega^2 s)\right\}\right]^{[1]} \end{aligned} \tag{75}$$

where we have included constant terms in the logarithms so that there is no constant term arising from the upper limit as $\Xi\to\infty$. In fact, by using (55), (49) and the expansion method by which (49) itself was derived, we find that

$$\int_0^\Xi Y_1(\xi,\eta)\,\mathrm{e}^{-\xi\eta}\,\mathrm{d}\xi \sim \mathrm{e}^{\frac{1}{3}\eta^3}\left[\tfrac{2}{3}\Xi^{\frac{3}{2}} + \eta^2\Xi^{\frac{1}{2}} - \tfrac{1}{4}\ln\Xi + O(\Xi^{-\frac{1}{2}})\right] + Z_1(\eta), \quad \text{as} \quad \Xi\to\infty, \tag{76}$$

where

$$\begin{aligned} Z_1(\eta) &= \int_0^\infty \mathrm{d}s\,\mathrm{e}^{\eta s}\{\omega\,\mathrm{Ai}\,(\omega s)\ln\left[2\pi^{\frac{1}{2}}\,\mathrm{e}^{\frac{1}{6}\pi\mathrm{i}}\,\mathrm{Ai}\,(\omega s)\right] + \omega^2\,\mathrm{Ai}\,(\omega^2 s)\ln\left[2\pi^{\frac{1}{2}}\,\mathrm{e}^{-\frac{1}{6}\pi\mathrm{i}}\,\mathrm{Ai}\,(\omega^2 s)\right]\} \\ &\quad - \int_0^\infty \mathrm{d}s\,\mathrm{Ai}(s)\left\{\mathrm{e}^{\eta\omega^2 s}\ln\left[2\pi^{\frac{1}{2}}\,\mathrm{e}^{\frac{1}{6}\pi\mathrm{i}}\,\mathrm{Ai}(s)\right] + \mathrm{e}^{\eta\omega s}\ln\left[2\pi^{\frac{1}{2}}\,\mathrm{e}^{-\frac{1}{6}\pi\mathrm{i}}\,\mathrm{Ai}\,(s)\right]\right\}. \end{aligned} \tag{77}$$

Using (76) we readily find that (74) may be rewritten

$$Q_s = \frac{2^{\frac{1}{3}}}{\alpha^{\frac{7}{3}}}\int_0^\infty \mathrm{d}\eta\left[Z_1(\eta)\,\mathrm{e}^{-\frac{1}{3}\eta^3} + Z_1(-\eta)\,\mathrm{e}^{\frac{1}{3}\eta^3} + \tfrac{4}{3}\eta^3 - \ln\eta + 1\right]. \tag{78}$$

By (77) and (78), Q_s may be expressed as a double integral over s and η. By reversing the order of these integrations and making use of the integral representations

$$\mathrm{Ai}\,(s) = \frac{1}{\pi}\int_0^\infty \cos\tfrac{1}{3}(\eta^3+\eta s)\,\mathrm{d}\eta, \tag{79}$$

$$\mathrm{Ai}\,(\omega s) = \frac{1}{2\pi}\omega^2\int_0^\infty \left[\mathrm{i}\mathrm{e}^{-\frac{1}{3}\eta^3+\eta s} - \mathrm{e}^{-\mathrm{i}(\frac{1}{3}\eta^3+\eta s)}\right]\mathrm{d}\eta, \tag{80}$$

$$\mathrm{Ai}\,(\omega^2 s) = -\frac{1}{2\pi}\omega\int_0^\infty \left[\mathrm{i}\mathrm{e}^{-\frac{1}{3}\eta^3+\eta s} + \mathrm{e}^{\mathrm{i}(\frac{1}{3}\eta^3+\eta s)}\right]\mathrm{d}\eta, \tag{81}$$

it is possible to perform the η-integration explicitly, as follows. We first divide (78) into four parts:

$$Q_s = \frac{2^{\frac{1}{3}}}{\alpha^{\frac{2}{3}}}[I_1 + I_2 + I_2^* + I_3], \tag{82}$$

where
$$I_1 = \int_0^\infty \mathrm{d}\eta \left[\left\{\int_0^\infty \mathrm{e}^{-\frac{1}{3}\eta^3+\eta s}[\omega \operatorname{Ai}(\omega s)\ln\{2\pi^{\frac{1}{2}}\mathrm{e}^{\frac{1}{6}\pi \mathrm{i}}\operatorname{Ai}(\omega s)\}] + \omega^2 \operatorname{Ai}(\omega^2 s)\ln\{2\pi^{\frac{1}{2}}\mathrm{e}^{-\frac{1}{6}\pi \mathrm{i}}\operatorname{Ai}(\omega^2 s)\}\right\} + \tfrac{2}{3}\eta^3 - \tfrac{1}{2}\ln\eta + 1\right], \tag{83}$$

$$I_2 = -\int_0^\infty \mathrm{d}\eta \int_0^\infty \mathrm{e}^{-\frac{1}{3}\eta^3+\eta\omega^2 s}\operatorname{Ai}(s)\ln\{2\pi^{\frac{1}{2}}\mathrm{e}^{\frac{1}{6}\pi \mathrm{i}}\operatorname{Ai}(s)\}\,\mathrm{d}s, \tag{84}$$

$$\begin{aligned} I_3 = \int_0^\infty \mathrm{d}\eta \Bigg[\Bigg\{\int_0^\infty & \mathrm{e}^{\frac{1}{3}\eta^3-\eta s}[\omega \operatorname{Ai}(\omega s)\ln\{2\pi^{\frac{1}{2}}\mathrm{e}^{\frac{1}{6}\pi \mathrm{i}}\operatorname{Ai}(\omega s)\} \\ & + \omega^2 \operatorname{Ai}(\omega^2 s)\ln\{2\pi^{\frac{1}{2}}\mathrm{e}^{-\frac{1}{6}\pi \mathrm{i}}\operatorname{Ai}(\omega^2 s)\}]\,\mathrm{d}s \\ & - \int_0^\infty \mathrm{e}^{\frac{1}{3}\eta^3-\eta\omega^2 s}\operatorname{Ai}(s)\ln\{2\pi^{\frac{1}{2}}\mathrm{e}^{\frac{1}{6}\pi \mathrm{i}}\operatorname{Ai}(s)\}\,\mathrm{d}s \\ & - \int_0^\infty \mathrm{e}^{\frac{1}{3}\eta^3-\eta\omega s}\operatorname{Ai}(s)\ln\{2\pi^{\frac{1}{2}}\mathrm{e}^{-\frac{1}{6}\pi \mathrm{i}}\operatorname{Ai}(s)\}\,\mathrm{d}s\Bigg\} + \tfrac{2}{3}\eta^3 - \tfrac{1}{2}\ln\eta\Bigg]. \end{aligned} \tag{85}$$

It may be shown that the three integrals concerned are convergent. It is clear that I_1 and I_2 arise from the integration of $Z_1(\eta)$, and I_3 from $Z_1(-\eta)$.

The inversion of the order of integration in (83) to (85) is not entirely straightforward and, in what follows, we will find it convenient to regard each integral as being the limit, as $\rho \to \infty$, of an η-integration from 0 to ρ and an s-integration from 0 to ρ^2. We see immediately that

$$I_1 = \lim_{\rho\to\infty}\int_0^{\rho^2} \mathrm{d}s \left[\left\{[\omega \operatorname{Ai}(\omega s)\ln\{2\pi^{\frac{1}{2}}\mathrm{e}^{\frac{1}{6}\pi \mathrm{i}}\operatorname{Ai}(\omega s)\} + \omega^2 \operatorname{Ai}(\omega^2 s)\ln\{2\pi^{\frac{1}{2}}\mathrm{e}^{-\frac{1}{6}\pi \mathrm{i}}\operatorname{Ai}(\omega^2 s)\}] \times \int_0^\rho \mathrm{e}^{\eta s-\frac{1}{3}\eta^3}\,\mathrm{d}\eta\right\} + \tfrac{1}{3}s - \frac{1}{8s^{\frac{1}{2}}}\ln s + \frac{1}{2s^{\frac{1}{2}}}\right]. \tag{86}$$

For I_2, we distort the contour in the η plane from the real axis to a path consisting of the ray $\eta = \eta'\,\mathrm{e}^{\frac{1}{6}\pi \mathrm{i}}$ from $\eta' = 0$ to $\eta' = \rho$, followed by a ϕ-integration round the circular arc $\eta = \rho\,\mathrm{e}^{\mathrm{i}\phi}$ from $\phi = 0$ to $\phi = \frac{1}{6}\pi$. The contribution made by the latter is

$$\mathrm{i}\rho\int_0^{\frac{1}{6}\pi}\mathrm{d}\phi\,\mathrm{e}^{\mathrm{i}\phi}\int_0^{\rho^2}\mathrm{d}s\operatorname{Ai}(s)\ln\{2\pi^{\frac{1}{2}}\mathrm{e}^{\frac{1}{6}\pi \mathrm{i}}\operatorname{Ai}(s)\}\exp[-\tfrac{1}{3}\rho^3\mathrm{e}^{3\mathrm{i}\phi} + \rho s\,\mathrm{e}^{-\mathrm{i}(\frac{2}{3}\pi-\phi)}].$$

The exponent is negative for the range of ϕ involved, and the integral tends to zero in the limit $\rho \to \infty$, leaving the contribution from the ray alone:

$$I_2 = \lim_{\rho\to\infty}\int_0^{\rho^2}\mathrm{d}s\left[\mathrm{i}\omega \operatorname{Ai}(s)\ln\{2\pi^{\frac{1}{2}}\mathrm{e}^{\frac{1}{6}\pi \mathrm{i}}\operatorname{Ai}(s)\}\int_0^\rho \mathrm{e}^{-\mathrm{i}(\frac{1}{3}\eta^3+\eta s)}\mathrm{d}\eta\right]. \tag{87}$$

The evaluation of I_3 presents greater difficulty. If we divide the integrals concerned into ray and arc contributions as above, we discover that the integral, W, defined by

$$\begin{aligned} W \equiv \int_0^\rho \mathrm{d}\eta \Bigg[\int_0^{\rho^2} & \mathrm{e}^{\frac{1}{3}\eta^3-\eta s}\,\omega \operatorname{Ai}(\omega s)\ln\{2\pi^{\frac{1}{2}}\mathrm{e}^{\frac{1}{6}\pi \mathrm{i}}\operatorname{Ai}(\omega s)\}\,\mathrm{d}s \\ & - \int_0^{\rho^2}\mathrm{e}^{\frac{1}{3}\eta^3-\eta\omega^2 s}\operatorname{Ai}(s)\ln\{2\pi^{\frac{1}{2}}\mathrm{e}^{\frac{1}{6}\pi \mathrm{i}}\operatorname{Ai}(s)\}\,\mathrm{d}s\Bigg] \end{aligned} \tag{88}$$

may be rewritten as

$$W = V + \int_0^{\rho^3} \mathrm{d}s \left[\mathrm{i}\omega \operatorname{Ai}(\omega s) \ln \{2\pi^{\frac{1}{2}} \mathrm{e}^{\frac{1}{6}\pi \mathrm{i}} \operatorname{Ai}(\omega s)\} \int_0^{\rho} \mathrm{e}^{-\mathrm{i}(\frac{1}{3}\eta^3 + \eta s)} \mathrm{d}\eta \right.$$
$$\left. + \mathrm{i}\omega \operatorname{Ai}(s) \ln \{2\pi^{\frac{1}{2}} \mathrm{e}^{\frac{1}{6}\pi \mathrm{i}} \operatorname{Ai}(s)\} \int_0^{\rho} \mathrm{e}^{\mathrm{i}(\frac{1}{3}\eta^3 + \eta s)} \mathrm{d}\eta \right], \tag{89}$$

where V represents the contribution of the arc integrals:

$$V = -\mathrm{i}\rho \int_0^{\frac{1}{2}\pi} \mathrm{e}^{\mathrm{i}\phi} \mathrm{d}\phi \int_0^{\rho^3} \omega \operatorname{Ai}(\omega s) \ln \{2\pi^{\frac{1}{2}} \mathrm{e}^{\frac{1}{6}\pi \mathrm{i}} \operatorname{Ai}(\omega s)\} \exp [\tfrac{1}{3}\rho^3 \mathrm{e}^{3\mathrm{i}\phi} - \rho s \mathrm{e}^{\mathrm{i}\phi}] \mathrm{d}s$$
$$+ \mathrm{i}\rho \int_0^{\frac{1}{6}\pi} \mathrm{e}^{\mathrm{i}\phi} \mathrm{d}\phi \int_0^{\rho^3} \operatorname{Ai}(s) \ln \{2\pi^{\frac{1}{2}} \mathrm{e}^{\frac{1}{6}\pi \mathrm{i}} \operatorname{Ai}(s)\} \exp [\tfrac{1}{3}\rho^3 \mathrm{e}^{3\mathrm{i}\phi} - \rho s \mathrm{e}^{-\mathrm{i}(\frac{2}{3}\pi - \phi)}] \mathrm{d}s$$
$$= V_1 + V_2, \text{ say.} \tag{90}$$

The first integral of (90) has a saddle point at $s = \rho^2 \mathrm{e}^{2\mathrm{i}\phi}$; the second at $s = \omega\rho^2 \mathrm{e}^{2\mathrm{i}\phi}$. Along the ray $s = s' \mathrm{e}^{2\mathrm{i}\phi}$, the integrand of V_1 is asymptotically proportional to

$$\exp [(\tfrac{1}{3}\rho^3 - \rho s' + \tfrac{2}{3}s'^{\frac{3}{2}}) \mathrm{e}^{3\mathrm{i}\phi}]. \tag{91}$$

If $\frac{1}{6}\pi < \phi < \frac{1}{2}\pi$, the real part of this exponent is negative everywhere on the ray $s' > 0$, except at the point $s' = \rho^2$ where it is zero. The integral is, in fact, asymptotically equal to the contribution from the line of steepest descent over this saddle point. If $0 < \phi < \frac{1}{6}\pi$, the exponent (91) has a positive real part everywhere on the ray except at the saddle point. In this case, however, the integral along the ray to the saddle point cancels exactly with a like ray integral arising from V_2. Each integral is therefore reduced to the contributions from the lines of steepest descent from these saddle points. Their sum, it appears, is equal and opposite to the contribution to V_1 from $\frac{1}{6}\pi < \phi < \frac{1}{2}\pi$ discussed above. Thus, V is asymptotically zero, and by (89)

$$I_3 = \lim_{\rho \to \infty} \int_0^{\rho^3} \mathrm{d}s \left[\left\{ \mathrm{i}\omega \operatorname{Ai}(\omega s) \ln \{2\pi^{\frac{1}{2}} \mathrm{e}^{\frac{1}{6}\pi \mathrm{i}} \operatorname{Ai}(\omega s)\} \int_0^{\rho} \mathrm{e}^{-\mathrm{i}(\frac{1}{3}\eta^3 + \eta s)} \mathrm{d}\eta \right.\right.$$
$$- \mathrm{i}\omega^2 \operatorname{Ai}(\omega^2 s) \ln \{2\pi^{\frac{1}{2}} \mathrm{e}^{-\frac{1}{6}\pi \mathrm{i}} \operatorname{Ai}(\omega^2 s)\} \int_0^{\rho} \mathrm{e}^{\mathrm{i}(\frac{1}{3}\eta^3 + \eta s)} \mathrm{d}\eta$$
$$+ \mathrm{i}\omega \operatorname{Ai}(s) \ln \{2\pi^{\frac{1}{2}} \mathrm{e}^{\frac{1}{6}\pi \mathrm{i}} \operatorname{Ai}(s)\} \int_0^{\rho} \mathrm{e}^{\mathrm{i}(\frac{1}{3}\eta^3 + \eta s)} \mathrm{d}\eta$$
$$- \mathrm{i}\omega^2 \operatorname{Ai}(s) \ln \{2\pi^{\frac{1}{2}} \mathrm{e}^{-\frac{1}{6}\pi \mathrm{i}} \operatorname{Ai}(s)\} \int_0^{\rho} \mathrm{e}^{-\mathrm{i}(\frac{1}{3}\eta^3 + \eta s)} \mathrm{d}\eta$$
$$\left.\left. + \tfrac{1}{3}s - \frac{1}{8s^{\frac{1}{2}}} \ln s \right\} \right]. \tag{92}$$

By (79) to (82), (86), (87) and (92), we now obtain

$$Q_s = \frac{2^{\frac{4}{3}}\pi}{\alpha^{\frac{7}{3}}} \int_0^{\infty} \mathrm{d}s \left[\mathrm{i}\omega \operatorname{Ai}^2(\omega^2 s) \ln \{2\pi^{\frac{1}{2}} \mathrm{e}^{-\frac{1}{6}\pi \mathrm{i}} \operatorname{Ai}(\omega^2 s)\} - \mathrm{i}\omega^2 \operatorname{Ai}^2(\omega s) \ln \{2\pi^{\frac{1}{2}} \mathrm{e}^{\frac{1}{6}\pi \mathrm{i}} \operatorname{Ai}(\omega s)\} \right.$$
$$- \mathrm{i}\omega^2 \operatorname{Ai}^2(s) \ln \{2\pi^{\frac{1}{2}} \mathrm{e}^{-\frac{1}{6}\pi \mathrm{i}} \operatorname{Ai}(s)\} + \mathrm{i}\omega \operatorname{Ai}^2(s) \ln \{2\pi^{\frac{1}{2}} \mathrm{e}^{\frac{1}{6}\pi \mathrm{i}} \operatorname{Ai}(s)\}$$
$$\left. + \frac{1}{\pi} \left(\tfrac{1}{3}s - \frac{1}{8s^{\frac{1}{2}}} \ln s + \frac{1}{4s^{\frac{1}{2}}} \right) \right]. \tag{93}$$

By using the result

$$\frac{\mathrm{d}}{\mathrm{d}s}[s\,\mathrm{Ai}^2(s)-\mathrm{Ai}'^2(s)]=\mathrm{Ai}^2(s),$$

we may conveniently integrate (93) by parts, giving

$$Q_s=\frac{2^{\frac{4}{3}}\pi}{\alpha^{\frac{7}{3}}}\int_0^\infty \mathrm{d}s\left\{\mathrm{i}\omega^2\frac{\mathrm{Ai}'(\omega s)}{\mathrm{Ai}(\omega s)}[\omega s\,\mathrm{Ai}^2(\omega s)-\mathrm{Ai}'^2(\omega s)]-\mathrm{i}\omega\frac{\mathrm{Ai}'(\omega^2 s)}{\mathrm{Ai}(\omega^2 s)}[\omega^2 s\,\mathrm{Ai}^2(\omega^2 s)-\mathrm{Ai}'^2(\omega^2 s)]\right.$$
$$\left.+\sqrt{3}\frac{\mathrm{Ai}'(s)}{\mathrm{Ai}(s)}[s\,\mathrm{Ai}^2(s)-\mathrm{Ai}'^2(s)]-\frac{1}{\pi}\left(s-\frac{1}{2s^{\frac{1}{2}}}\right)\right\}, \tag{94}$$

or, alternatively,

$$Q_s=\frac{2^{\frac{4}{3}}}{\alpha^{\frac{7}{3}}}\int_0^\infty \mathrm{d}s\left\{\left[\pi\frac{\mathrm{Ai}(s)\,\mathrm{Ai}'(s)+\mathrm{Bi}(s)\,\mathrm{Bi}'(s)}{\mathrm{Ai}^2(s)+\mathrm{Bi}^2(s)}[s\,\mathrm{Ai}(s)\,\mathrm{Bi}(s)-\mathrm{Ai}'(s)\,\mathrm{Bi}'(s)]\right.\right.$$
$$+\frac{1}{2}\frac{[s\,\mathrm{Ai}^2(s)-\mathrm{Ai}'^2(s)]-[s\,\mathrm{Bi}^2(s)-\mathrm{Bi}'^2(s)]}{\mathrm{Ai}^2(s)+\mathrm{Bi}^2(s)}$$
$$\left.\left.+\pi\sqrt{3}\frac{\mathrm{Ai}'(s)}{\mathrm{Ai}(s)}[s\,\mathrm{Ai}^2(s)-\mathrm{Ai}'^2(s)]-s+\frac{1}{2s^{\frac{1}{2}}}\right)\right]. \tag{95}$$

There does not appear to be any obvious method by which (94) or (95) may be evaluated analytically. A numerical integration has given

$$Q_s\sim 1{\cdot}732\alpha^{-\frac{7}{3}}. \tag{96}$$

Summing Q_{ms}, Q_h and Q_s we see that, to order $\alpha^{-\frac{7}{3}}$,

$$Q=\frac{4}{3\alpha}-\frac{\pi}{\alpha^2}+\frac{1{\cdot}732}{\alpha^{\frac{7}{3}}}. \tag{97}$$

For $M=100$, the ratio of the third term to the second is 0·1497 while that of the second to the first is 0·04712.

The calculation of the potential difference, Φ, across the duct requires the evaluation of the integral

$$\int_0^1\left[Mv+\frac{\partial b}{\partial z}\right]_{z=0}\mathrm{d}y=\int_0^1\left[2\alpha X+\frac{\partial X}{\partial z}\right]_{z=0}\mathrm{d}y$$
$$=\frac{\pi}{4}-\frac{\ln\alpha}{6\alpha}-\frac{1}{2\alpha}[1+\tfrac{2}{3}\ln 2-2Z_1(0)], \tag{98}$$

with the aid of (33), (34) and (76). A numerical integration has given $Z_1(0)=-0{\cdot}1830$. In this way we obtain

$$\Phi=\frac{\pi}{\alpha}\Phi_0\left[1-\frac{2}{3\pi\alpha}\ln\alpha-\frac{1{\cdot}164}{\alpha}\right], \tag{99}$$

where Φ_0 is the potential difference in the limit $M\to 0$.

References

Gold, R. R. 1962 *J. Fluid Mech.* **13**, 505.
Hunt, J. C. R. & Stewartson, K. 1965 *J. Fluid Mech.* **23**, 563.
Shercliff, J. A. 1962 *J. Fluid Mech.* **13**, 513.
Watson, G. N. 1944 *A treatise on the theory of Bessel functions* (2nd Ed.). Cambridge University Press.

THE CRITICAL LAYER IN STRATIFIED SHEAR FLOW

P. BALDWIN AND P. H. ROBERTS

Summary. The study of linear stability of a layer of stratified fluid in horizontal shearing motion leads, in the absence of diffusive effects, to a second order differential equation, often called the Taylor–Goldstein equation. This equation possesses a singularity at any critical point, i.e. at any point at which the flow speed, U, is equal to the wave speed, c. If c is complex, a similar singularity arises at any point at which the analytic extension of U into the complex plane is equal to c. Assuming the stratification is thermal in origin, the introduction of a small viscosity and heat conductivity removes this singularity, but leads to a governing equation of sixth order, four solutions being of rapidly varying WKBJ form. The circumstances in which the remaining two solutions can be uniformly represented in the limit of small viscosity and conductivity by the solutions of the Taylor–Goldstein equation are examined in this paper.

1. *Introduction.* Incompressible fluid occupies the layer $y_1 < y < y_2$, where y is a coordinate against the direction of gravity, $\mathbf{g}$, and flows in the x-direction with velocity $U(y)$. The density $\rho(y)$ of the fluid nearly takes a constant value, ρ_0, everywhere, i.e. $\beta(y) \equiv -d\rho/\rho dy$ is small compared with all reciprocal lengths of interest. The Brunt frequency $\mathrm{N}(y) = \sqrt{(g\beta)}$ is real for stable stratification $\beta > 0$. Adopting the Boussinesq approximation, the linear stability of this equilibrium is governed by the equations

$$\left(\frac{\partial}{\partial t} + U\frac{\partial}{\partial x}\right)\mathbf{u} + U' u_y \mathbf{1}_x = -\nabla\varpi - \sigma\mathbf{1}_y + \nu\nabla^2\mathbf{u}, \tag{1.1}$$

$$\left(\frac{\partial}{\partial t} + U\frac{\partial}{\partial x}\right)\sigma - \mathrm{N}^2 u_y = \kappa\nabla^2\sigma, \tag{1.2}$$

$$\nabla.\mathbf{u} = 0. \tag{1.3}$$

Here $\mathbf{u}$ denotes the departure of the fluid velocity from $U\mathbf{1}_x$; $\rho\varpi$ denotes the perturbation in pressure; σ the buoyance force per unit mass; ν the kinematic viscosity; κ the thermal diffusivity; and $\mathbf{1}_s$ denotes the unit vector in the s direction.

Operating on (1.1) by $-\mathbf{1}_y.\mathrm{curl}^2$, and using (1.3), we obtain

$$\left(\frac{\partial}{\partial t} + U\frac{\partial}{\partial x} - \nu\nabla^2\right)\nabla^2 u_y = U''\frac{\partial u_y}{\partial x} - \nabla_{\mathrm{H}}{}^2\sigma, \tag{1.4}$$

where $\nabla_{\mathrm{H}} \equiv (\partial/\partial x, 0, \partial/\partial z)$ is the horizontal gradient operator. On eliminating σ between (1.2) and (1.4), we see that

$$\left(\frac{\partial}{\partial t} + U\frac{\partial}{\partial x} - \kappa\nabla^2\right)\left[\left(\frac{\partial}{\partial t} + U\frac{\partial}{\partial x} - \nu\nabla^2\right)\nabla^2 u_y - U''\frac{\partial u_y}{\partial x}\right] = \mathrm{N}^2\nabla_{\mathrm{H}}{}^2 u_y. \tag{1.5}$$

Seeking normal mode solutions of the form

$$u_y = \phi(y)\exp\left[i\gamma(x - ct) + iz\delta\right], \tag{1.6}$$

[MATHEMATIKA 17 (1970), 102–119]

where $\gamma(\neq 0)$, δ and c are constants, we obtain

$$\left[\mathrm{D}^2 - \alpha^2 - \frac{i\gamma}{\kappa}(U-c)\right]\left\{\left[\mathrm{D}^2 - \alpha^2 - \frac{i\gamma}{\nu}(U-c)\right](\mathrm{D}^2-\alpha^2)\phi + \frac{i\gamma U''}{\nu}\phi\right\} = \frac{\mathrm{N}^2\alpha^2}{\nu\kappa}\phi, \tag{1.7}$$

where $\mathrm{D} = d/dy$ and $\alpha^2 = \gamma^2 + \delta^2$. In non-dimensional form, we may rewrite (1.7) as

$$[(\mathrm{D}^2-\alpha^2) - i\gamma\mathrm{RP}(U-c)]\{(\mathrm{D}^2-\alpha^2)^2\phi - i\gamma\mathrm{R}[(U-c)(\mathrm{D}^2-\alpha^2)\phi - U''\phi]\} = \alpha^2\mathrm{PR}^2\,\mathrm{Ri}\,\mathrm{N}^2\phi, \tag{1.8}$$

where P, R and Ri denote the Prandtl, Reynolds and Richardson numbers

$$\mathrm{P} = \frac{\nu}{\kappa}, \quad \mathrm{R} = \frac{\mathscr{U}\mathscr{L}}{\nu}, \quad \mathrm{Ri} = \left(\frac{\mathscr{N}\mathscr{L}}{\mathscr{U}}\right)^2,$$

and $\mathscr{L}$, $\mathscr{U}$ and $\mathscr{N}$ are the scales of length, velocity and Brunt frequency used in the non-dimensionalization.

The Taylor–Goldstein equation is obtained from (1.8) by taking the formal limit $\mathrm{R} \to \infty$, giving

$$(U-c)^2(\mathrm{D}^2-\alpha^2)\phi = [(U-c)U'' - (\alpha/\gamma)^2\,\mathrm{Ri}\,\mathrm{N}^2]\phi. \tag{1.9}$$

It is of interest to know the domains of y over which solutions of this equation provide uniformly valid solutions of (1.8). It is evident that equation (1.9) is singular at each zero, $y = y_c$, of $U - c$. Assuming, as will generally be the case, that $U - c = U_c'(y - y_c)$ as $y \to y_c$, where $U_c' \neq 0$, equation (1.9) reduces, to leading order in that limit, to

$$(y-y_c)^2\,\mathrm{D}^2\phi = -\mathrm{J}\phi, \tag{1.10}$$

where

$$\mathrm{J} = \left(\frac{\alpha}{\gamma}\right)^2 \mathrm{Ri}\left(\frac{\mathrm{N}_c}{U_c'}\right)^2.$$

This shows that as $y \to y_c$

$$\phi = a_1(y-y_c)^{\varepsilon_1} + a_2(y-y_c)^{\varepsilon_2}, \tag{1.11}$$

where

$$\left.\begin{matrix}\varepsilon_1\\ \varepsilon_2\end{matrix}\right\} = \tfrac{1}{2} \pm \sqrt{(\tfrac{1}{4} - \mathrm{J})}, \tag{1.12}$$

and a_1 and a_2 are constants. For brevity and in analogy with a terminology prevalent in the theory of homogeneous shear flows, we shall refer to solutions of (1.8) which are asymptotic to solutions of (1.9) as "the inviscid solutions", ignoring the role of heat conduction.

Near $y = y_c$, equation (1.8) has no singularity, and to leading order in the limit $y \to y_c$ it reduces to

$$\left(\frac{1}{\mathrm{P}}\frac{d^2}{dz^2} - z\right)\left(\frac{d^2}{dz^2} - z\right)\frac{d^2\phi}{dz^2} = -\mathrm{J}\phi, \tag{1.13}$$

where, here and from now onwards, z is taken to be

$$z = (i\gamma\mathrm{R}U_c')^{\frac{1}{3}}(y-y_c), \tag{1.14}$$

and assumed to run over complex values. The WKBJ method shows that the asymptotic forms of the six solutions of (1.13) as $|z| \to \infty$ are

$$z^{\varepsilon_1},\ z^{\varepsilon_2},\ z^{-\frac{1}{4}}\exp(\tfrac{2}{3}z^{\frac{3}{2}}),\ z^{-\frac{1}{4}}\exp(-\tfrac{2}{3}z^{\frac{3}{2}}),\ z^{-\frac{9}{4}}\exp(\tfrac{2}{3}\mathrm{P}^{\frac{1}{2}}z^{\frac{3}{2}})\ \text{and}\ z^{-\frac{9}{4}}\exp(-\tfrac{2}{3}\mathrm{P}^{\frac{1}{2}}z^{\frac{3}{2}}). \tag{1.15}$$

The first two are clearly related to the $y \to y_c$ form (1.11) of the Taylor–Goldstein solution. For brevity, we will refer to the remaining four solutions as "the viscous solutions".

Although (for $|z| \to \infty$) solutions of (1.13) behave as (1.15) indicates in a neighbourhood of z, it is clearly quite possible in principle that a solution asymptotic to z^{ε_1} for one value of $\arg z$ might be asymptotic to $z^{-\frac{1}{4}}\exp(\frac{2}{3}z^{\frac{3}{2}})$ for another. If these values of $\arg z$ corresponded to the positive and negative y-axes (for example), physical nonsense might result. It is the object of this paper to clarify the relationships between (1.15) for different values of $\arg z$, and at the same time provide corroboration for the procedure of Booker and Bretherton (1967). The present approach is however more akin to that of Koppel (1964) and of Gage and Reid (1968), who however only reached firm conclusions in the special case $\mathrm{P} = 1$. It is also related to the numerical study of Hazel (1967).

2. *Solution of the critical layer equation; the saddle points.* To simplify the solution of

$$\left(\frac{1}{\mathrm{P}}\frac{d^2}{dz^2} - z\right)\left(\frac{d^2}{dz^2} - z\right)\frac{d^2\phi}{dz^2} + \mathrm{J}\phi = 0, \tag{2.1}$$

we introduce constants

$$\xi = \tfrac{1}{3}\left(1 - \frac{1}{\mathrm{P}}\right),\quad m = \tfrac{1}{3}(\tfrac{1}{4} - \mathrm{J})^{\frac{1}{2}}\ \text{and}\ k = 3(\tfrac{1}{2} - m), \tag{2.2}$$

which we initially suppose are all positive. The basic solutions of (2.1) may then be written as Laplace integrals (Koppel, 1964). Three are given by

$$U_i(z) = \int_{C_i} \exp(zs - \tfrac{1}{3}s^3)\frac{f(s)}{s^k}\,ds,\quad (i = 1, 2, 3) \tag{2.3}$$

where

$$f(s) = {}_1\mathrm{F}_1(\tfrac{1}{6} + m;\ 1 + 2m;\ \xi s^3). \tag{2.4}$$

The base of six solutions may be completed by reversing the sign of m. These will be denoted by $V_i(z)$. The contour C_1 starts from $\infty\exp(2\pi i/3)$ and terminates at $\infty\exp(4\pi i/3)$; C_2 starts from $\infty\exp(4\pi i/3)$ and terminates at $\infty\exp(2\pi i)$; and C_3 starts from $\infty\exp(0)$ and terminates at $\infty\exp(2\pi i/3)$; see Fig. 1. It is important to observe that, although, by the series expansion of (2.4), these six solutions are integral functions of z, so that

$$U_i(z) \equiv U_i\big(z\exp(2\pi n i)\big), \tag{2.5}$$

for integral n, some care is necessary over the s arguments. For example, if $C_2{}^*$ were defined to start at $\infty\exp(-2\pi i/3)$ and terminate at $\infty\exp(0)$ and if $C_3{}^*$ were defined to start at $\infty\exp(2\pi i)$ and terminate at $\infty\exp(8\pi i/3)$, we would find that

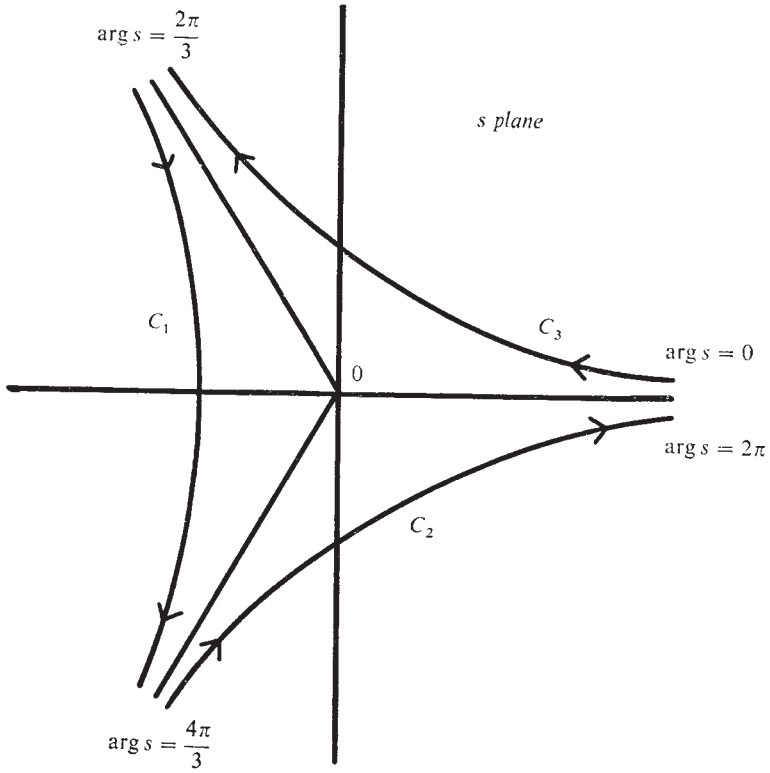


Fig. 1

the corresponding solutions (2.3) would be related by

$$\left.\begin{aligned} U_2^* &= A^{-3}U_2, \quad U_3^* = A^3 U_3, \\ V_2^* &= B^{-3}V_2, \quad V_3^* = B^3 V_3, \end{aligned}\right\} \tag{2.6}$$

where

$$A = \exp\left(-(\tfrac{1}{3} - 2m)\,\pi i\right), \text{ and } B = \exp\left(-(\tfrac{1}{3} + 2m)\,\pi i\right). \tag{2.7}$$

(Note $A^3 = B^{-3}$.) Another elementary property of the definitions (2.3) follows from rotations of the s-plane by $\pm 2\pi/3$:

$$U_2(z) \equiv AU_1(z\exp(2\pi i/3)), \text{ and } U_3(z) = A^{-1}U_1(z\exp(-2\pi i/3)). \tag{2.8}$$

We now evaluate U_i asymptotically for large $|z|$. Referring to Slater (1960, eq. 4.1.6), we find

$$f(s) \sim \frac{\Gamma(1+2m)\,A^n \exp(\xi s^3)}{\Gamma(\frac{1}{6}+m)\,\xi^{m+\frac{5}{6}}\,s^{3m+\frac{5}{2}}}\;{}_2\mathrm{F}_0\left(\tfrac{5}{6}+m, \tfrac{5}{6}-m;;\frac{1}{\xi s^3}\right)$$

$$+\frac{\Gamma(1+2m)\,B^{-(n-\frac{1}{2}\varepsilon)}}{\Gamma(\frac{5}{6}+m)\,\xi^{m+\frac{1}{6}}\,s^{3m+\frac{1}{2}}}\;{}_2\mathrm{F}_0\left(\tfrac{1}{6}+m, \tfrac{1}{6}-m;;-\frac{1}{\xi s^3}\right), \tag{2.9}$$

provided

$$\left.\begin{aligned} &\text{either } (2n-1)\frac{\pi}{3} < \arg s < \frac{2n\pi}{3}, \text{ and } \varepsilon = 1, \\ &\text{or } \frac{2n\pi}{3} < \arg s < (2n+1)\frac{\pi}{3}, \text{ and } \varepsilon = -1, \end{aligned}\right\} \tag{2.10}$$

where n is an integer. In fact, it is clear that the Stokes lines for this asymptotic expansion are the rays $\arg s = \pi n/3$, while the anti-Stokes lines are

$$\arg s = (2n+1)\,\pi/6.$$

On substituting into (2.3), we readily see that the second term of (2.9) gives rise to two saddle points, located asymptotically at

$$s = s_1 \equiv z^{\frac{1}{2}} \exp(2\pi l i), \text{ and } s = s_2 \equiv z^{\frac{1}{2}} \exp\left((2l+1)\,\pi i\right), \tag{2.11}$$

where l is an integer, while the first term of (2.9) is associated with saddle points at

$$s = s_{P1} \equiv (zP)^{\frac{1}{2}} \exp(2\pi l i), \text{ and } s = s_{P2} \equiv (zP)^{\frac{1}{2}} \exp\left((2l+1)\,\pi i\right). \tag{2.12}$$

These four "outer saddle points" as we shall term them, are related to the viscous solutions of (1.15). Inspection of (2.3) also reveals directly the presence of an "inner saddle point" at

$$s = s_0 \equiv \frac{k}{z} \exp(2\pi l i). \tag{2.13}$$

This is related to the inviscid solution z^{ε_2} in (1.15); the other inviscid solution arises from an inner saddle point for the V solutions. Naturally, the locations of the saddle points given in (2.11) to (2.13) are only correct asymptotically in the limit $|z| \to \infty$. The purpose of the integer l appearing in (2.11) to (2.13) is clear when it is recalled that, since $\arg s$ is unambiguously defined at infinity on each C_i, it is also determined unambiguously at any saddle point over which C_i (suitably distorted) passes. Later, in Table 1, we list the asymptotic forms of U_i for $0 < \arg z < 2\pi$. In deriving these, it was necessary to take $l = 1$ for s_0 and $l = 0$ for the remainder, apart from two major exceptions: (i) for U_2 in the sector $0 < \arg z < 2\pi/3$, $l = 1$ must be used in the definitions of s_1 and s_{P1}, and (ii) for U_3 in the sector $4\pi/3 < \arg z < 2\pi$, $l = -1$ must be used in the definitions of s_2 and s_{P2}.

We now evaluate, for each distinct saddle point in turn, the contribution to an integral of the form (2.3) from the neighbourhood of that saddle point when C_i is distorted to pass along the steepest line over it. We include only the leading term (namely, unity) in each ${}_2F_0$ function appearing in (2.9):

(a) *The saddle point*, s_1. Supposing z lies in the range

$$\frac{4\pi}{3}(n-1) < \arg z < \frac{4\pi n}{3}, \tag{2.14a, d}$$

integrate the second term of (2.9) over $s = s_1$ in the (steepest) direction

$$\arg(s - s_1) = -\tfrac{1}{4}\arg z \textit{ from } \arg(s - s_1) = \pi - \tfrac{1}{4}\arg z.$$

The contribution is

$$B^{-(3l+n-\frac{1}{2})}\left[\frac{\sqrt{\pi}\,\Gamma(1+2m)}{\xi^{m+\frac{1}{4}}\,\Gamma(\frac{5}{6}+m)}\right] z^{-\frac{5}{4}} \exp(\tfrac{4}{3}z^{\frac{3}{4}}) \equiv B^{-(3l+n-1)}H_1(z), \text{ say.} \tag{2.15a}$$

(b) *The saddle point*, s_2. Supposing z lies in the range

$$\frac{2\pi}{3}(2n-1) < \arg z < \frac{2\pi}{3}(2n+1), \tag{2.14b, c}$$

integrate the second term of (2.9) over $s = s_2$ in the (steepest) direction

$$\arg(s - s_2) = -\tfrac{1}{2}\pi - \tfrac{1}{4}\arg z \textit{ from } \arg(s - s_2) = \tfrac{1}{2}\pi - \tfrac{1}{4}\arg z.$$

The contribution made is

$$-iB^{-(3l+n+\frac{1}{2})}\left[\frac{\sqrt{\pi}\,\Gamma(1+2m)}{\xi^{m+\frac{1}{6}}\,\Gamma(\frac{5}{6}+m)}\right] z^{-\frac{5}{4}}\exp(-\tfrac{2}{3}z^{\frac{3}{2}}) \equiv B^{-(3l+n+1)}H_2(z), \quad \text{say.} \tag{2.15b}$$

(c) *The saddle point*, s_{P1}. Supposing z lies in the range (2.14b, c) above, integrate the first term of (2.9) over $s = s_{P1}$ in the (steepest) directions defined in (a). The contribution made is

$$A^{3l+n}\left[\frac{\sqrt{\pi}\,\Gamma(1+2m)}{\xi^{m+\frac{5}{6}}\,\mathrm{P}_{\frac{1}{6}}^{\frac{1}{2}}\,\Gamma(\frac{1}{6}+m)}\right] z^{-\frac{9}{4}}\exp(\tfrac{2}{3}\mathrm{P}^{\frac{1}{2}}z^{\frac{3}{2}}) \equiv A^{3l+n}H_{P1}(z), \text{ say.} \tag{2.15c}$$

(d) *The saddle point*, s_{P2}. Supposing z lies in the range (2.14a, d) above, integrate the first term of (2.9) over $s = s_{P2}$ in the (steepest) directions defined in (b). The contribution made is

$$-iA^{3l+n+1}\left[\frac{\sqrt{\pi}\,\Gamma(1+2m)}{\xi^{m+\frac{5}{6}}\,\mathrm{P}_{\frac{1}{6}}^{\frac{1}{2}}\,\Gamma(\frac{1}{6}+m)}\right] z^{-\frac{9}{4}}\exp(-\tfrac{2}{3}\mathrm{P}^{\frac{1}{2}}z^{\frac{3}{2}}) \equiv A^{3l+n+1}H_{P2}(z), \text{ say.} \tag{2.15d}$$

It is worth noticing that, according to the definitions,

$$\left.\begin{aligned} H_1(z) &\equiv \pm(AB)^{\pm 1}H_2\big(z\exp(\pm 2\pi i/3)\big),\\ H_2(z) &\equiv \mp(AB)^{\pm 1}H_1\big(z\exp(\pm 2\pi i/3)\big),\\ H_{P1}(z) &\equiv \pm H_{P2}\big(z\exp(\pm 2\pi i/3)\big),\\ H_{P2}(z) &\equiv \mp H_{P1}\big(z\exp(\pm 2\pi i/3)\big). \end{aligned}\right\} \tag{2.16}$$

(e) *The saddle point*, s_0. Although the level curves following the lines of steepest descent from s_0 are perpendicular to the ray $\arg s = \arg s_0$ at s_0, they are both asymptotically anti-parallel to it when $|s_0| \ll |s| \ll |z|^{\frac{1}{2}}$, after the manner sketched in Fig. 2 for the case $\arg z = \pi/4$. Of course, the contour there illustrated cannot necessarily be continued in these directions once $|s| \gg |z|^{\frac{1}{2}}$, since the convergence requirements of the $\exp(-s^3/3)$ term in (2.3) might then be violated. Nevertheless, if any segment of C_i has to be brought into the $|s| \ll |z|^{\frac{1}{2}}$ neighbourhood of the origin and over s_0, the corresponding contributions to the integral U_i may be obtained, to leading order, by ignoring the $\exp(-s^3/3)$ term and extending this segment to infinity along the rays $\arg s = (2l \pm 1)\pi - \arg z$. Define, therefore, a contour $\mathscr{C}_s$ consisting of a path from ∞ on $\arg s = (2l-1)\pi - \arg z$ which passes round the origin in the positive sense and returns to ∞ on $\arg s = (2l+1)\pi - \arg z$. The leading order contribution to an integral, such as (2.3), from this segment is then found to be

$$\int_{\mathscr{C}_s} \exp(zs)\frac{ds}{s^k},$$

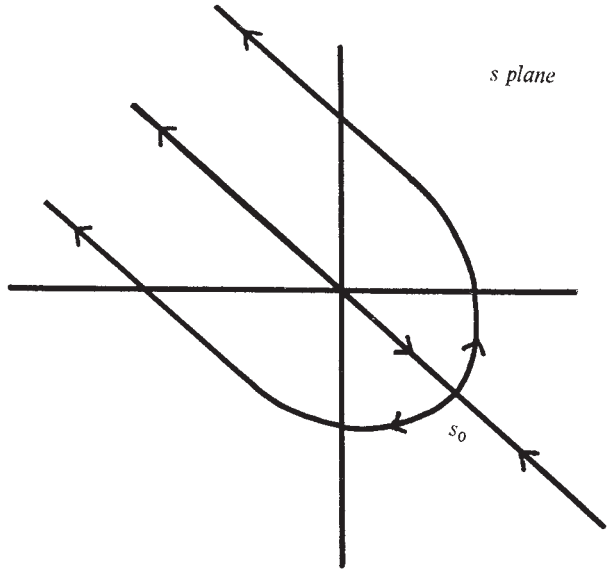


Fig. 2. Level curve, near s_0 for arg $z = \frac{1}{4}\pi$

where we have replaced the ${}_1F_1$ function in (2.4) by its leading term, viz. unity. To present this integral in a more familiar guise, we write $s = -(t/z)\exp(2\pi li)$, and adopt the convention that $\log(-t)$ is purely real when t is on the negative real t-axis. The contour $\mathscr{C}_s$ transforms to a contour, $\mathscr{C}_t$, which starts and finishes at ∞ on the positive t-axis and encircles $t = 0$ counterclockwise. We find the integral becomes

$$-A^{3l} z^{\frac{1}{4}-3m} \int_{\mathscr{C}_t} (-t)^{-k} \exp(-t)\, dt.$$

Referring now to Hankel's representation of the gamma function (*cf.* e.g. Whittaker and Watson, 1927, §12.22), we reduce the contribution to

$$A^{3l} \frac{2\pi i}{\Gamma(k)} z^{\frac{1}{4}-3m} \equiv A^{3l} L(z), \quad \text{say.} \tag{2.15e}$$

3. *Behaviour of solutions in different Stokes sectors.* As arg z varies, the location of the saddle points s_i varies and the geometry of the steepest lines through them changes continuously. The situation is illustrated in the sequence of six cases sketched in Figs. 3 (a) to (f) for arg $z = 0$ to $5\pi/3$ in steps of $\pi/3$. (For simplicity, the s_P saddle points are excluded.) The obvious relationship between cases (a), (c) and (e) and between cases (b), (d) and (f) brings home once more the fundamental $2\pi/3$ periodicity of (2.3) and (2.8). Cases (a), (c) and (e) are atypical in the sense that (as we will later see) each marks a transition between two topologically different distortions of C_i over the saddle points, i.e. the corresponding arguments of z, viz. 0, $4\pi/3$ and $2\pi/3$ define the Stokes lines, here labelled D_1, D_2 and D_3. (See Fig. 4 (a).)

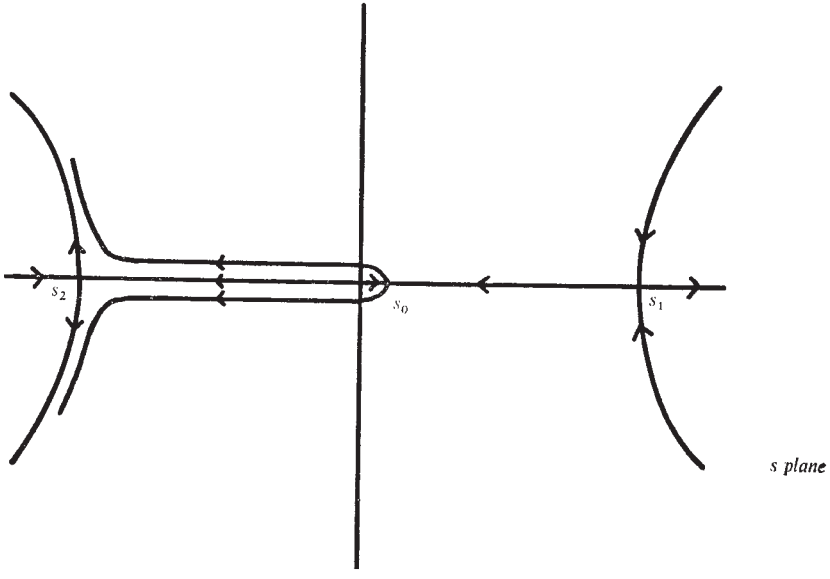


Fig. 3(a). arg $z = 0$

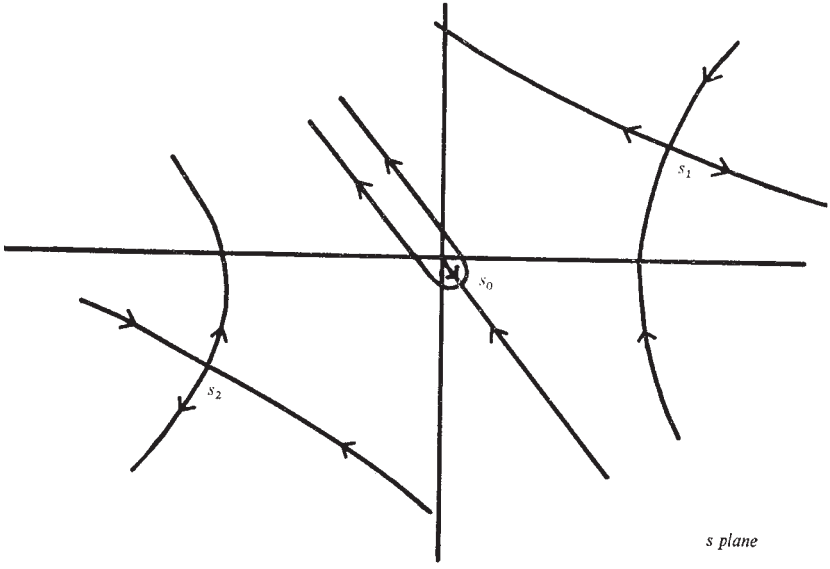


Fig. 3(b). arg $z = \pi/3$

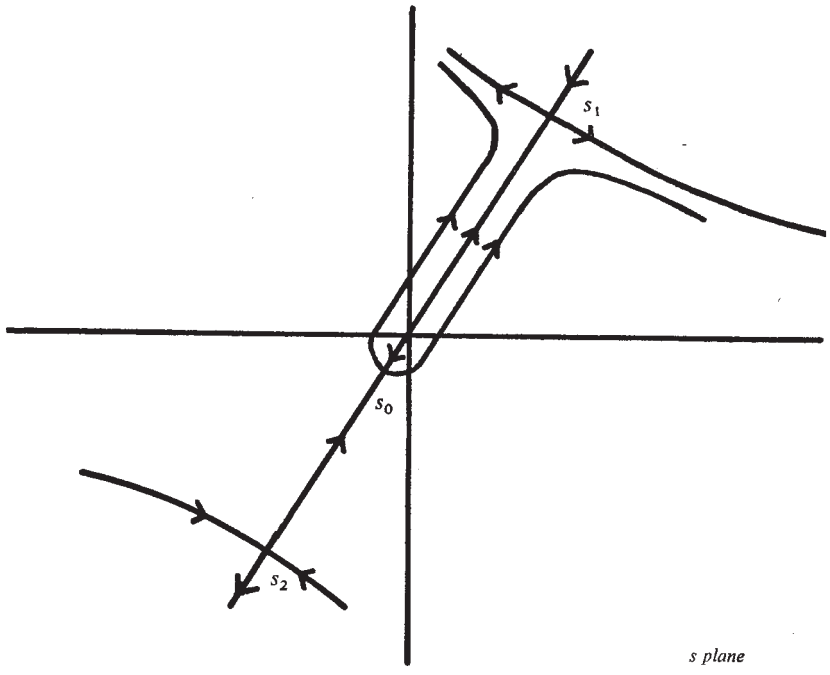


Fig. 3(c). arg $z = 2\pi/3$

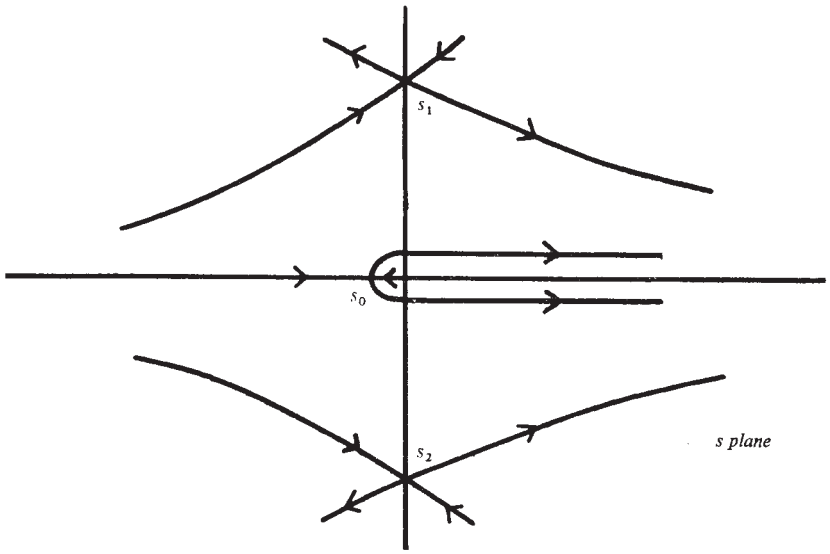


Fig. 3(d). arg $z = \pi$

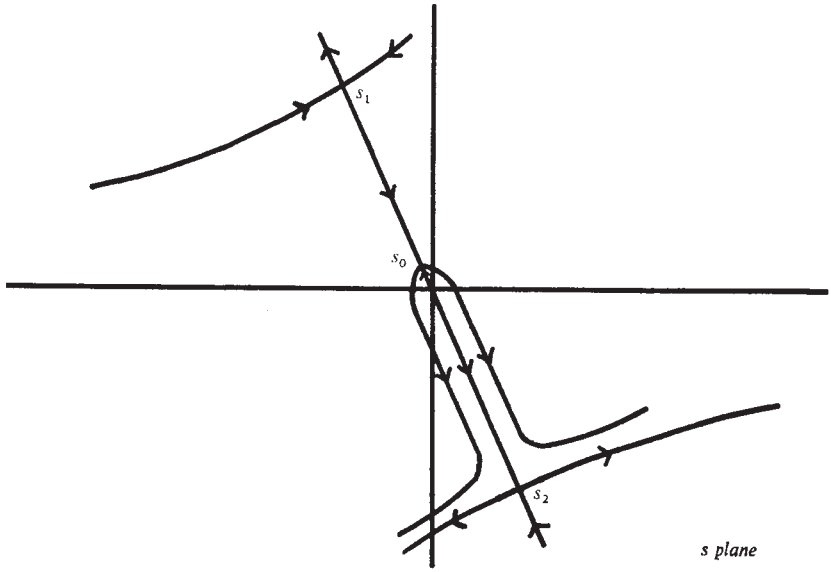


Fig. 3(e). arg $z = 4\pi/3$

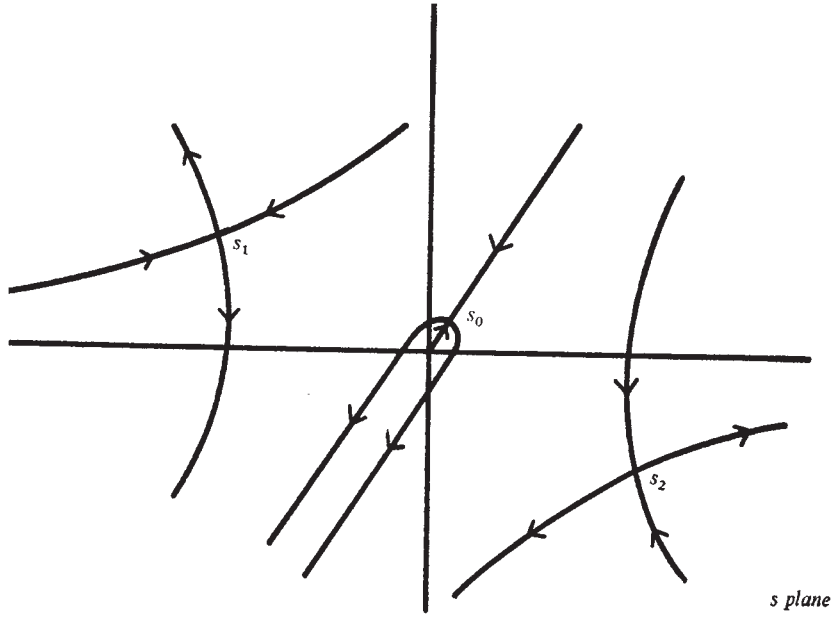


Fig. 3(f). arg $z = 5\pi/3$

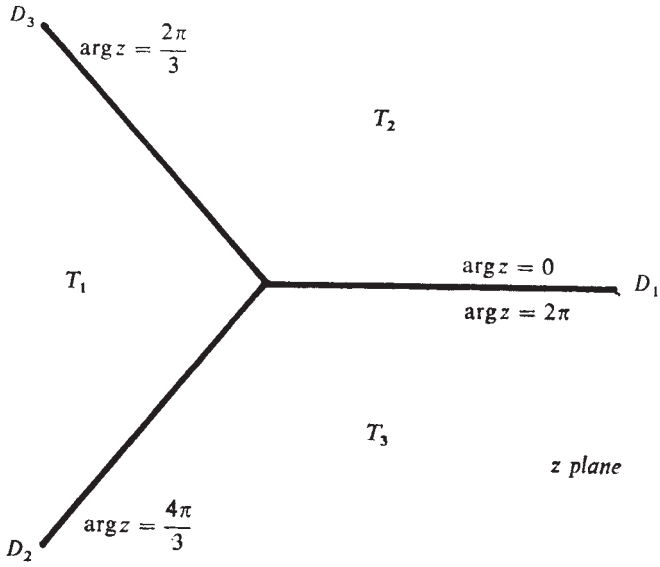


Fig. 4(a). Stokes lines and sectors

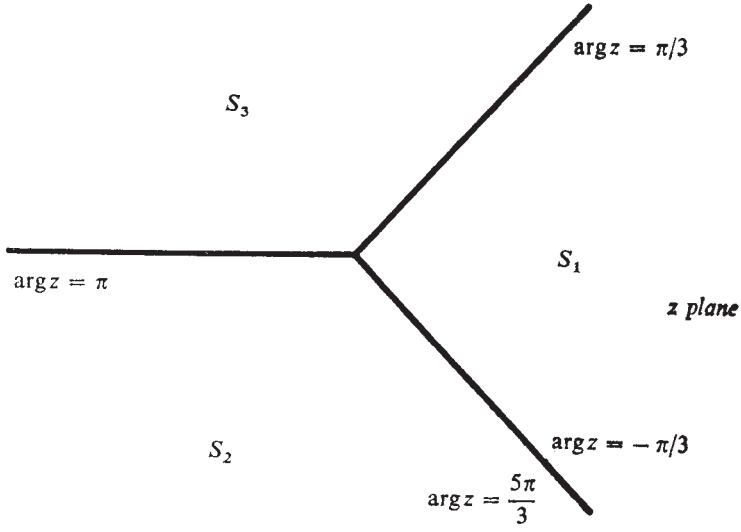


Fig. 4(b). Anti-Stokes lines and sectors

We denote the sector opposite D_i by T_i:

$$T_1 : 2\pi/3 < \arg z < 4\pi/3,$$
$$T_2 : \quad 0 < \arg z < 2\pi/3,$$
$$T_3 : 4\pi/3 < \arg z < 2\pi.$$

It should be noted that it would not be correct to use the tables below if the arguments listed here are altered by taking, for example, T_3 to be $-2\pi/3 < \arg z < 0$ (although this case could, of course, be covered from the one listed by replacing z by $z \exp(2\pi i)$).

We now consider the form of C_1 in each sector T_i. In T_2, C_1 passes on the left of the origin from $\infty \exp(2\pi i/3)$ to $\infty \exp(4\pi i/3)$ and over the steepest line at s_{P2} or s_2, depending on whether the first or second term of (2.9) is under discussion. Thus the asymptotic expansion of U_1 contains only the H_{P2} ($n = 1, l = 0$) and H_2 ($n = l = 0$) contributions of (b) and (d) above, both with the positive sign since the direction over the saddle points is the same as that listed when defining these terms. For $0 < \arg z < \pi/3$ both terms are exponentially large, the $B^{-1}H_2$ dominating the $A^2 H_{P2}$ (for $P > 1$). This relative dominancy is reversed on crossing into the sector $\pi/3 < \arg z < 2\pi/3$ where each term is exponentially small. The line $\arg z = \pi/3$ is therefore recognised as an anti-Stokes line; two others are $\arg z = \pi$ and $\arg z = 5\pi/3$. These cases are illustrated in Figs. 3 (b), (d) and (f). If we pass from sector T_2 into sector T_3, the topology of C_1 is changed; it now passes over either s_{P1} or s_1, and the asymptotic expansion therefore involves only the H_{P1} ($n = 1$, $l = 0$) and H_1 ($n = 2, l = 0$) contributions of (a) and (c) above, both again in a positive sense. For $5\pi/6 < \arg z < 2\pi$, the $B^{-1}H_1$ dominates the AH_{P1} (for $P > 1$), both being exponentially large. As a check we notice that the dominating terms on each side of D_1 namely $B^{-1}H_1$ and $B^{-1}H_2$ are numerically equal; the subdominant parts, AH_{P1} and $A^2 H_{P2}$ are discontinuous, as one expects at a Stokes line. On moving into T_1 the contour C_1 becomes highly distorted and passes over all five saddle points, each in the positive sense. First, from $\infty \exp(2\pi i/3)$ it crosses over either s_{P1} or s_1 to $\infty \exp(0)$, giving rise to H_{P1} ($n = 1, l = 0$) and H_1 ($n = 1, l = 0$) contributions to U_1. Next, it passes from $\infty \exp(0)$ to $\infty \exp(2\pi i)$ via a $\mathscr{C}_s$ contour over s_0 and produces an L ($l = 1$) term as evaluated in (e) above. Finally, it passes from $\infty \exp(2\pi i)$ to $\infty \exp(4\pi i/3)$ over either s_{P2} or s_2 and gives rise to H_{P2} ($n = 1$, $l = 0$) and H_2 ($n = 1, l = 0$) contributions to U_1. Similar expansions can be obtained for U_2 and U_3. The final results are listed in Table 1 in which, for $0 < \arg z < \pi/3$ in T_2, for $2\pi/3 < \arg z < \pi$ in T_1, and for $4\pi/3 < \arg z < 5\pi/3$ in T_3, the order shown is that of decreasing dominancy (largest first). For the remainder of these sectors, on the other sides of the respective anti-Stokes lines, the order is that of increasing dominancy (largest last). Table 1 also lists solutions, u_i, to be defined in §4.

At the commencement of §2 it was supposed, for simplicity, that ξ, m and k were positive, i.e. that $P > 1$ and $-2 < J < \frac{1}{4}$. Some of these restrictions may now be lifted: the analysis holds almost as it stands when $J > \frac{1}{4}$, for the complex k then arising merely changes $\arg s_0$ by an amount numerically less than $\pi/2$. Although this results in a displacement of the level lines in the neighbourhood of $s = 0$, they are essentially undistorted for $|s| \gg 1$, and the topology of their connection is as before. The analysis therefore goes through unchanged, apart from the replacement of m by $i\mu$, where $\mu = [\sqrt{(4J - 1)}]/6$.

TABLE 1 (P > 1)

Sector $T_1 : 2\pi/3 < \arg z < 4\pi/3$

$$U_1 \sim A^2 H_{P2} + B^{-2} H_2 + A^3 L + H_1 + A H_{P1},$$
$$U_2 \sim -A^2 H_{P2} - B^{-2} H_2,$$
$$U_3 \sim -H_1 - A H_{P1},$$
$$u_1 \sim A^3 L,$$
$$u_2 \sim A^3 L + (1 - A^3) H_1 + A(1 - A^3) H_{P1},$$
$$u_3 \sim A^2(1 - A^{-3}) H_{P2} + B^{-2}(1 - A^{-3}) H_2 + A^3 L,$$

Sector $T_2 : 0 < \arg z < 2\pi/3$

$$U_1 \sim B^{-1} H_2 + A^2 H_{P2},$$
$$U_2 \sim A^3 H_{P1} + B^{-3} H_1 + A^3 L - B^{-1} H_2 - A^2 H_{P2},$$
$$U_3 \sim -H_{P1} - H_1,$$
$$u_1 \sim -(1 - A^3) H_{P1} - (1 - A^3) H_1 + A^3 L,$$
$$u_2 \sim A^3 L,$$
$$u_3 \sim L + B^{-1}(1 - A^{-3}) H_2 + A^2(1 - A^{-3}) H_{P2},$$

Sector $T_3 : 4\pi/3 < \arg z < 2\pi$

$$U_1 \sim A H_{P1} + B^{-1} H_1,$$
$$U_2 \sim -B^{-2} H_2 - A^3 H_{P2},$$
$$U_3 \sim -A H_{P1} - B^{-1} H_1 + A^3 L + B H_2 + H_{P2},$$
$$u_1 \sim A^3 L + B(1 - A^3) H_2 + (1 - A^3) H_{P2},$$
$$u_2 \sim A(1 - A^3) H_{P1} + B^{-1}(1 - A^3) H_1 + A^6 L,$$
$$u_3 \sim A^3 L,$$

where

$$H_1 = h z^{-\frac{1}{4}} \exp(\tfrac{2}{3} z^{\frac{3}{2}}),\ H_2 = -ih z^{-\frac{1}{4}} \exp(-\tfrac{2}{3} z^{\frac{3}{2}}) \equiv H_1(z \exp(2\pi i)),$$

$$H_{P1} = h_P z^{-\frac{9}{4}} \exp(\tfrac{2}{3} P^{\frac{1}{2}} z^{\frac{3}{2}}),\ H_{P2} = -ih_P z^{-\frac{9}{4}} \exp(-\tfrac{2}{3} P^{\frac{1}{2}} z^{\frac{3}{2}}) \equiv H_{P1}(z \exp(2\pi i)),$$

$$h = \frac{\pi^{\frac{1}{2}} \Gamma(1 + 2m)\, B^{-\frac{1}{4}}}{\xi^{m+\frac{1}{4}} \Gamma(\frac{5}{6} + m)},\quad h_P = \frac{\pi^{\frac{1}{2}} \Gamma(1 + 2m)}{\xi^{m+\frac{1}{4}} P^{\frac{1}{2}} \Gamma(\frac{1}{6} + m)},\quad L = \frac{2\pi i}{\Gamma(k)} z^{\frac{1}{4} - 3m}.$$

Asymptotic forms for the solutions V_i and v_i may be obtained by replacing m by $-m$ (which also involves exchanging A and B).

Considering next the restriction $\xi > 0$, we note that the case $P = 1$ has already been completely solved by Koppel (1964) and, very elegantly, by Gage and Reid (1967). The case $0 < P < 1$ requires a slight reformulation, since, for $\arg s$ in the range (2.10), relation (2.9) must be replaced by

$$f(s) \sim \frac{\Gamma(1+2m)\,B^{-n}}{\Gamma(\frac{5}{6}+m)|\xi|^{m+\frac{1}{2}}\,s^{3m+\frac{1}{2}}}\,{}_2F_0\left(\tfrac{1}{6}+m,\ \tfrac{1}{6}-m;;\ \frac{1}{|\xi|s^3}\right)$$

$$-\frac{\Gamma(1+2m)\,A^{n-\frac{1}{2}\varepsilon}\exp(\xi s^3)}{\Gamma(\frac{1}{6}+m)|\xi|^{m+\frac{5}{6}}\,s^{3m+\frac{1}{2}}}\,{}_2F_0\left(\tfrac{5}{6}+m,\ \tfrac{5}{6}-m;;-\frac{1}{|\xi|s^3}\right). \quad (3.1)$$

The remaining analysis is much as before, and leads to the results listed in Table 2.

As in the case $J = \frac{1}{4}$ (i.e. $m = 0$), negative values of J may lead to integral values of $2m$. In these cases, the solutions U_i and V_i are not independent. This difficulty is trivial, and may be overcome by a familiar limiting argument. If $J = -N(N+1)$ where N is a positive integer, k is zero or a negative integer, and the branch point at $s = 0$ in (2.3) disappears, to be replaced by a zero, i.e. the three U_i solutions are no longer independent (their sum vanishes). A new contour can, however, be chosen from $s = 0$ to ∞ and provides a new independent solution. For any $J \leqslant -2$, the lines of steepest descent and ascent through s_0 are exchanged, and $\arg s_0$ changes by π; the level lines of Figs. 3 are substantially modified. In view of all these complications, it is perhaps fortunate that little physical interest attaches to negative J, since the instabilities in these cases of unstable stratification should be rolls in the x-direction ($\gamma = 0$) completely unaffected by the shear U. For example, Gage and Reid (*loc. cit*) observed that (when $P = 1$) such modes were preferred for the case of plane Poiseuille flow of a uniformly stratified fluid, provided $J < -0{\cdot}92\,10^{-6}$, a numerically tiny value. In what follows, we will ignore the possibility of negative J.

4. *Application to gravity waves.* In stable situations, c is real and lies in the range of U. There will, therefore, be one or more critical points in the real interval $y_1 \leqslant y \leqslant y_2$ under consideration. Far from these critical points, ϕ is given by the Taylor–Goldstein equation (1.9) and appropriate boundary conditions at $y = y_1$ and $y = y_2$. Near the critical points, however, these solutions become non-analytic (according to (1.11), $D\phi$ becomes infinite), and the question arises as to how the solutions on each side may be properly connected. Booker and Bretherton (1967) overcame this difficulty by considering the associated initial value problem. They concluded that, for $\gamma U_c' > 0$, the components of (1.11) behaving as $(y - y_c)^{\frac{1}{2}\pm 3m}$ for $y > y_c$ should, for $y < y_c$, be matched to the solution behaving as

$$Q(\mp m)(y_c - y)^{\frac{1}{2}\pm 3m}, \quad (4.1)$$

where

$$Q(m) = -i\exp(3\pi m i). \quad (4.2)$$

This result is one that could be obtained formally by replacing $y - y_c$ by $\exp(-\pi i)(y_c - y)$ for points to the left of y_c. Hazel (1967) integrated (2.1) numerically for several values of P and J, and confirmed Booker and Bretherton's findings. We now wish to demonstrate that the conclusion is in fact true for all P and $J(> 0)$.

Although the result may be argued directly from the solutions already discussed,

TABLE 2 (P < 1)

Sector $T_1 : 2\pi/3 < \arg z < 4\pi/3$

$$U_1 \sim B^{-2}\bar{H}_2 - A^2\bar{H}_{P2} + A^3 L - \bar{H}_{P1} + B^{-1}\bar{H}_1,$$
$$U_2 \sim -B^{-2}\bar{H}_2 + A^2\bar{H}_{P2},$$
$$U_3 \sim \bar{H}_{P1} - B^{-1}\bar{H}_1,$$
$$u_1 \sim A^3 L,$$
$$u_2 \sim A^3 L - (1 - A^3)\bar{H}_{P1} + B^{-1}(1 - A^3)\bar{H}_1,$$
$$u_3 \sim B^{-2}(1 - A^{-3})\bar{H}_2 - A^2(1 - A^{-3})\bar{H}_{P2} + A^3 L,$$

Sector $T_2 : 0 < \arg z < 2\pi/3$

$$U_1 \sim -A\bar{H}_{P2} + B^{-2}\bar{H}_2,$$
$$U_2 \sim B^{-3}\bar{H}_1 - A^3\bar{H}_{P1} + A^3 L + A\bar{H}_{P2} - B^{-2}\bar{H}_2,$$
$$U_3 \sim -\bar{H}_1 + \bar{H}_{P1},$$
$$u_1 \sim -(1 - A^3)\bar{H}_1 + (1 - A^3)\bar{H}_{P1} + A^3 L,$$
$$u_2 \sim A^3 L,$$
$$u_3 \sim L - A(1 - A^{-3})\bar{H}_{P2} + B^{-2}(1 - A^{-3})\bar{H}_2,$$

Sector $T_3 : 4\pi/3 < \arg z < 2\pi$

$$U_1 \sim B^{-1}\bar{H}_1 - A\bar{H}_{P1},$$
$$U_2 \sim A^2\bar{H}_{P2} - B^{-3}\bar{H}_2,$$
$$U_3 \sim -B^{-1}\bar{H}_1 + A\bar{H}_{P1} + A^3 L - A^{-1}\bar{H}_{P2} + \bar{H}_2,$$
$$u_1 \sim A^3 L - A^{-1}(1 - A^3)\bar{H}_{P2} + (1 - A^3)\bar{H}_2,$$
$$u_2 \sim B^{-1}(1 - A^3)\bar{H}_1 - A(1 - A^3)\bar{H}_{P1} + A^6 L,$$
$$u_3 \sim A^3 L,$$

where

$$\bar{H}_1 = \bar{h} z^{-\frac{1}{6}} \exp(\tfrac{2}{3}z^{\frac{3}{2}}), \quad \bar{H}_2 = -i\bar{h} z^{-\frac{1}{6}} \exp(-\tfrac{2}{3}z^{\frac{3}{2}}) \equiv \bar{H}_1(z \exp(2\pi i)),$$
$$\bar{H}_{P1} = \bar{h}_P z^{-\frac{1}{6}} \exp(\tfrac{2}{3}P^{\frac{1}{2}} z^{\frac{3}{2}}), \quad \bar{H}_{P2} = -i\bar{h}_P z^{-\frac{1}{6}} \exp(-\tfrac{2}{3}P^{\frac{1}{2}} z^{\frac{3}{2}}) \equiv \bar{H}_{P1}(z \exp(2\pi i)),$$

$$\bar{h} = \frac{\pi^{\frac{1}{2}}\,\Gamma(1 + 2m)}{|\xi|^{m+\frac{1}{6}}\,\Gamma(\frac{5}{6} + m)}, \quad \bar{h}_P = \frac{\pi^{\frac{1}{2}}\,\Gamma(1 + 2m)\,A^{\frac{1}{2}}}{|\xi|^{m+\frac{1}{6}}\,P^{\frac{1}{2}}\,\Gamma(\frac{1}{6} + m)}, \quad L = \frac{2\pi i}{\Gamma(k)} z^{\frac{1}{2}-3m}.$$

Asymptotic forms for the solutions V_i and v_i may be obtained by replacing m by $-$, (which also involves exchanging A and B).

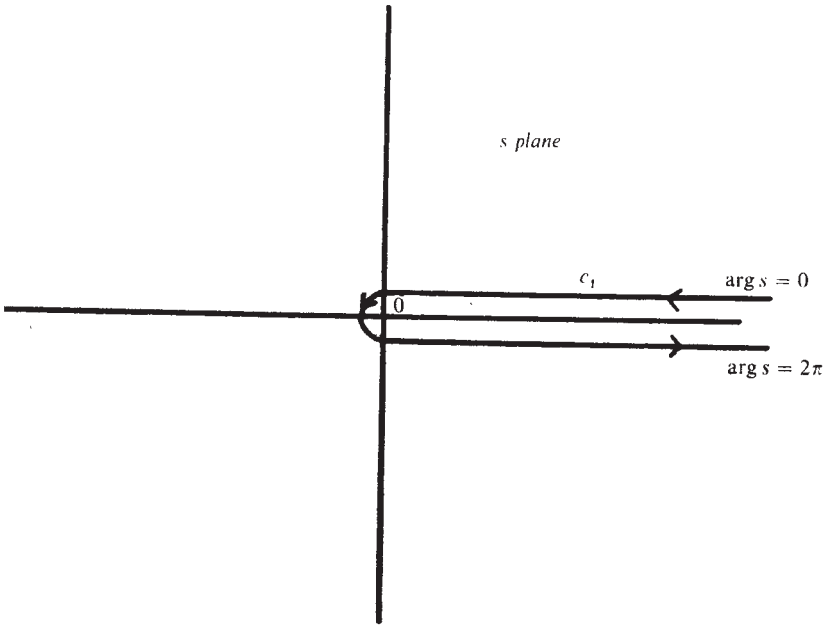


Fig. 5(a)

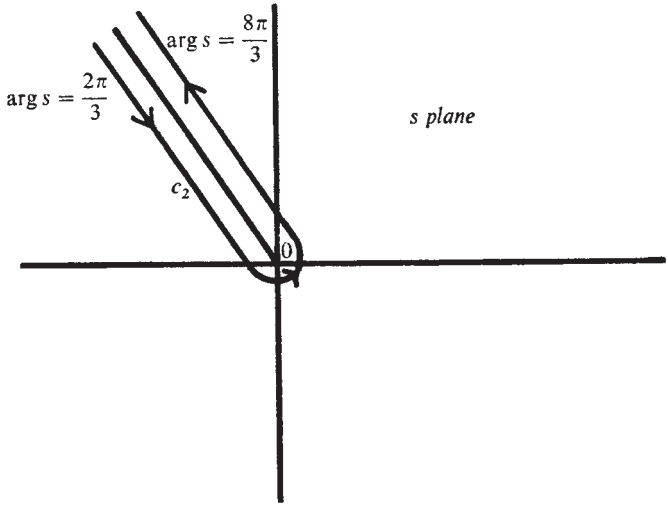


Fig. 5(b)

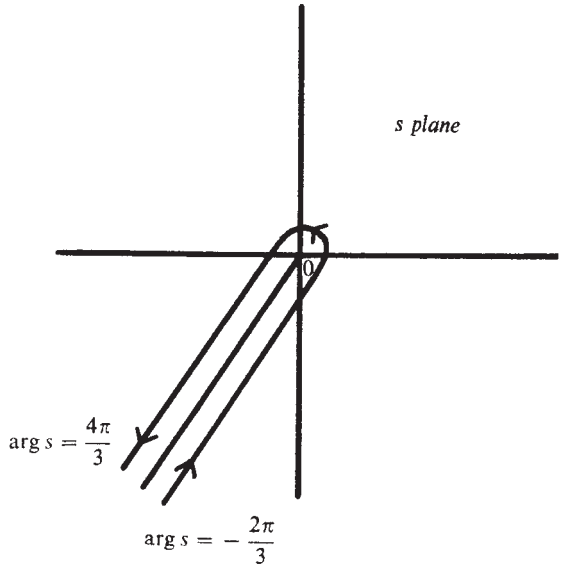


Fig. 5(c)

it is more convenient to consider the alternative forms

$$u_i(z) = \int_{c_i} \exp(zs - \tfrac{1}{3}s^3)\frac{f(s)}{s^k}\,ds \quad (i = 1, 2, 3), \tag{4.3}$$

where $f(s)$ is defined by (2.4), and the associated solutions $v_i(z)$ are obtained by reversing the sign of m, as before. The contour c_1 starts at $\infty \exp(0)$, encircles the origin once in the positive sense, and terminates at $\infty \exp(2\pi i)$; c_2 starts and finishes at $\infty \exp(2\pi i/3)$ and $\infty \exp(8\pi i/3)$, respectively, having also encircled $s = 0$; and c_3 starts and finishes at $\infty \exp(-2\pi i/3)$ and $\infty \exp(4\pi i/3)$, respectively. The contours are sketched in Figs. 5. It is easy to see that

$$u_1 = U_1 + U_2 + U_3, \quad u_2 = U_1 + U_2 + U_3^*, \quad u_3 = U_1 + U_2^* + U_3, \tag{4.4}$$

and similarly for the v_i.

The asymptotic forms of the solutions u_i are listed for each sector T_j in Tables 1 and 2, for $P > 1$ and $P < 1$, respectively, the order of dominancy being as described in §3. It may be seen immediately that the dominant contribution to u_i is inviscid in $S_j \cup S_k$, where (i, j, k) is a permutation of $(1, 2, 3)$, and the S_i are the anti-Stokes sectors illustrated in Fig. 4 (b).

In the case $\gamma U_c' > 0$, the real y-axis through the critical point is mapped, according to (1.14) into the rays $\arg z = \pi/6$ for $y > y_c$, and $\arg z = 7\pi/6$ for $y < y_c$, and these both lie in $S_1 \cup S_2$. It follows that the inviscid solutions to the physical problem are provided directly by u_3 and v_3. Also, referring to Tables 1 and 2, we have

$$u_3 \sim \phi_+ \equiv \frac{2\pi i}{\Gamma(k)}(i\gamma \mathrm{R}U_c')^{\frac{1}{3}-m}(y - y_c)^{\frac{1}{3}-3m}, \tag{4.5}$$

for $y - y_c \to +\infty$, while for $y - y_c \to -\infty$ we have

$$u_3 \sim \phi_- \equiv A^{\frac{1}{2}} \frac{2\pi i}{\Gamma(k)} (i\gamma \mathrm{R} U_c')^{\frac{1}{2}-m} (y_c - y)^{\frac{1}{2}-3m}, \tag{4.6}$$

so that

$$\phi_- = Q(m) \left\{ \frac{2\pi i}{\Gamma(k)} (i\gamma \mathrm{R} U_c')^{\frac{1}{2}-m} (y_c - y)^{\frac{1}{2}-3m} \right\}. \tag{4.7}$$

Thus, on the left of y_c the solution has the same form as on the right except that its amplitude is multiplied by the factor $Q(m)$ and, of course, $y_c - y$ replaces $y - y_c$. In a similar way, v_3 experiences an amplitude change of $Q(-m)$ in crossing y_c from right to left.

If $\gamma U_c' < 0$, the inviscid solutions are given by u_2 and v_2 provided $\gamma U_c'$ is interpreted in (1.14) as $|\gamma U_c'| \exp(-\pi i)$. If instead it is taken to be $|\gamma U_c'| \exp(\pi i)$, solutions u_1 and v_1 must be used. In either case, a similar argument shows that the amplitude enhancement in the inviscid parts of u and v in crossing y_c from right to left are $-Q(-m)$ and $-Q(m)$ respectively. All these conclusions are independent of the size of P, as may be verified by using Table 2 in place of Table 1.

References

J. R. Booker and F. P. Bretherton, 1967, *J. Fluid Mech.*, 27, 513.
K. S. Gage and W. H. Reid, 1968, *J. Fluid Mech.*, 33, 21.
P. Hazel, 1967, *J. Fluid Mech.*, 14, 257.
D. Koppel, 1964, *J. Math. Phys.*, 5, 963.
L. J. Slater, 1960, *Confluent hypergeometric functions*, (Cambridge University Press).
E. T. Whittaker and G. N. Watson, 1927, *A course of modern analysis*, 4th Ed. (Cambridge University Press).

The Department of Engineering Mathematics,
The University,
Newcastle upon Tyne.

and

The School of Mathematics,
The University,
Newcastle upon Tyne.

(*Received on the 22nd of December*, 1969.)

Printed by C. F. Hodgson & Son, Ltd., 23 Pakenham Street, London, W.C.1.

THE ASTROPHYSICAL JOURNAL, 162:699–701, November 1970

NOTES

ON THE HYDROMAGNETICS OF ROTATING FLUIDS

IBRAHIM A. ELTAYEB AND PAUL H. ROBERTS
School of Mathematics, University of Newcastle upon Tyne, England
Received 1970 May 18

ABSTRACT

There are two essentially different states of motion which can occur when a conducting fluid is driven by given forces (e.g., buoyancy): either (*a*) a state in which the magnetic field is zero and the amplitude of motion is such that viscous friction disposes of all energy available or (*b*) one in which a magnetic field is present, and in which viscous and ohmic dissipation jointly balance the energy supplied. Apparently, nature often "chooses" solution (*b*), and, as a result of the study of a simple model, it is suggested here that Coriolis forces may be responsible for this.

It is generally accepted that the magnetic fields of cosmical bodies such as the Earth are the result of the dynamo action of motions in their electrically conducting interiors. Since this induction process is probably closely linear, it cannot by itself determine the magnitude of the resulting steady (or quasi-steady) field, $\boldsymbol{B}$; this is decided by a balance between the rate of working of the forces (e.g., buoyancy) driving the fluid motions and the rate of dissipation of energy by resistive and frictional losses. The *dynamical* reason is not clear why the system should "prefer" this final state of nonzero $\boldsymbol{B}$ to the equally valid solution in which the field vanishes and the driving and viscous forces alone balance. It is suggested here that Coriolis forces may provide the answer, and their effect is illustrated by a simple convective system, the so-called Bénard layer (Chandrasekhar 1961).

Consider a plane horizontal layer of uniform conducting fluid which is of depth d and rotates as a solid body, with angular velocity $\boldsymbol{\Omega}$, about the vertical, there being also a uniform magnetic field, $\boldsymbol{B}_0$, present in that direction. The top and bottom surfaces of the fluid are assumed to be maintained at fixed temperatures, the lower surface being at a temperature βd higher than the upper. This variant of the well-known Bénard situation was first studied by Chandrasekhar (1961, chap. 4, for example), who showed that in the Boussinesq approximation the stability of the layer to convective perturbations depends on five dimensionless parameters: the Rayleigh number, R; the Taylor number, T; the Hartmann number, M; the Prandtl number, p; and the magnetic Prandtl number, p_m, where

$$R = \frac{g\alpha\beta d^4}{\kappa\nu}, \quad T = \frac{4\Omega^2 d^4}{\nu^2}, \quad M = B_0 d\left(\frac{\sigma}{\rho\nu}\right)^{1/2}, \quad p = \frac{\nu}{\kappa}, \quad p_m = \mu\sigma\nu, \quad (1)$$

where g is the acceleration due to gravity, α is the coefficient of volume expansion, ν is the kinematic viscosity, κ is the thermal diffusivity, ρ is the density, μ is the permeability, and σ is the electrical conductivity of the fluid (mks units). If R is gradually increased from zero, no permanent convective motions can occur until R reaches a critical value, R_c, which, together with the wavenumber a_c of the motions which then ensue, depends on T, M, p, and p_m in general. The motions may be steady (principle of exchange of stabilities) or oscillatory ("overstability"), and in the former case R_c does not depend on p or p_m. Since T and M are generally large in cosmical bodies, we will restrict our atten-

tion to the limits $T \to \infty$, $M \to \infty$. In one important respect, our layer does not faithfully model the cosmical situation: in the astrophysical and geophysical context, $\boldsymbol{B}_0$ is not applied externally but is maintained by the body itself through dynamo action. To include this self-induction here would, however, obscure the physical effect we wish to examine, and would also introduce severe analytical complications.

In the absence of a magnetic field, Chandrasekhar showed (1961, chap. 2) that to leading order

$$R_c \sim KT^{2/3}, \quad a_c \sim CT^{1/6}, \quad T \to \infty, \tag{2}$$

where K and C are constants. If $p > 0.67659\ldots$, the principle of exchange of stabilities applies, and $K = 3(\pi^2/2)^{2/3}$, $C = (\pi^2/2)^{1/6}$. If $p < 0.67659\ldots$, overstability occurs and K and C depend on p. In the absence of rotation, Chandrasekhar showed (chap. 3) that to leading order

$$R_c \sim KM^2, \quad a_c \sim CM^{1/3}, \quad M \to \infty. \tag{3}$$

If $p_m < p$, the principle of exchange of stability applies, and $K = \pi^2$, $C = (\pi^4/2)^{1/6}$. If $p_m > p$, overstability occurs and K and C depend on p and p_m. Since p_m is probably small in fluids such as the Earth's core, we will restrict the discussion below to the limit $p_m \to 0$. This has the effect of inhibiting overstability except in the rotationally dominant cases. Also, for definiteness (and simplicity) we will suppose both boundaries to be free and electrically insulating. It should be noted, however, that, apart from numerical constants of order unity, the conclusions we reach are not dependent on this choice of surface conditions.

Consider now the joint effects of rotation and magnetic field. It is immediately clear that the asymptotic form of R_c in the double limit $T \to \infty$, $M \to \infty$ will depend on the relative magnitudes of T and M. We can, in fact, distinguish three main cases:

1. *Magnetic field dominating:* $T < 7.30\ M^{8/3}$. Steady convection occurs, and equation (3) applies, to leading order.

2. *Magnetic field and rotation equally influential:* $7.30\ M^{8/3} \leq T \leq 0.01\ M^6$ (or, more precisely, $[2\pi^2]^{2/3}M^{8/3} \leq T \leq 27M^6/256\pi^2$). Steady convection occurs, and

$$\left.\begin{aligned} R_c &\sim \pi^2 \frac{[(M^4+T)^{1/2}+T^{1/2}]}{(M^4+T)^{1/2}} \left\{M^2 + [(M^4+T)^{1/2}+T^{1/2}]\frac{\pi^2 T^{1/2}}{M^2}\right\}, \\ a_c &\sim \pi \frac{(M^4+T)^{1/4}}{T^{1/4}}, \end{aligned}\right\} T \to \infty. \tag{4}$$

It should be noted that, if $T = O(M^4)$, this implies that $R_c = O(T^{1/2})$ and $a_c = O(1)$, for $T \to \infty$.

3. *Rotation dominates:* $T \geq 0.01\ M^6$. The situation is essentially that described above in the field-free case, and overstability again occurs if $p < 0.67659$.

We wish to draw attention to the fact that, except for $T < O(M^3)$ and $T > O(M^6)$, *the critical Rayleigh number, R_c, in the presence of the magnetic field is smaller* (in fact, asymptotically infinitely smaller) *than its value in the absence of magnetic field* (as given in eq. [2]); i.e., *convection is more easily excited with a field present than without one.* For example, in the interesting range $T = O(M^4)$, R_c is $O(T^{1/2})$ rather than $O(T^{2/3})$, for $T \to \infty$. As a corollary, we observe that the magnetic field actually facilitates the transport of heat, a fact which may have some relevance in the evolution of a protostar.

The physical explanation of the phenomenon may be sought in the observation that, where no magnetic field is present, the "deflecting" acceleration created by the Coriolis force can only be balanced by viscous shear, and this shear can only reach the required magnitude by the creation of large velocity gradients associated with small-scale motion ($a_c = O(T^{1/6})$). When magnetic fields of sufficient strength, however, are present, the

Coriolis deflections in the main body of the layer are met by Lorentz forces. The associated motions are *not* necessarily of small scale (in fact, $a_c = O(1)$ for $T = O(M^4)$; see also Chandrasekhar 1961, § 54); the ohmic losses which buoyancy must make good are therefore not so great and the critical Rayleigh number is correspondingly smaller. The fact that viscosity plays little part in the internal dynamics of the $T = O(M^4)$ range is also apparent when it is realized that T/M^4 and $R_c/T^{1/2}$ are independent of ν.

At a general point in a cosmical body, the directions of gravity, rotation, and magnetic field will not of course be parallel, but will be at arbitrary angles to each other. It is therefore of interest to observe that the main conclusions of this note hold good for other orientations of these fields and for other boundary conditions on flow and field. This will be made clearer in a more detailed forthcoming study by one of us (I. A. E.).

REFERENCES

Chandrasekhar, S. 1961, *Hydrodynamic and Hydromagnetic Stability* (Oxford: Clarendon Press).

Astrophysical Letters, 1971, Vol. 9, pp. 79–80

A Transformation of the Stellar Wind Equations

P. H. ROBERTS *National Center for Atmospheric Research, Boulder, Colorado 80302, USA*

A simple transformation is demonstrated which reduces the number of parameters in the stellar wind equations from two to one.

It seems not to have been noticed that the inviscid stellar wind equations admit a simple transformation which reduces them to a form involving only one parameter.

The temperature, kinetic energy density, and gravitational energy in a stellar wind are often cast in a non-dimensional form due to Chamberlain (1961). Let

$$\tau = \frac{T}{T_1}, \qquad \psi = \frac{mw^2}{kT_1}, \qquad \lambda = \frac{GmM}{kT_1 r}, \tag{1}$$

where T is the temperature, w is the radial expansion velocity at a distance r from the stellar center, and T_1 is a reference temperature corresponding in this case to that of the base of the corona where $r = r_1$; M is the mass of the star, m is half the mass of a hydrogen atom, G is the universal constant of gravitation, and k is Boltzmann's constant. The energy and heat conduction equations then become, respectively,

$$\frac{1}{2}\frac{\mathrm{d}\psi}{\mathrm{d}\lambda} = \frac{1-2\tau/\lambda-\mathrm{d}\tau/\mathrm{d}\lambda}{1-\tau/\psi}, \tag{2}$$

$$\frac{1}{2}A\tau^{5/2}\frac{\mathrm{d}\tau}{\mathrm{d}\lambda} = \varepsilon_\infty + \lambda - \frac{1}{2}\psi - \frac{5}{2}\tau, \tag{3}$$

where $\varepsilon_\infty kT_1$ is the residual energy per particle at infinity, and

$$A = \frac{2GmM\kappa_0 T_1^{3/2}}{k^2 F}. \tag{4}$$

Here $\kappa_0 T^{5/2}$ is the thermal electron conductivity, and F is the net outward particle flux, Nwr^2.

It may be verified that a further transformation

$$\tau = \varepsilon_\infty \tau', \qquad \psi = \varepsilon_\infty \psi', \qquad \lambda = \varepsilon_\infty \lambda', \tag{5}$$

reduces equations (2) and (3) to the forms

$$\frac{1}{2}\frac{\mathrm{d}\psi'}{\mathrm{d}\lambda'} = \frac{1-2\tau'/\lambda'-\mathrm{d}\tau'/\mathrm{d}\lambda'}{1-\tau'/\psi'}, \tag{6}$$

$$K\tau'^{5/2}\frac{\mathrm{d}\tau'}{\mathrm{d}\lambda'} = 1 + \lambda' - \frac{1}{2}\psi' - \frac{5}{2}\tau', \tag{7}$$

where

$$K = \frac{1}{2}A\varepsilon_\infty^{3/2} = \frac{GmM\kappa_0}{k^2 F}(\varepsilon_\infty T_1)^{3/2}. \tag{8}$$

The advantages of employing equations (6) and (7) are obvious. The entire set of stellar winds are contained in a single one parameter family. (The Chamberlain solar breeze solutions still require separate treatment, however, although an analogous transformation eliminates all parameters from equations 6 and 7.) In a practical case one could proceed as follows.

Suppose that r_1, the radius of the corona, and the mass of the star are known. Choose K arbitrarily. Select the solution to (6) and (7) which passes through a critical point, λ_c', and is supersonic at $r = \infty$. Let $\lambda' = \lambda_1'$ be any value of λ' on the subsonic side of λ_c'. This point can be put into correspondence with the coronal surface by observing that $\tau = 1$ thereat, i.e., by (5), $\varepsilon_\infty = 1/\tau_1'$. Manipulation of (1), (5) and (8) now gives

$$T_1 = \frac{GmM}{kr_1}\left(\frac{\tau_1'}{\lambda_1'}\right), \tag{9}$$

$$w_1 = \left(\frac{GM}{r_1}\right)^{1/2}\left(\frac{\psi_1'}{\lambda_1'}\right)^{1/2}, \tag{10}$$

$$N_1 = \left[\kappa_0\left(\frac{m}{k}\right)^{1/2}\frac{(GmM)^2}{(kr_1)^3}\right]\frac{1}{K\lambda_1'\psi_1'^{1/2}}. \tag{11}$$

If one or other of T_1 or N_1 is known, a search over the solution may often be expected to locate the λ_1' for which equation (9) or (11) agrees with that observation. If both T_1 and N_1 are known, as for example in the case of the Sun, then a search over K is also involved. It is possible, using the ideas

presented here, to reduce considerably the labor involved in surveying stellar wind solutions. This will be considered further by B. Durney and the author in a paper to be presented elsewhere.

The author thanks Dr. B. Durney for interesting him in this problem and for some helpful discussions. The permanent address of the author is the School of Mathematics, University of Newcastle upon Tyne, England. The National Center for Atmospheric Research is sponsored by the National Science Foundation.

REFERENCE

Chamberlain, J. W., 1961, *Astrophys. J.*, **133,** 675.

Received in final form 28 June 1971

A HAMILTONIAN THEORY FOR WEAKLY INTERACTING VORTICES

P. H. ROBERTS

It is shown that the equations governing the motion of a set of circular vortex rings, separated by distances large compared with their dimensions, can be written in canonical form.

1. *Introduction.* The reader of Part II of J. J. Thomson's award winning Adams Prize essay of 1882 can scarcely fail to be struck by the remarkable results of his analysis of the interaction of two widely separated vortex rings moving in an incompressible fluid. The relations governing the angle of deflection and the exchange of impulse (or, as we shall prefer to say, "momentum") are strongly reminiscent of those given in standard texts for scattering under central forces (*e.g.* Landau and Lifshitz, 1960, §§17–18). The suggestion that the phenomenon can be described from the standpoint of classical hamiltonian dynamics of interacting particles is strong. Of course, vortex rings cannot strictly be likened to elastic particles (*i.e.* those which have no internal degrees of freedom); a vortex has an infinite number of degrees of freedom, as is evinced by the infinite variety of vortex waves they can carry, and which are excited in any "collision" with another ring, a fact J. J. Thomson (1883) clearly recognized. But when vortex rings scatter only weakly, as happens when their separation is always large compared with their overall diameters during the collision, the energy in these internal motions need not be considered in deciding the orbits of the rings, *i.e.* the collision can be considered to be elastic.

In this short paper, we present a hamiltonian dynamics for weakly interacting vortices. Apart from its intrinsic interest, we also have in mind applications to quantum fluids. Here vortices represent a class of elementary excitations of the entire fluid system and, as such, they ought to obey hamiltonian dynamics. It has, in fact, long been recognized that the large circular vortex ring, moving in isolation in a fluid at rest at infinity, *approximately* obeys the relation

$$\frac{dq_i}{dt} = \frac{\partial H}{\partial p_i}, \tag{1}$$

where $\mathbf{q}$ is the centre of the ring, $\mathbf{p}$ its momentum, and H is the kinetic energy of fluid motion, $\mathbf{u}$, in the laboratory frame, $\mathscr{F}$, in which the fluid is at rest at infinity. More recently, Roberts and Donnelly (1970) have observed that, correctly interpreted, (1) is in fact *exact* to the leading terms in the asymptotic expansion in a/R for H, $\mathbf{p}$ and $d\mathbf{q}/dt$, for the large circular ring. Here R is the overall radius of the ring, and a is the radius of its core. Rings of moderate a/R do not have circular cross-sections, and it is necessary to redefine a and R; for example, the cross-sectional area can be taken to be πa^2, and R can be chosen to be the mean of the closest and furthest points of the vortex core from the axis of symmetry (Fraenkel, 1970). We show below (§2) that (1) holds exactly (*i.e.* for all a/R) for circular vortex rings, no matter what their core structure may be. We also present (§3) a hamiltonian theory for the weak

[MATHEMATIKA **19** (1972), 169–179]

interaction of rings; in doing so, we locate (§4) and correct an error made by J. J. Thomson (*loc. cit.*).

2. *The single vortex hamiltonian.* Our objective in this section is that of proving (1) for any axially symmetric distribution of vorticity (confined to a finite region, $\mathscr{R}$, containing the origin, O, in an infinite fluid at rest at infinity) which moves steadily *without change of form* with speed dq/dt along the axis of symmetry, Oz.

We must first decide how (1) should be understood, and particularly how the partial derivative signs displayed there should be interpreted. We consider an infinitesimal axisymmetric displacement $\delta\boldsymbol{\xi}(\mathbf{x}, t)$ of the equilibrium which, in obedience to Kelvin's theorem, advects the vorticity $\boldsymbol{\omega}$ (= curl $\mathbf{u}$) of the initial state to another axially symmetric distribution. In general, the new state will not be steady in any frame of reference, since vortex waves will have been excited by $\delta\boldsymbol{\xi}$. We suppose, however, that a $\delta\boldsymbol{\xi}$ can be found for which the displaced configuration of vorticity moves steadily along Oz, again without change of form. We explicitly rule out all members of the infinite family of $\delta\boldsymbol{\xi}$ which merely rearrange the vortex lines of the undisturbed configuration without change of overall structure. In other words, we suppose that the new configuration has an (infinitesimally) *different* energy, $H + \delta H$, a *different* momentum $\mathbf{p} + \delta\mathbf{p}$, and also a different velocity. The existence of one $\delta\boldsymbol{\xi}$ with this property implies the existence of an infinite family, whose members differ from each other only by a displacement which rearranges vortex lines within the perturbed configuration without change of overall structure, and which yield the same δH and $\delta\mathbf{p}$. Apart from this trivial degeneracy, we shall regard $\delta\boldsymbol{\xi}$ as unique, up to a multiplication factor.

We are now in a position to interpret the partial derivative in (1): $\partial H/\partial p_i$ is the limiting value of $\delta H/\delta p_i$ as $\delta\boldsymbol{\xi} \to 0$.

Since the remainder of this section rests on the assumption that a $\delta\boldsymbol{\xi}$ exists with the properties described above, we should enquire how reasonable the hypothesis is. Dyson (1893) and Fraenkel (1970, 1972) have made it plain that, for one particular core structure (the so-called "standard model" $f(\bar{\psi}) =$ constant; see (7) below), a sequence of vortex rings exists which range continuously from a large circular vortex ($R \gg a$) at one extreme, to the Hill's spherical vortex at the other. The required $\delta\boldsymbol{\xi}$ transforms one member of this sequence into an adjacent member. The existence of the necessary $\delta\boldsymbol{\xi}$ is therefore certainly established for this model. Moreover, it is elementary to show that the first four† terms in the asymptotic expansion in a/R of H, $\mathbf{p}$ and $d\mathbf{q}/dt$ obey (1); see appendix. Further, Fraenkel (1972) showed that, for all sufficiently small a/R, similar families of vortex rings exist for a wide variety of core structures (*i.e.* for an infinite number of choices of $f(\bar{\psi})$ in (7) below). Moreover, it is easy to demonstrate, using his expressions (4·3) and (4·6), that (1) is obeyed for the first two† terms in the a/R expansions of H, $\mathbf{p}$ and $d\mathbf{q}/dt$.

To proceed more formally, we define cylindrical coordinates (ϖ, ϕ, z) with Oz as axis of symmetry, and introduce the Stokes streamfunction, $\bar{\psi}(\varpi, z)$, in terms of which

$$\bar{u}_\varpi = \frac{1}{\varpi}\frac{\partial\bar{\psi}}{\partial z}, \quad \bar{u}_\phi = 0, \quad \bar{u}_z = -\frac{1}{\varpi}\frac{\partial\bar{\psi}}{\partial\varpi}. \tag{2}$$

† In making this remark, we count terms in, for example, $(a/R)^n$ and $(a/R)^n \log(R/a)$ as being of different order in a/R.

Here and elsewhere, all variables with superimposed bars refer to the co-moving frame, $\bar{\mathscr{F}}$, in which the flow is steady; variables without bars refer to the (instantaneously coinciding) laboratory frame, $\mathscr{F}$, in which $\mathbf{u} \to 0$ for $r \equiv \sqrt{(\varpi^2 + z^2)} \to \infty$. It is only in $\bar{\mathscr{F}}$ that the integral

$$H = \tfrac{1}{2}\rho \int_{\substack{\text{all} \\ \text{space}}} \mathbf{u}^2 \, dV \qquad (\rho = \text{density}), \tag{3}$$

for the hydrodynamic energy of flow converges, and therefore has meaning. Clearly, since $\mathbf{u} = \bar{\mathbf{u}} + d\mathbf{q}/dt$, we have

$$\bar{\psi} = \psi + \tfrac{1}{2}\varpi^2 \, dq/dt, \tag{4}$$

where $\psi = O(r^{-2})$ for $r \to \infty$.

Only the ϕ-component, ω, of the vorticity, curl $\mathbf{u}$, of the flow (2) can be non-zero; it is

$$\omega = \bar{\omega} = \varpi^{-1} \Delta \bar{\psi}, \tag{5}$$

where Δ is the modified Laplacian operator:

$$\Delta = \varpi \frac{\partial}{\partial \varpi} \left(\frac{1}{\varpi} \frac{\partial}{\partial \varpi} \right) + \frac{\partial^2}{\partial z^2} . \tag{6}$$

It is well known (*e.g.* Lamb, 1932, p. 245) that, when a form preserving vorticity distribution occurs, a function $f(\bar{\psi})$ exists such that

$$\omega = \varpi f(\bar{\psi}). \tag{7}$$

The function f defines the core structure. Outside $\mathscr{R}$, f is zero.

The simplest choice for f is a constant value throughout $\mathscr{R}$; this defines the "standard model" which has been extensively studied. Whenever, as in that case, f is discontinuous at the surface S of $\mathscr{R}$, the solution $\bar{\psi}$ of (5) and (7) must be constant and continuous on S; its normal derivative must also be continuous. For simplicity, we will suppose that f approaches zero continuously as S is approached from within. This restriction can later be lifted, but it does not preclude f which are small only in a thin layer near S.

We now introduce the hypothesis discussed at the beginning of the section. We consider an infinitesimal axisymmetric displacement, $\delta\boldsymbol{\xi}$, defined by a singlevalued streamfunction $\delta\psi$, so that

$$\delta\boldsymbol{\xi} = \left(\frac{1}{\varpi} \frac{\partial}{\partial z} \delta\psi, \; 0, \; -\frac{1}{\varpi} \frac{\partial}{\partial \varpi} \delta\psi \right), \tag{8}$$

and suppose that the vortex lines are merely advected by this displacement. Since $\delta\psi$ is infinitesimal, the corresponding change, $\delta\boldsymbol{\omega}$, in $\boldsymbol{\omega}$ is given by

$$\delta\boldsymbol{\omega} = \nabla \times (\delta\boldsymbol{\xi} \times \boldsymbol{\omega}), \tag{9}$$

of which only the ϕ-component,

$$\delta\omega = \frac{\partial(\delta\psi, \omega/\varpi)}{\partial(\varpi, z)} , \tag{10}$$

can be non-zero. We may note, using (7), that

$$\bar{\psi}\delta\omega = \frac{\partial(\delta\psi, G(\bar{\psi}))}{\partial(\varpi, z)} , \tag{11}$$

where

$$G(\bar{\psi}) = \int^{\bar{\psi}} \bar{\psi} f'(\bar{\psi})\, d\bar{\psi}. \tag{12}$$

By (2), (3), (4), (8), (11) and (12), we have

$$\begin{aligned}
\delta H &= \rho \int (u_\varpi \delta u_\varpi + u_z \delta u_z)\, dV \\
&= \rho \int \left(\delta u_\varpi \frac{\partial \psi}{\partial z} - \delta u_z \frac{\partial \psi}{\partial \varpi} \right) \frac{dV}{\varpi} \\
&= -\rho \int_{\mathscr{R}} \psi \delta\omega \frac{dV}{\varpi} \\
&= -\rho \int_{\mathscr{R}} \bar{\psi} \delta\omega \frac{dV}{\varpi} + \tfrac{1}{2}\rho \frac{dq}{dt} \int_{\mathscr{R}} \varpi \delta\omega dV \\
&= -2\pi\rho \int\int \frac{\partial(\delta\psi, G(\bar{\psi}))}{\partial(\varpi, z)}\, d\varpi dz + \delta p \frac{dq}{dt},
\end{aligned} \tag{13}$$

where

$$\mathbf{p} = \tfrac{1}{2}\rho \int_{\mathscr{R}} (\mathbf{x} \times \boldsymbol{\omega})\, dV \tag{14}$$

is the momentum (impulse) of the vorticity distribution. Now, if A is a meridian section of $\mathscr{R}$, and P is its periphery, we have

$$\int\int_A \frac{\partial(\delta\psi, G)}{\partial(\varpi, z)}\, d\varpi dz = \int_A d\mathbf{A} \cdot \nabla \times (G \nabla \delta\psi) = \int_P G d\mathbf{p} \cdot \nabla(\delta\psi). \tag{15}$$

Since $\bar{\psi}$ is constant over P, so is $G(\bar{\psi})$, and this term can be withdrawn from the integral on the right of (15). The remaining expression clearly vanishes on integration. (For the same reason, the lower limit in definition (12) may be left indefinite.) Thus, finally, (13) reduces to

$$\delta H = \delta p \frac{dq}{dt}, \tag{16}$$

and the result (1) follows in the limit.

If f is discontinuous across S, a new integral appears on the right of (13)†, which is proportional to the value $\bar{\psi}_s$ of $\bar{\psi}$ on S, to the discontinuity in f, and to the change in $\delta\psi$ round P. The last of these vanishes, and (16) therefore follows once more.

3. *The interaction hamiltonian.* In this section we set up a hamiltonian for a set of small circular vortex rings, where by "small" we mean that the separation between rings is large compared with their dimensions. We give two arguments, one in the spirit of J. J. Thomson's analysis, and one which is slightly more direct.

† I am grateful to Dr. John Norbury for this remark.

When all other vortices are remote from the particular ring, $\mathscr{R}$, under consideration, the fluid velocity, $\mathbf{v}$, they create can be represented over $\mathscr{R}$ by the approximate expression

$$v_i = a_i + b_{ij} x_j, \tag{17}$$

where the coefficients

$$a_i = v_i(0, t), \qquad b_{ij} = (\partial v_i/\partial x_j)_{(0,\, t)}, \tag{18, 19}$$

are functions of t alone. To adequate approximation, we may assume $\mathscr{R}$ to be a circular *filament* of radius R, and in writing (18) and (19) we have supposed that this circle is instantaneously centred at the origin, and lies in the $x_1 x_2$-plane. We may note that, since the fluid is of uniform density,

$$b_{ii} = 0, \tag{20}$$

Here, and below, we adopt the summation convention for Roman suffices.

The a_i term in (17) merely advects $\mathscr{R}$ without alteration in orientation or shape, *i.e.* without change of momentum; we will temporarily ignore it. Consider the $i = 3$ terms in (17). These act to tip $\mathscr{R}$ out of the $x_1 x_2$-plane, *i.e.* to change the *direction* of $\mathbf{p}$. In a time δt, the point P on $\mathscr{R}$ with coordinates $(x_1, x_2, 0)$ is carried by the flow (17) to the point (x_1, x_2, x_3), where

$$x_3 = (b_{31} x_1 + b_{32} x_2)\, \delta t. \tag{21}$$

This is the equation of a plane through O with unit normal $\mathbf{n} = (-b_{31}\, \delta t, -b_{32}\, \delta t, 1)$. The vortex lies in this plane at time δt, and its momentum is then $p\mathbf{n}$. Taking the limit $\delta t \to 0$, we have

$$\frac{dp_1}{dt} = -pb_{31}, \qquad \frac{dp_2}{dt} = -pb_{32}. \tag{22, 23}$$

Consider next the $i = 1$ and $i = 2$ terms in (17). These act to change the cross-sectional area encircled by the filament, *i.e.* they change the *magnitude* of $\mathbf{p}$. (They also excite waves on $\mathscr{R}$). Since $p = \kappa \pi \rho R^2$, where κ is the circulation round the core of $\mathscr{R}$, we have

$$\frac{dp}{dt} = \frac{2p}{R} \frac{dR}{dt}. \tag{24}$$

The point P on $\mathscr{R}$ with coordinates $(R\cos\theta,\ R\sin\theta,\ 0)$ has the radial velocity

$$v_R = v_1 \cos\theta + v_2 \sin\theta, \tag{25}$$

from the origin and in the $x_1 x_2$-plane. By (17) we have

$$v_R = R[b_{11}\cos^2\theta + b_{22}\sin^2\theta + (b_{12} + b_{21})\sin\theta\cos\theta]. \tag{26}$$

Averaging round $\mathscr{R}$, we obtain from (20) and (26)

$$\frac{dR}{dt} = \frac{1}{2\pi} \int_0^{2\pi} v_R\, d\theta = \tfrac{1}{2} R(b_{11} + b_{22}) = -\tfrac{1}{2} R b_{33}, \tag{27}$$

or, using (24),

$$\frac{dp_3}{dt} = -pb_{33}. \tag{28}$$

On returning to general axes, we obtain from (22), (23) and (28) the astonishingly simple vector result

$$\frac{dp_i}{dt} = -p_j \frac{\partial v_j}{\partial x_i}, \tag{29}$$

where $\partial v_j/\partial x_i$ is to be evaluated at the instantaneous centre of the vortex. The expression (29) can be put into correspondence with results given on pages 65 and 68 of J. J. Thomson's essay. Equation (29) may also be obtained by recalling that the momentum (14) may, for a vortex filament, be transformed into $\rho\kappa$ times the vector surface area of an open surface S of which it is the periphery (see Lamb, 1932, Art. 152). Taking S to be a material surface, we see by standard kinematical arguments that each element $\mathbf{dS}$ of S satisfies (29) with p set equal to $\mathbf{dS}$. Assuming the rates of strain are constant over the ring, as is appropriate for distant collisions, (29) follows on integration over S†.

Consider now a set of widely separated vortex rings, $\mathscr{R}_\alpha$, situated at $\mathbf{q}_\alpha$ and having momenta $\mathbf{p}_\alpha$. (Greek letters are used for vortex labels, and the summation convention is *not* adopted for these.) The velocity potential ($\mathbf{u} = -\nabla\Phi$) of the flow created by $\mathscr{R}_\alpha$ is $\kappa_\alpha/4\pi$ times the solid angle subtended by the area circled by $\mathscr{R}_\alpha$ at the point $\mathbf{x}$ considered. At great distances ($|\mathbf{x} - \mathbf{q}_\alpha| \gg \mathscr{R}_\alpha$), this gives to leading order

$$\Phi_\alpha(\mathbf{x}, t) = -\frac{1}{4\pi\rho}(\mathbf{p}_\alpha \cdot \nabla_\alpha)\left[\frac{1}{|\mathbf{x} - \mathbf{q}_\alpha|}\right], \tag{30}$$

where the differentiation is with respect to $\mathbf{q}_\alpha$ as the suffix α on the ∇ signifies. It follows that the flow created at $\mathscr{R}_\alpha$ by the remaining vortices is

$$\mathbf{v}_\alpha(\mathbf{x}, t) = -\frac{1}{4\pi\rho}\sum_\beta{}' (\mathbf{p}_\beta \cdot \nabla_\beta)\nabla\left[\frac{1}{|\mathbf{x} - \mathbf{q}_\beta|}\right], \tag{31}$$

where, as is usual in statistical physics, the dash against the summation sign indicates that the sum over β excludes the ring $\mathscr{R}_\alpha$ itself. Equation (29) now becomes

$$\frac{dp_{\alpha i}}{dt} = -\frac{1}{4\pi\rho}\sum_\alpha{}' p_{\alpha j} p_{\beta k} \frac{\partial^3}{\partial q_{\alpha i}\partial q_{\alpha j}\partial q_{\alpha k}}\left[\frac{1}{|\mathbf{q}_\alpha - \mathbf{q}_\beta|}\right]. \tag{32}$$

Denote the one-vortex hamiltonian of §2 for $\mathscr{R}_\alpha$ by H_α, and reserve H from now onwards for the quantity

$$H = \sum_\alpha H_\alpha + \sum_\alpha \sum_\beta{}' H_{\alpha\beta}, \tag{33}$$

where

$$H_{\alpha\beta} = -\frac{1}{8\pi\rho} p_{\alpha i} p_{\beta j} \frac{\partial^2}{\partial q_{\alpha i}\partial q_{\beta j}}\left[\frac{1}{|\mathbf{q}_\alpha - \mathbf{q}_\beta|}\right] = H_{\beta\alpha}. \tag{34}$$

Clearly, by (32), we have

$$\frac{dp_{\alpha i}}{dt} = -\frac{\partial H}{\partial q_{\alpha i}}. \tag{35}$$

Moreover, the velocity of $\mathscr{R}_\alpha$ is compounded from the self-induced motion given by

† I am grateful to Dr. A. Soward for this remark.

(1), and the flow $\mathbf{v}_\alpha(\mathbf{q}_\alpha, t)$ created by the remaining rings, and given by (31). Thus, by (33) and (34), we have

$$\frac{dq_{\alpha i}}{dt} = \frac{\partial H}{\partial p_{\alpha i}}. \tag{36}$$

Equations (35) and (36) are the canonical equations of hamiltonian dynamics. They indicate clearly that $H_{\alpha\beta}$ should be interpreted as an interaction hamiltonian for two well-separated vortex rings. Since $H_{\alpha\beta}$ and H_α are unaffected by translation of frame, the total momentum of the system $\Sigma_\alpha \mathbf{p}_\alpha$, is conserved.

Since H does not involve t explicitly, it is a constant of the motion. It is, in fact, the total hydrodynamic energy of the system, as the following alternative derivation of (34) clearly shows. Let $\mathbf{A}$ be the vector potential of $\mathbf{u}$, for which

$$\mathbf{u} = \nabla \times \mathbf{A}, \quad \nabla \cdot \mathbf{A} = 0, \tag{37, 38}$$

where $\mathbf{n} \times \mathbf{A}$ is continuous at every vortex surface. Since, by (37) and (38),

$$\nabla^2 \mathbf{A} = -\boldsymbol{\omega}, \tag{39}$$

we have, for the infinite fluid,

$$\mathbf{A}(\mathbf{x}, t) = \frac{1}{4\pi} \int \frac{\boldsymbol{\omega}(\mathbf{x}', t)}{|\mathbf{x} - \mathbf{x}'|} dV'. \tag{40}$$

We may use (40) to divide $\mathbf{A}$ into parts, $\mathbf{A}_\alpha$, arising from each ring, $\mathscr{R}_\alpha$, separately. Recalling (37), we can write (3) in the form

$$H = \tfrac{1}{2}\rho \int \mathbf{A} \cdot \boldsymbol{\omega} dV, \tag{41}$$

which gives an expression of the form (33) with

$$H_\alpha = \tfrac{1}{2}\rho \int_{\mathscr{R}_\alpha} \mathbf{A}_\alpha \cdot \boldsymbol{\omega}_\alpha dV_\alpha, \tag{42}$$

$$H_{\alpha\beta} = \tfrac{1}{2}\rho \int_{\mathscr{R}_\beta} \mathbf{A}_\alpha \cdot \boldsymbol{\omega}_\beta dV_\beta = \frac{1}{8\pi} \rho \int_{\mathscr{R}_\alpha} \int_{\mathscr{R}_\beta} \frac{\boldsymbol{\omega}_\alpha(\mathbf{x}_\alpha, t) \cdot \boldsymbol{\omega}_\beta(\mathbf{x}_\beta, t)}{|\mathbf{x}_\alpha - \mathbf{x}_\beta|} dV_\alpha dV_\beta = H_{\beta\alpha}. \tag{43}$$

When the vortices are well-separated, the distribution of $\boldsymbol{\omega}_\alpha$ within $\mathscr{R}_\alpha$ is not sensibly affected by the remaining rings. To the first approximation, therefore, (42) is simply the energy of the isolated vortex as used in §2. To show that, in the same approximation, (43) and (34) are equivalent, we choose points, $\mathbf{x}_\alpha = \mathbf{q}_\alpha$ in $\mathscr{R}_\alpha$ and $\mathbf{x}_\beta = \mathbf{q}_\beta$ in $\mathscr{R}_\beta$, to represent "the" position of these rings, and expand $|\mathbf{x}_\alpha - \mathbf{x}|^{-1}$ as

$$\frac{1}{|\mathbf{x}_\alpha - \mathbf{x}|} = \sum_{n=0}^{\infty} \frac{1}{n!} (\mathbf{x}_\alpha - \mathbf{q}_\alpha)_{i_1} \cdots (\mathbf{x}_\alpha - \mathbf{q}_\alpha)_{i_n} \frac{\partial^n}{\partial q_{\alpha i_1} \cdots \partial q_{\alpha i_n}} \left[\frac{1}{|\mathbf{x} - \mathbf{q}_\alpha|} \right], \tag{44}$$

so that (40) gives

$$A_{\alpha j}(\mathbf{x}, t) = \sum_{n=0}^{\infty} \frac{1}{n!} M^{(n)}_{\alpha j;\, i_1 \ldots i_n} \frac{\partial^n}{\partial q_{\alpha i_1} \cdots \partial q_{\alpha i_n}} \left[\frac{1}{|\mathbf{x} - \mathbf{q}_\alpha|} \right], \tag{45}$$

where the moments of $\boldsymbol{\omega}_\alpha$ are

$$M^{(n)}_{\alpha j;\, i_1 \ldots i_n} = \frac{1}{4\pi} \int_{\mathscr{R}_\alpha} (\mathbf{x}_\alpha - \mathbf{q}_\alpha)_{i_1} \ldots (\mathbf{x}_\alpha - \mathbf{q}_\alpha)_{i_n} \omega_{\alpha_j}(\mathbf{x}_\alpha, t)\, dV_\alpha. \tag{46}$$

On using the fact that $\boldsymbol{\omega} \cdot dS$ vanishes over the surface of each vortex, it is easy to show that

$$M^{(0)}_{\alpha i} = \frac{1}{4\pi} \int_{S_\alpha} (\mathbf{x}_\alpha - \mathbf{q}_\alpha)_i\, \omega_{\alpha j}\, dS_{\alpha j} = 0, \tag{47}$$

$$M^{(1)}_{\alpha i;\, j} = -\frac{1}{8\pi} \varepsilon_{ijk} \int_{\mathscr{R}_\alpha} [(\mathbf{x}_\alpha - \mathbf{q}_\alpha) \times \boldsymbol{\omega}_\alpha]_k\, dV_\alpha = -\frac{1}{4\pi\rho} \varepsilon_{ijk}\, p_{\alpha k}. \tag{48}$$

The first non-vanishing term of (45) is therefore

$$\mathbf{A}_\alpha(\mathbf{x}, t) \simeq \frac{\mathbf{p}_\alpha \times (\mathbf{x} - \mathbf{q}_\alpha)}{4\pi\rho |\mathbf{x} - \mathbf{q}_\alpha|^3}, \tag{49}$$

which is equivalent to (30). Since, by (47), $M^{(1)}_{\alpha i;\, j}$ is independent of the choice of $\mathbf{q}_\alpha$ within $\mathscr{R}_\alpha$, the same is true of (49), to the order warranted. By (43), we now have

$$H_{\alpha\beta} \simeq \frac{1}{8\pi} \mathbf{p}_\alpha \cdot \left[\nabla_\alpha \times \int_{\mathscr{R}_\beta} \frac{\boldsymbol{\omega}_\beta(\mathbf{x}_\beta, t)}{|\mathbf{x}_\beta - \mathbf{q}_\alpha|}\, dV_\beta \right]. \tag{50}$$

To evaluate the integral on the right of (50), we may again expand, writing as for (44)

$$\frac{1}{|\mathbf{x}_\beta - \mathbf{q}_\alpha|} = \sum_{n=0}^{\infty} \frac{1}{n!} (\mathbf{x}_\beta - \mathbf{q}_\beta)_{i_1} \ldots (\mathbf{x}_\beta - \mathbf{q}_\beta)_{i_n} \frac{\partial^n}{\partial q_{\beta i_1} \ldots \partial q_{\beta i_n}} \left[\frac{1}{|\mathbf{q}_\alpha - \mathbf{q}_\beta|} \right]. \tag{51}$$

On substitution into (50), the first term again vanishes, and the second yields, on using (47)

$$H_{\alpha\beta} \simeq \tfrac{1}{2} \varepsilon_{ijk}\, p_{\alpha i}\, M^{(1)}_{\beta k;\, l} \frac{\partial^2}{\partial q_{\alpha j} \partial q_{\beta l}} \left[\frac{1}{|\mathbf{q}_\alpha - \mathbf{q}_\beta|} \right], \tag{52}$$

which, on applying (48), gives (34). The present argument also indicates that the infinite set of moments (46) defined for each ring may be useful generalized momenta for describing their internal degrees of freedom: it is, for example, possible to express $H_{\alpha\beta}$ as an infinite bilinear form in the moments, with coefficients that are derivatives of $|\mathbf{q}_\alpha - \mathbf{q}_\beta|^{-1}$.

It will be seen that from (30) or (49) that, in the present approximation, the flow created by each ring is that of a doublet. The magnitude of the resulting "force" between rings is identical to that existing between similarly situated electrostatic dipoles, but its direction is opposite, the hydrodynamical dipoles attracting, for example, when facing "head to head". It may be noted that an angular momentum integral exists, which shows that $\Sigma_\alpha \mathbf{q}_\alpha \times \mathbf{p}_\alpha$ is conserved.

4. *The distant collision of two rings.* J. J. Thomson examined the small angle scattering of two vortex rings (α and β, say) by assuming that, to leading order, each describes a straight line (L_a or L_β) with a speed appropriate to its undisturbed momentum. He evaluated, for example, $\Delta\mathbf{p}_\alpha$ (the net change in momentum $\mathbf{p}_\alpha$ of the α ring) by

integrating $d\mathbf{p}_\alpha/dt$ along L_α. Although in principle his procedure seems to be correct for sufficiently distant encounters, it appears that his analysis contains one significant technical error: a factor of 3 has been omitted in deriving (71), on p. 48 of his essay, from the previous line. Unfortunately, this mistake persists, and destroys the hamiltonian character of his final results.

By (32), we have

$$\Delta p_{\alpha i} = -\frac{1}{4\pi\rho}\int_{-\infty}^{\infty} p_{\alpha j}\, p_{\beta k}\frac{\partial^3}{\partial q_{\alpha i}\partial q_{\alpha j}\partial q_{\alpha k}}\left[\frac{1}{|\mathbf{q}_\alpha - \mathbf{q}_\beta|}\right] dt. \tag{53}$$

Let $\mathbf{c}$ be the vector of closest approach of the vortices (which is of course not the same as the common perpendicular to L_α and L_β), and let

$$\mathbf{k} = \frac{d}{dt}(\mathbf{q}_\alpha - \mathbf{q}_\beta), \tag{54}$$

be the relative velocity of $\mathscr{R}_\alpha$ and $\mathscr{R}_\beta$. Let $t = 0$ be the moment of closest approach, and substitute

$$\mathbf{q}_\alpha - \mathbf{q}_\beta = \mathbf{c} + \mathbf{k}t \tag{55}$$

in (53). Since

$$\int_{-\infty}^{\infty}\frac{(q_{\alpha i} - q_{\beta i})\,dt}{|\mathbf{q}_\alpha - \mathbf{q}_\beta|^5} = c_i\int_{-\infty}^{\infty}\frac{dt}{(c^2 + k^2 t^2)^{5/2}} = \frac{4c_i}{3kc^4}, \tag{56}$$

$$\int_{-\infty}^{\infty}\frac{(q_{\alpha i} - q_{\beta i})(q_{\alpha j} - q_{\beta j})(q_{\alpha k} - q_{\beta k})\,dt}{|\mathbf{q}_\alpha - \mathbf{q}_\beta|^7}$$

$$= \frac{16}{15kc^6}c_i c_j c_k + \frac{4}{15k^3 c^4}(c_i k_j k_k + c_j k_k k_i + c_k k_i k_j), \tag{57}$$

equation (53) reduces to

$$\Delta p_{\alpha i} = -\frac{p_{\alpha j}\, p_{\beta k}}{\pi\rho k c^4}\left[c_i\left(\delta_{jk} - \frac{k_j k_k}{k^2}\right) + c_j\left(\delta_{ki} - \frac{k_k k_i}{k^2}\right) + c_k\left(\delta_{ij} - \frac{k_i k_j}{k^2}\right) - \frac{4}{c^2}c_i c_j c_k\right]. \tag{58}$$

The corresponding result of J. J. Thompson is given, in component form, in §32 on p. 56 of his essay. His second expression, for the component along $\mathbf{p}_\alpha \times \mathbf{p}_\beta$ is too small by the factor of 3 mentioned earlier.

Acknowledgment. I am grateful to Dr. John Norbury for discussions, and for the specific point referred to in §2. The interest of Professor Donnelly and Dr. Derek Moore was also appreciated. Professor Steve Childress and Dr. A. Soward kindly criticised the paper in draft.

Appendix: the hamiltonian relation for the large circular ring in the standard model. We confirm (1) for the first four terms in the asymptotic expansion in a/R of H, $\mathbf{p}$ and $d\mathbf{q}/dt$. We first derive a result, which is believed to be new, and which refers to the standard model, defined from (7) by

$$f(\bar{\psi}) = k = \text{constant}. \tag{A.1}$$

According to standard theory, (3) can be transformed in two ways (Lamb, 1932, Art. 162)

$$H = -\tfrac{1}{2}\rho \int_{\mathscr{R}} \psi f(\bar{\psi})\, dV, \tag{A.2}$$

$$H = -\rho \int_{\mathscr{R}} f(\bar{\psi})\, \mathbf{x}.\nabla\psi dV. \tag{A.3}$$

Combining these linearly, it follows that, when (A.1) holds,

$$H = -\frac{1}{7}\rho k \int_{\mathscr{R}} \nabla.(\psi\mathbf{x})\, dV = -\frac{1}{7}\rho k \int_{S} \psi(\mathbf{x}.dS). \tag{A.4}$$

But $\bar{\psi}(=\bar{\psi}_s)$ is constant over S, the surface of the ring, and (A.4) therefore gives, by (4),

$$H = \frac{1}{7}\rho k\left[-\bar{\psi}_s \int_S \mathbf{x}.dS + \frac{1}{2}\frac{dq}{dt}\int_S \varpi^2\, \mathbf{x}.dS\right],$$

or

$$H = \frac{5}{7}p\frac{dq}{dt} - \frac{6\pi\rho\kappa}{7}\bar{\psi}_s. \tag{A.5}$$

For a Hill's spherical vortex of radius a, for example,

$$\bar{\psi}_s = 0, \qquad p = 2\pi\rho a^3\, dq/dt, \tag{A.6, A.7}$$

(Lamb, 1932, Art. 165), and the expression

$$H = \frac{10}{7}\pi\rho a^3(dq/dt)^2, \tag{A.8}$$

for the kinetic energy is correctly obtained.

For the large circular ring, Dyson (1893) gives

$$\bar{\psi}_s = -\frac{\kappa R}{8\pi}\left[\left\{3\log\left(\frac{8R}{a}\right) - \frac{31}{4}\right\} + \frac{3}{8}\left(\frac{a}{R}\right)^2\left\{3\log\left(\frac{8R}{a}\right) - \frac{5}{4}\right\} + \ldots\right], \tag{A.9}$$

$$p = \pi\rho\kappa R^2\left[1 + \frac{3}{4}\left(\frac{a}{R}\right)^2 + \ldots\right], \tag{A.10}$$

$$\frac{dq}{dt} = \frac{\kappa}{4\pi R}\left[\left\{\log\left(\frac{8R}{a}\right) - \frac{1}{4}\right\} - \frac{3}{8}\left(\frac{a}{R}\right)^2\left\{\log\left(\frac{8R}{a}\right) - \frac{5}{4}\right\} + \ldots\right]. \tag{A.11}$$

Equations (A.5), and (A.9) to (A.11), yield

$$H = \tfrac{1}{2}\rho\kappa^2 R\left[\left\{\log\left(\frac{8R}{a}\right) - \frac{7}{4}\right\} + \frac{3}{8}\left(\frac{a}{R}\right)^2\log\left(\frac{8R}{a}\right) + \ldots\right]. \tag{A.12}$$

Equations (A.10) to (A.12) agree with results (6.14) to (6.18) given by Fraenkel (1972).

To verify (1) we observe that the displacement $\delta\xi$ considered in §2 must, *inter alia*, preserve the volume $V = 2\pi^2 a^2 R$ of the ring. We therefore write (A.10) and (A.12) as

$$p = \pi\rho\kappa R^2 \left[1 - \frac{3V}{8\pi^2 R^3} + \ldots\right], \tag{A.13}$$

$$H = \tfrac{1}{2}\rho\kappa^2 R \left[\left\{\tfrac{1}{2}\log\left(\frac{128\pi^2 R^3}{V}\right) - \frac{7}{4}\right\} + \frac{3V}{32\pi^2 R^3}\log\left(\frac{128\pi^2 R^3}{V}\right) + \ldots\right]. \tag{A.14}$$

Now, on applying $\delta\xi$, only R will change in (A.13) and (A.14), and therefore

$$\frac{\delta H}{\delta p} = \frac{\delta H/\delta R}{\delta p/\delta R} = \tfrac{1}{2}\rho\kappa^2 \left[\left\{\tfrac{1}{2}\log\left(\frac{128\pi^2 R^3}{V}\right) - \frac{1}{4}\right\} - \frac{3V}{8\pi^2 R^3}\left\{\frac{1}{2}\log\left(\frac{128\pi^2 R^3}{V}\right) - \frac{3}{4}\right\} + \ldots\right] \Big/ \left[2\pi\rho\kappa R\left(1 - \frac{3V}{16\pi R^3}\right) + \ldots\right]$$

which, to the order required, is the right hand side of (A.11).

References

F. W. Dyson, 1893, *Phil. Trans. Roy. Soc. Lond.* A, 185, 43 and 1041.
L. E. Fraenkel, 1970, *Proc. Roy. Soc. Lond.*, A, 316, 29.
L. E. Fraenkel, 1972, *J. Fluid Mech.*, 51, 119.
H. Lamb, 1932, *Hydrodynamics* (Cambridge: University Press (6th edition)).
L. D. Landau and E. M. Lifshitz, 1960, *Mechanics* (Oxford: Pergamon Press).
P. H. Roberts and R. J. Donnelly, 1970, *Physics Letters*, 31A, 137.
J. J. Thomson, 1883 *A treatise on the motion of vortex rings* (London: Macmillan & Co.).

76C05: *Fluid mechanics; Incompressible inviscid fluids; Vorticity flows.*

School of Mathematics,
University of Newcastle upon Tyne.

(*Received on the 23rd of June* 1972.)

Printed by C. F. Hodgson & Son, Ltd., 50 Holloway Road, London N7 4JL.

A Unified Approach to Mean Field Electrodynamics

P. H. ROBERTS and A. M. SOWARD

University of Newcastle upon Tyne, England

(Received 1974 July 2)

Using the first order smoothing approximation and a novel technique (double Fourier transformation and expansion) a number of results, new and old, in the theory of mean field electrodynamics and magnetohydrodynamics are given a systematic and general derivation. They are cast into forms which bring into new prominence the role of the helicity spectrum in induction processes. The situations in which the results may be expected to be accurate are delineated. New expressions are given for the reduction in the mean electromotive force created by the LORENTZ forces acting on the microscale turbulence.

Mit Hilfe der Korrelationsapproximation zweiter Ordnung und einer neuen Technik (doppelte Fouriertransformation und -entwicklung) wird eine Reihe von neuen und alten Ergebnissen in der Theorie der Elektrodynamik gemittelter Felder und der Magnetohydrodynamik durch eine systematische und allgemeine Ableitung dargestellt. Sie werden in Formen gebracht, die die Rolle des Helicity-Spektrums in Induktionsprozessen in neuem Licht erscheinen lassen. Die Situationen, in denen die Ergebnisse exakt gültig sind, werden angegeben. Neue Ausdrücke für die Bestimmung der mittleren elektromotorischen Kraft, die durch die Wirkung der LORENTZ-Kräfte auf die Turbulenz hervorgerufen werden, werden mitgeteilt.

1. Introduction

In a notable paper, STEENBECK, KRAUSE and RÄDLER (1966) observed that constitutive relationships, such as OHMS law, between the means of quantities such as the magnetic field, $\boldsymbol{B}$, electric field, $\boldsymbol{E}$, and electric current, $\boldsymbol{j}$, would generally differ in a turbulent conductor from those familiar in classical electromagnetic theory. They were able to show that the presence of rotation and of gradients in turbulent intensity (and/or density) would introduce helicity (Schraubensinn) into the turbulence. Such helicity, they demonstrated, would create a component of mean current, $\bar{\boldsymbol{j}}$, parallel or antiparallel to the mean magnetic field, $\bar{\boldsymbol{B}}$. A theory of this phenomenon, which they called 'the α-effect', had been adumbrated earlier by PARKER (1955). In a series of papers, STEENBECK and KRAUSE were able to demonstrate convincingly that the α-effect could sustain the mean field in a turbulent conductor, and they constructed such mean field dynamos in very simple (e.g. axisymmetric) form. They also applied their theory to the solar cycle, to the geomagnetic field, and to stellar magnetism, often with striking success. It was also shown that mirror symmetric turbulence would add to the molecular diffusivity, η, a new turbulent diffusivity, η_T. It was surmised, on physical grounds, that η_T would be positive and that the decay of mean field would be enhanced by the turbulence. This attitude was vindicated by KRAUSE and ROBERTS (1973 a, b) in limiting cases (essentially those in which first order smoothing is valid — see below).

In parallel with these developments, RÄDLER (1968a, 1968b, 1969) probed the theory of mean field electrodynamics more deeply, and located new effects and in particular a new source of dynamo activity, the $\boldsymbol{\Omega} \times \bar{\boldsymbol{j}}$ effect; that is, a mean electromotive force perpendicular to the mean current and to the mean local angular velocity, $\boldsymbol{\Omega}$, of the medium. The electromagnetic effect of local straining of the mean flow, $\bar{\boldsymbol{u}}$, was examined by KRAUSE (1967, 1968, 1973). KRAUSE (1969) also applied the methods of mean field electrodynamics to purely hydrodynamic flow and was able to lend support to the popular belief that REYNOLDS stresses created by locally isotropic turbulence would add to the kinematic molecular viscosity, ν, a new turbulent viscosity, ν_T, which is positive in first order smoothing theory (KRAUSE and RÜDIGER, 1974).

Some of these ideas were taken up by MOFFATT (1970a, 1970b, 1972) who concentrated particularly on induction by a random distribution of inertial waves. These waves are individually highly helical and, provided their energy is propagated predominantly upwards or downwards, they define collectively a motion having non-zero mean helicity. MOFFATT's ideas have recently been taken further by SOWARD (1975). While the mean field remains weak, the inertial waves give rise to a generalized α-effect in which the components of mean current are linearly related to those of the mean field:

$$\bar{j}_i = \alpha_{ij} \bar{B}_j \tag{1.1}$$

where the pseudo-tensor α_{ij} is symmetric. If the mean field strengthens through dynamo action, it will alter the microscale motions and therefore α_{ij} also. In other words α_{ij} is a function of $\overline{B}^2$, and MOFFATT and SOWARD were able to compute this function quantitatively for inertial waves. We may regard this work as the beginnings of a new subject: mean field magnetohydrodynamics. We should note, however, that it was predated by several other observations. See for example § 10.5 of KRAUSE and RÄDLER (1971) and § 5.2 of ROBERTS (1971). Some recent studies by RÄDLER (1974), VAINSHTEIN (1972) and RÜDIGER (1973) should also be mentioned.

We may define mean field magnetohydrodynamics as the marriage of the mean field electrodynamics initiated by STEENBECK, KRAUSE and RÄDLER (1966) and the mean field hydrodynamics of turbuent transport coefficients.

In mean field magnetohydrodynamics, we seek to express the mean REYNOLDS and MAXWELL stresses, $-\varrho\overline{u_i u_j}$ and $\overline{B_i B_j}/\mu$ (where ϱ is the density and μ the permeability), and the mean electromotive forces, as functions of $\overline{\boldsymbol{u}}$ and $\overline{\boldsymbol{B}}$. We may then study the evolution of these mean quantities, unencumbered by the complexities of the microscale motions and fields.

When working in the framework of mean field electrodynamics we may, as for the kinematic dynamo problem, legitimately regard $\boldsymbol{u}$ as given. When working in the framework of mean field magnetohydrodynamics however we should strictly, as for the magnetohydrodynamic dynamo problem, provide a theory for $\boldsymbol{u}$. This would raise difficulties more severe than those still unresolved in classical turbulence theory, and it is necessary to limit our objectives in order to make progress. We will adopt the attitude of STEENBECK, KRAUSE and RÄDLER (1966). We will suppose that, to a first approximation, the turbulence is homogeneous, isotropic and mirror-symmetric, the existence of stochastic body forces with the same statistical properties being implicitly or explicitly postulated. This zero-order motion is regarded as being 'given'. We will then suppose that all other effects, such as CORIOLIS, LORENTZ or buoyancy forces are so weak (when measured by dimensionless parameters based on the microscale) that they can be treated as small perturbations of the basic isotropic mirror-symmetric state. With this assumption, progress is possible and interesting effects can be isolated, though admittedly not always in the cases of prime physical import (see § 5).

The plan of presentation is as follows. After reviewing in § 2 some of our mathematical apparatus, we show in § 3 how the principal results of STEENBECK, KRAUSE and RÄDLER (1966) and of RÄDLER (1969) can be quickly recovered by its use. Its further application in § 3 leads to a new discussion of induction by weakly sheared turbulent flow, a situation examined by KRAUSE (1967, 1968, 1973). In § 4 the full magnetohydrodynamic situation is considered, especially the influence of the mean field on the microscale motions, and the concomitant modification of their inductive effect on the mean field. In § 3, some remarks are made about the limit of small correlation times. In § 5, we summarize and discuss our principal findings.

2. Locally homogeneous turbulence

We make some preliminary remarks about notation and mathematical techniques.

In future, we will abandon the use of overbars for averages, and instead use capital letters for mean fields and small letters for fluctuating fields. Thus, we will denote the magnetic field, electric current, electric field, fluid velocity, vorticity, and so forth by $\boldsymbol{B}+\boldsymbol{b}, \boldsymbol{J}+\boldsymbol{j}, \boldsymbol{E}+\boldsymbol{e}, \boldsymbol{U}+\boldsymbol{u}, \boldsymbol{\Omega}+\boldsymbol{\omega}$, etc. When we do require the ensemble average of products, we will use angle brackets, e.g.

$$\langle u_i(\boldsymbol{x}_1, t_1)\, u_j(\boldsymbol{x}_2, t_2)\rangle \tag{2.1}$$

will be the mean of the velocity product between brackets. We will find it convenient to regard quantities such as (2.1) as functions of mean and relative coordinates:

$$\boldsymbol{X} = \tfrac{1}{2}(\boldsymbol{x}_1 + \boldsymbol{x}_2)\,, \qquad \boldsymbol{x} = \boldsymbol{x}_1 - \boldsymbol{x}_2\,, \qquad T = \tfrac{1}{2}(t_1 + t_2)\,, \qquad t = t_1 - t_2\,, \tag{2.2}$$

and write

$$\boldsymbol{\Phi}_{ij}(\boldsymbol{X}, T; \boldsymbol{x}, t) = \langle u_i(\boldsymbol{X} + \tfrac{1}{2}\boldsymbol{x}, T + \tfrac{1}{2}t)\, u_j(\boldsymbol{X} - \tfrac{1}{2}\boldsymbol{x}, T - \tfrac{1}{2}t)\rangle\,, \tag{2.3}$$

and similarly for other variables. Clearly we have

$$\boldsymbol{\Phi}_{ij}(\boldsymbol{X}, T; \boldsymbol{x}, t) = \boldsymbol{\Phi}_{ji}(\boldsymbol{X}, T; -\boldsymbol{x}, -t)\,, \tag{2.4}$$

and we will say that the turbulence is locally steady if

$$\boldsymbol{\Phi}_{ij}(\boldsymbol{X}, T; \boldsymbol{x}, t) = \boldsymbol{\Phi}_{ij}(\boldsymbol{X}, T; \boldsymbol{x}, -t)\,, \tag{2.5}$$

and locally mirror-symmetric if

$$\boldsymbol{\Phi}_{ij}(\boldsymbol{X}, T; \boldsymbol{x}, t) = \boldsymbol{\Phi}_{ij}(\boldsymbol{X}, T; -\boldsymbol{x}, t)\,. \tag{2.6}$$

It will often be convenient to use FOURIER transformation by writing

$$\hat{\Phi}_{ij}(\boldsymbol{X}, T; \boldsymbol{k}, \omega) = (2\pi)^{-4} \iint \Phi_{ij}(\boldsymbol{X}, T; \boldsymbol{x}, t) \exp[-i(\boldsymbol{k}\cdot\boldsymbol{x} - \omega t)]\, \mathrm{d}\boldsymbol{x}\, \mathrm{d}t\,, \tag{2.7}$$

where, by the reality of Φ_{ij}, we have

$$\hat{\Phi}_{ij}(\boldsymbol{X}, T; -\boldsymbol{k}, -\omega) = \hat{\Phi}^{*}_{ij}(\boldsymbol{X}, T; \boldsymbol{k}, \omega) \tag{2.8}$$

the asterisk denoting the conjugate complex function. Also, by (2.4) and (2.8), we have

$$\hat{\Phi}_{ij}(\boldsymbol{X}, T; \boldsymbol{k}, \omega) = \hat{\Phi}^{*}_{ji}(\boldsymbol{X}, T; \boldsymbol{k}, \omega)\,. \tag{2.9}$$

When the turbulence is locally steady and mirror-symmetric, (2.5), (2.6) and (2.9) show that $\hat{\Phi}_{ij}$ is real.

It is sometimes convenient to take FOURIER transforms with respect to $\boldsymbol{X}$ and T also, writing

$$\tilde{\Phi}_{ij}(\boldsymbol{K}, \Omega; \boldsymbol{k}, \omega) = (2\pi)^{-4} \iint \hat{\Phi}_{ij}(\boldsymbol{X}, T; \boldsymbol{k}, \omega) \exp[-i(\boldsymbol{K}\cdot\boldsymbol{X} - \Omega T)]\, \mathrm{d}\boldsymbol{X}\, \mathrm{d}T\,. \tag{2.10}$$

We will not set down the conditions that correspond to (2.7) and (2.8). We will note however that, if

$$u_i(\boldsymbol{x}, t) = \iint \hat{u}_i(\boldsymbol{k}, \omega) \exp[i(\boldsymbol{k}\cdot\boldsymbol{x} - \omega t)]\, \mathrm{d}\boldsymbol{k}\, \mathrm{d}\omega\,, \tag{2.11}$$

then (2.3) and (2.10) give

$$\tilde{\Phi}_{ij}(\boldsymbol{K}, \Omega; \boldsymbol{k}, \omega) = \langle \hat{u}_i(\boldsymbol{k} + \tfrac{1}{2}\boldsymbol{K}, \omega + \tfrac{1}{2}\Omega)\, \hat{u}_j(-\boldsymbol{k} + \tfrac{1}{2}\boldsymbol{K}, -\omega + \tfrac{1}{2}\Omega)\rangle\,. \tag{2.12}$$

We may obtain the mean square velocity, $V^2(\boldsymbol{X}, T)$, by using the result

$$V^2(\boldsymbol{X}, T) = \langle \boldsymbol{u}\cdot\boldsymbol{u}\rangle = \Phi_{ii}(\boldsymbol{X}, T; 0, 0) = \iint \hat{\Phi}_{ii}(\boldsymbol{X}, T; \boldsymbol{k}, \omega)\, \mathrm{d}\boldsymbol{k}\, \mathrm{d}\omega\,. \tag{2.13}$$

A quantity that recurs below with some frequency is

$$H(\boldsymbol{X}, T; \boldsymbol{x}, t) = \tfrac{1}{2}\langle u_i(\boldsymbol{X} + \tfrac{1}{2}\boldsymbol{x}, T + \tfrac{1}{2}t)\, \omega_i(\boldsymbol{X} - \tfrac{1}{2}\boldsymbol{x}, T - \tfrac{1}{2}t) + \omega_i(\boldsymbol{X} + \tfrac{1}{2}\boldsymbol{x}, T + \tfrac{1}{2}t)\, u_i(\boldsymbol{X} - \tfrac{1}{2}\boldsymbol{x}, T - \tfrac{1}{2}t)\rangle\,, \tag{2.14}$$

where $\boldsymbol{\omega}$ (= curl $\boldsymbol{u}$) is the vorticity of the small scale motion. We may derive the helicity, $h(\boldsymbol{X}, T)$ of the motion from (2.14) by using the result

$$h(\boldsymbol{X}, T) = \langle \boldsymbol{u}\cdot\boldsymbol{\omega}\rangle = H(\boldsymbol{X}, T; 0, 0) = \iint \hat{H}(\boldsymbol{X}, T; \boldsymbol{k}, \omega)\, \mathrm{d}\boldsymbol{k}\, \mathrm{d}\omega\,. \tag{2.15}$$

The helicity spectrum, $\hat{H}$, provides the key contribution to the mean electromotive force, making regeneration of the magnetic field possible. It is immediately apparent from their definitions that $\hat{H}$ and $\hat{\Phi}_{ij}$ are related by

$$\hat{H} = ik_i \varepsilon_{ijk} \hat{\Phi}_{jk}\,. \tag{2.16}$$

We will generally be concerned with globally steady, locally homogeneous, turbulence. By 'globally steady' we mean that Φ_{ij} and $\hat{\Phi}_{ij}$ are independent of T, and $\tilde{\Phi}_{ij}$ is proportional to $\delta(\Omega)$. In reality, we are mainly interested in situations in which the mean magnetic field, $\boldsymbol{B}$, varies over a long time scale, T_B (say), which may be comparable with the time scale, T_V (say), over which Φ_{ij} varies. Provided both T_B and T_V are large compared with the correlation time, τ, we are justified, at least to the order of accuracy considered here, in assuming that Φ_{ij} and $\boldsymbol{B}$ are steady. We would, however, lose many results of interest were we to suppose that turbulence were globally homogeneous, that is if we postulated that Φ_{ij} and $\hat{\Phi}_{ij}$ were independent of $\boldsymbol{X}$ and $\tilde{\Phi}_{ij}$ was proportional to $\delta(\boldsymbol{K})$. Instead we will consider locally homogeneous turbulence, by which we mean that variations in $\boldsymbol{X}$ occur over a macroscale, of dimension L_V, which is large compared with the correlation length, λ, over which the microscale variables $\boldsymbol{x}$ change.

The occurrence of two length scales L_V and λ is, of course, of special significance in mean field electrodynamics. It is common in these circumstances to suppose separability, by postulating for example that

$$\Phi_{ii}(\boldsymbol{X}, T; \boldsymbol{x}, t) = V^2(\boldsymbol{X}, T)\, \psi(\boldsymbol{x}, t)\,, \tag{2.17}$$

where $\psi(0, 0) = 1$. We wish to emphasize that we make no such ad hoc hypothesis below. Our technique will be one of expansion of quantities such as $\tilde{\Phi}_{ij}$ in TAYLOR series about $\boldsymbol{K} = 0$, or equivalently of expressing $\hat{\Phi}_{ij}$ or Φ_{ij} in power series in $\partial/\partial X_i$ derivatives, i.e. in powers of the small parameter λ/L_V, or λ/L_B, where L_B is the macroscale of field variation.

The dangers of applying hypotheses such as (2.17) to tensor quantities has been recognized by several writers. For example, suppose that the fluid is incompressible, so that

$$(k_i + \tfrac{1}{2}K_i)\, \tilde{\Phi}_{ij}(\boldsymbol{K}, \Omega; \boldsymbol{k}, \omega) = (k_j - \tfrac{1}{2}K_j)\, \tilde{\Phi}_{ij}(\boldsymbol{K}, \Omega; \boldsymbol{k}, \omega) = 0\,. \tag{2.18}$$

Expanding in powers of $\boldsymbol{K}$ as indicated above, we see that a form for $\tilde{\Phi}_{ij}$ which is isotropic and mirror symmetric to leading order and which satisfies (2.18) is

$$\tilde{\Phi}_{ij}(\boldsymbol{K}, \Omega; \boldsymbol{k}, \omega) = \tilde{\Phi}(\boldsymbol{K}, \Omega; k, \omega) \left[P_{ij}(\boldsymbol{k}) + \tfrac{1}{2} k^{-2} (K_i k_j - K_j k_i)\right], \tag{2.19}$$

where

$$P_{ij}(\boldsymbol{k}) = \delta_{ij} - k_i k_j / k^2 . \tag{2.20}$$

On performing one pair of FOURIER inversions, we obtain

$$\hat{\boldsymbol{\Phi}}_{ij}(\boldsymbol{X}, T; \boldsymbol{k}, \omega) = \hat{\Phi} P_{ij}(\boldsymbol{k}) + \frac{i}{2k^2}\left(k_i \frac{\partial \hat{\Phi}}{\partial X_j} - k_j \frac{\partial \hat{\Phi}}{\partial X_i}\right), \tag{2.21}$$

where, since $\hat{\Phi}$ depends on $\boldsymbol{k}$ in magnitude, k, but not direction, (2.13) gives

$$V^2 = 16\pi \int_0^\infty \int_0^\infty \hat{\Phi} k^2 \, \mathrm{d}k \, \mathrm{d}\omega . \tag{2.22}$$

It is clear, after performing the second pair of FOURIER inversions, that a relationship of the form (2.17) will not generally hold beyond the leading order. An expression equivalent to (2.19) has been given by VAINSHTEIN and ZEL'DOVICH (1972, eq. (25)).

Evidently (2.19) is not, even to order $\boldsymbol{K}$, the most general form obeying (2.18). For example, we could add a term proportional to $P_{ij}(\boldsymbol{k})$ at the $O(\boldsymbol{K})$ level; we could however regard this as being absorbed into the leading $O(1)$ term. More significantly, we could include helical terms, such as

$$\hat{\Phi}^H_{ij} = -\frac{i}{2k^2} \hat{H} \varepsilon_{ijk} k_k + \frac{1}{4k^4} (\varepsilon_{ikp} k_j + \varepsilon_{jkp} k_i) \, k_k \frac{\partial \hat{H}}{\partial X_p}, \tag{2.23}$$

where the helicity spectrum, $\hat{H}(\boldsymbol{X}, T; \boldsymbol{k}, \omega)$, is here isotropic, i.e. dependent on the magnitude, k, of $\boldsymbol{k}$ only. Further arbitrariness will arise, of course, if we replace the assumption of local isotropy by the less restrictive one of local homogeneity. Except for these reservations, (2.21) does provide the most general locally mirror-symmetric, locally isotropic, turbulence for an incompressible fluid, to the order shown.

On one occasion below (§ 3), we will find it easy to recognize compressibility in the fluid, by adding the term $\hat{C}\delta_{ij}$ to the leading order part of (2.19), and adjusting the $O(\boldsymbol{K})$ terms appropriately. Reinterpretation of the root mean square velocity is also required. We will usually assume that the mean density, ϱ, is a constant, independent of $\boldsymbol{X}$ and T.

Returning to the incompressible fluid, we note that BOCHNER's theorem has proved to be useful in the mean field electrodynamics of homogeneous turbulence (KRAUSE and ROBERTS, 1973 a, b). We will postulate that BOCHNER's theorem continues to be valid in a locally homogeneous flow, and in particular that

$$\hat{\Phi}(\boldsymbol{X}, T; \boldsymbol{k}, \omega) \geqq 0 . \tag{2.24}$$

Sometimes authors assume empirical forms of $\hat{\Phi}$ obeying

$$\hat{\Phi}(\boldsymbol{X}, T; \boldsymbol{k}, \omega) \leqq \hat{\Phi}(\boldsymbol{X}, T; \boldsymbol{k}, 0) . \tag{2.25}$$

Although our analysis depends in no way on (2.25), we find on one occasion (§ 4) that we can determine the sign of a diffusion cofficient (η_B) by using (2.25).

Another point of a technical nature should be mentioned. We are to be governed by the equations of magnetohydrodynamics which (together with the relevant boundary conditions) can only determine the evolution of field and flow uniquely when the initial state is properly defined: it is strictly not correct to attempt solutions starting from $t = -\infty$. In making FOURIER transformations in time, we should therefore explicitly recognize that the character of the problem is of initial value type. We will therefore suppose that the fields and flow are known at $t = 0$, and are the same in every realization of the ensemble. We will hope that if t is sufficiently large the system will lose all "memory" of this initial state. We may add that, in addition to permitting a conceptually clearer discussion, this procedure can in principle facilitate the discussion of non-linear interactions (e.g. BENNEY and NEWELL, 1969), a topic we will not enter into here.

To illustrate the effect of this attitude, we will consider induction of microscale field by a turbulent motion of zero mean ($\boldsymbol{U} = 0$) from a *uniform*, but not necessarily constant field $\boldsymbol{B}$. For simplicity we adopt first order smoothing theory, variously called 'the second order correlation approximation' or 'the discard of third order cumulants'. In the present case it means that we will replace the correct induction equation

$$\partial \boldsymbol{b}/\partial t = \boldsymbol{B} \cdot \nabla \boldsymbol{u} + \eta \nabla^2 \boldsymbol{b} + \operatorname{curl} [\boldsymbol{u} \times \boldsymbol{b} - \langle \boldsymbol{u} \times \boldsymbol{b} \rangle] \tag{2.26}$$

by

$$\partial \boldsymbol{b}/\partial t = \boldsymbol{B} \cdot \nabla \boldsymbol{u} + \eta \nabla^2 \boldsymbol{b} . \tag{2.27}$$

The approximation is mathematically a most convenient one, and is frequently adopted in mean field electrodynamics. By comparing terms in (2.26), we see that it should be a good approximation provided the root mean square velocity V is small compared with λ/τ or η/λ or both, where τ is the microscale (correlation) time. (See also ROBERTS and SOWARD, 1975.)

We consider solutions of (2.27) subject to the initial conditions.

$$\boldsymbol{u} = \boldsymbol{b} = 0\,, \quad t < 0; \qquad \boldsymbol{u} = \hat{\boldsymbol{u}}(\boldsymbol{k}, \omega) \exp [i(\boldsymbol{k}\cdot\boldsymbol{x} - \omega t)]\,, \quad t > 0\,. \tag{2.28}$$

The assumption that $\boldsymbol{b} = 0$ at $t = 0$ is equivalent to stating that the initial magnetic field is the same in all realizations of the ensemble, and is therefore by definition identical to the mean field. By (2.27) and (2.28), we have, for $t > 0$,

$$\boldsymbol{b} = \hat{\boldsymbol{b}}(\boldsymbol{k}, \omega) \exp [i(\boldsymbol{k}\cdot\boldsymbol{x} - \omega t)]\,, \tag{2.29}$$

where

$$\hat{\boldsymbol{b}}(\boldsymbol{k}, \omega) = i(\boldsymbol{k}\cdot\boldsymbol{B})\, \Delta(\omega + i\eta k^2)\, \hat{\boldsymbol{u}}(\boldsymbol{k}, \omega)\,, \tag{2.30}$$

and

$$\Delta(\omega) = \{\exp [i\omega t] - 1\}/i\omega\,. \tag{2.31}$$

Solution (2.29) is valid only for times small compared with both T_B and T_V, the times over which $\boldsymbol{B}$ and V vary, if at all. Within this restriction we may distinguish two cases:

(a) If $\lambda^2/\eta \ll t \ll (T_B, T_V)$ as $t \to \infty$, then

$$\Delta(\omega + i\eta k^2) \to (-i\omega + \eta k^2)^{-1}\,; \tag{2.32a}$$

(b) if $\lambda^2/\eta \gg (T_B, T_V) \gg t$ as $t \to \infty$, then

$$\Delta(\omega + i\eta k^2) \sim \Delta(\omega) \to \pi\delta(\omega) + i/\omega\,. \tag{2.32b}$$

The limiting forms (2.32) imply that, in application, one is interested in a complete spectrum of motions and not merely one mode, as in (2.28). Then in place of (2.29) and (2.30) it is found that

$$\bar{\boldsymbol{b}}(\boldsymbol{k}, t) = i(\boldsymbol{k}\cdot\boldsymbol{B}) \int_{-\infty}^{\infty} \hat{\boldsymbol{u}}(\boldsymbol{k}, \omega)\, \Delta(\omega + i\eta k^2)\, e^{-i\omega t}\, d\omega\,, \tag{2.33}$$

where $\bar{\boldsymbol{b}}$ denotes the FOURIER inverse of $\hat{\boldsymbol{b}}$ with respect to ω but not $\boldsymbol{k}$. To understand (2.32 b) note that, when η is neglected, the contour of ω-integration (the real axis) may be deformed at the pole $\omega = 0$ into a semicircle of small radius $\bar{\varepsilon}$ in the upper-half plane. In the limit $t \to \infty$, there is no contribution to the integral (2.33) from the integration of $(i\omega)^{-1} \exp(i\omega t)$ along the deformed contour. Only the $-1/i\omega$ of (2.31) contributes to (2.33) in the limit $\bar{\varepsilon} \to 0$, and this gives rise to half the residue at the pole $\omega = 0$, together with the principal part of the integral along the real axis. The former is represented by the $\pi\delta(\omega)$ appearing on the right of (2.32b). It is worth noticing that, in the case of large (but not extremely large) conductivities in which τ is small compared with λ^2/η, the pole of (2.32a) approaches the origin and (2.32b) is effectively restored. Further details concerning the limit $t \to \infty$ are given by BENNEY and NEWELL (1969). It should be observed that, although not explicitly recognized by the notation both $\hat{\boldsymbol{b}}(\boldsymbol{k}, \omega)$ and $\Delta(\omega)$ in (2.30) and (2.31) depend on t. We will, however, usually be most interested in limiting forms, such as (2.32), which are independent of t.

It should be pointed out that the limit (b) can give rise to secularities that invalidate the results obtained by our present approach. For example, the treatment given in § 3 of the electromagnetic effects of CORIOLIS forces becomes meaningless in the limit of very large conductivities. In other contexts however, as for example when η_T is computed by (3.12), no secularities appear, and limit (b) is acceptable.

Two methods by which secularities may be avoided may be mentioned. First, the addition of diffusion can sometimes remove a secularity entirely. For example, in the calculation of CORIOLIS effects mentioned above, kinematic viscosity, ν, can be recognized and the secularitites avoided *provided* $\lambda^2/\nu \ll t \ll (T_B, T_V)$ as $t \to \infty$, so that a result analogous to (2.32a) holds for $\Delta(\omega + i\nu k^2)$. Second, even without this artifice, a reversal of limiting procedures can evade the difficulty. For example, in, the case of CORIOLIS forces, we may compute quantities such as α on the assumption that (a) applies, and subsequently take the limit $\tau \ll \lambda^2/\eta$. As an aftermath of the secularity that would have been encountered had the limits been taken in the reverse order, the expression obtained for α will become infinite as $\eta \to 0$.

Before concluding this section, we wish to record the little known identity

$$k^2 \varepsilon_{ijk} = \varepsilon_{pjk} k_i k_p + \varepsilon_{ipk} k_j k_p + \varepsilon_{ijp} k_k k_p\,, \tag{2.34}$$

which is of great utility in simplifying the analysis below. A useful consequence of (2.34) is

$$\varepsilon_{ijk} P_{i\alpha} P_{j\beta} = \varepsilon_{\alpha\beta\gamma} k_k k_\gamma / k^2\,. \tag{2.35}$$

3. Application to mean field electrodynamics

In this section we wish to illustrate how the apparatus described in § 2 may be used to recover several of the known results of mean field electrodynamics in a way that seems to us more concise and unified than the original derivations.

We first consider induction by a globally steady, locally homogeneous turbulence and a constant, nearly homogeneous, field $\boldsymbol{B}$ again using the first order smoothing approximation. In other words, we generalize the results that followed from (2.27) to the case in which $\hat{\boldsymbol{B}} = \hat{\boldsymbol{B}}(\boldsymbol{K})$ varies so slowly with $\boldsymbol{K}$ that we may represent it by the first two terms of its Taylor expansion. Starting from the induction equation

$$\frac{\partial b_i}{\partial t} = \varepsilon_{ijk} \frac{\partial}{\partial x_j} (\varepsilon_{klm} u_l B_m) + \eta \nabla^2 b_i \,, \tag{3.1}$$

and the initial conditions (2.28) we readily obtain (2.29) but with (2.30) replaced by

$$\hat{b}_i(\boldsymbol{k}, \omega) = \varepsilon_{ijk} \varepsilon_{klm} i k_j \int \hat{u}_l(\boldsymbol{k} - \boldsymbol{K}', \omega)\, \hat{B}_m(\boldsymbol{K}')\, \mathrm{d}\boldsymbol{K}'\, \Lambda^*(\omega + i\eta k^2) \,. \tag{3.2}$$

We will now compute the mean electromotive force, $\langle \boldsymbol{u} \times \boldsymbol{b} \rangle$, induced by the microscale field and motions. We will do this through the quantity

$$\boldsymbol{\mathcal{E}}(\boldsymbol{X}, T; \boldsymbol{x}, t) = \langle \boldsymbol{u}(\boldsymbol{X} + \tfrac{1}{2}\boldsymbol{x},\ T + \tfrac{1}{2}t) \times \boldsymbol{b}(\boldsymbol{X} - \tfrac{1}{2}\boldsymbol{x},\ T - \tfrac{1}{2}t) \rangle \,, \tag{3.3}$$

and will obtain $\langle \boldsymbol{u} \times \boldsymbol{b} \rangle$ through the result

$$\langle \boldsymbol{u} \times \boldsymbol{b} \rangle = \boldsymbol{\mathcal{E}}(\boldsymbol{X}, T; 0, 0) = \iint \hat{\boldsymbol{\mathcal{E}}}(\boldsymbol{X}, T; \boldsymbol{k}, \omega)\, \mathrm{d}\boldsymbol{k}\, \mathrm{d}\omega \,. \tag{3.4}$$

(In our case, $\boldsymbol{\mathcal{E}}$ is independent of T, but we wish to set down these results in general form.) It may be seen from (3.2) that

$$\begin{aligned} \hat{\mathcal{E}}_i = \varepsilon_{i\alpha\beta} \varepsilon_{\beta jk} \varepsilon_{klm} \big[-i(k_j - \tfrac{1}{2}K_j) \big] \Lambda^*\big(\omega - i\eta(\boldsymbol{k} - \tfrac{1}{2}\boldsymbol{K})^2\big) \times \\ \times \int \langle \hat{u}_\alpha(\boldsymbol{k} + \tfrac{1}{2}\boldsymbol{K}, \omega)\, \hat{u}_l(-\boldsymbol{k} + \tfrac{1}{2}\boldsymbol{K} - \boldsymbol{K}', -\omega) \rangle\, B_m(\boldsymbol{K}')\, \mathrm{d}\boldsymbol{K}' \,. \end{aligned} \tag{3.5}$$

Noting that the correlation appearing in (3.5) can be written as

$$\hat{\boldsymbol{\Phi}}_{\alpha l}\Big(\boldsymbol{K} - \boldsymbol{K}'; \boldsymbol{k} + \frac{1}{2}\boldsymbol{K}', \omega\Big) = \hat{\boldsymbol{\Phi}}_{\alpha l}(\boldsymbol{K} - \boldsymbol{K}'; \boldsymbol{k}, \omega) + \frac{1}{2} K'_p \frac{\partial}{\partial k_p} \hat{\boldsymbol{\Phi}}_{\alpha l}(\boldsymbol{K} - \boldsymbol{K}'; \boldsymbol{k}, \omega) + \mathrm{O}(\boldsymbol{K}^2) \,, \tag{3.6}$$

we see, on expanding (3.5) to order $\boldsymbol{K}$ and inverting,

$$\begin{aligned} \hat{\mathcal{E}}_i(\boldsymbol{X}; \boldsymbol{k}, \omega) = (\delta_{ij}\delta_{\alpha k} - \delta_{ik}\delta_{\alpha j})\, \varepsilon_{klm} \Big\{ -ik_j \Lambda^*(\omega - i\eta k^2)\, \hat{\boldsymbol{\Phi}}_{\alpha l} B_m + \\ + \frac{1}{2}\frac{\partial}{\partial k_p} \big[k_j \Lambda^*(\omega - i\eta k^2) \big] \Big(\frac{\partial \hat{\boldsymbol{\Phi}}_{\alpha l}}{\partial X_p} B_m + 2 \hat{\boldsymbol{\Phi}}_{\alpha l} \frac{\partial B_m}{\partial X_p} \Big) - \\ - \frac{1}{2}\frac{\partial}{\partial k_p} \big[k_j \hat{\boldsymbol{\Phi}}_{\alpha l} \Lambda^*(\omega - i\eta k^2) \big] \frac{\partial B_m}{\partial X_p} \Big\} \,, \end{aligned} \tag{3.7}$$

from which $\langle \boldsymbol{u} \times \boldsymbol{b} \rangle$ follows by (3.4).

The significance of the various terms on the right-hand side of (3.7) can be made more transparent by introducing the helicity spectrum, $\hat{H}$, and the large-scale divergence, $\partial \hat{\boldsymbol{\Phi}}_{il}/\partial X_l$, of the energy spectrum tensor. Reduction of (3.7) to the form

$$\begin{aligned} \hat{\mathcal{E}}_i = -\frac{k_i k_m}{k^2} \Lambda^*(\omega - i\eta k^2)\, \hat{H} B_m - \Big[\varepsilon_{mpq} P_{ip} \frac{\partial \hat{\boldsymbol{\Phi}}_{jq}}{\partial X_j} B_m + \varepsilon_{ijm} \hat{\boldsymbol{\Phi}}_{jq} \frac{\partial B_m}{\partial X_j} \Big] \Lambda^*(\omega - i\eta k^2) - \\ - \frac{i}{2k} \frac{\partial}{\partial k_p} \Big[\frac{k_i k_m}{k} \Lambda^*(\omega - i\eta k^2) \Big] \frac{\partial \hat{H}}{\partial X_p} B_m - \frac{ik_m}{k^2} \frac{\partial}{\partial k_p} \big[k_i \Lambda^*(\omega - i\eta k^2) \big] \hat{H} \frac{\partial B_m}{\partial X_p} + \\ + \frac{i}{2} \frac{\partial}{\partial k_p} \Big[\frac{k_i k_m}{k^2} \Lambda^*(\omega - i\eta k^2)\, \hat{H} \Big] \frac{\partial B_m}{\partial X_p} \,, \end{aligned} \tag{3.8}$$

is made possible by using the equation

$$ik^2 \varepsilon_{ijk} \hat{\boldsymbol{\Phi}}_{jk} = k_i \hat{H} - \tfrac{1}{2} \varepsilon_{ink} k_n\, \partial(\hat{\boldsymbol{\Phi}}_{jk} + \hat{\boldsymbol{\Phi}}_{kj})/\partial X_j \,, \tag{3.9}$$

which is itself a consequence of the identity (2.34) and the incompressibility conditions (2.18) in the forms

$$k_j \hat{\boldsymbol{\Phi}}_{ij} = \tfrac{1}{2}\, i\, \partial \hat{\boldsymbol{\Phi}}_{ij}/\partial X_j \,, \qquad k_j \hat{\boldsymbol{\Phi}}_{ij} = -\tfrac{1}{2}\, i\, \partial \hat{\boldsymbol{\Phi}}_{ij}/\partial X_j \,. \tag{3.10}$$

The last term in (3.8) provides no contribution to $\langle \boldsymbol{u} \times \boldsymbol{b} \rangle$.

It may be anticipated that, when helicity is present, the first term of (3.8) dominates $\hat{\mathscr{E}}_i$. In many cases of astrophysical interest, however, the helicity is small and the terms involving $\hat{\Phi}_{iq}$ are significant. It is only in special circumstances, in which the integral of the first term vanishes, that those involving $\hat{H}\partial B_m/\partial X_p$ and $B_m\partial\hat{H}/\partial X_p$ can provide an important part of $\hat{\mathscr{E}}_i$.

We now compute $\langle \boldsymbol{u} \times \boldsymbol{b} \rangle$ from (3.4) and (3.8) in some interesting cases. Suppose that the turbulence is locally-isotropic and mirror-symmetric, so that $\hat{\Phi}_{ij}$ is given by (2.21) and the helicity spectrum $\hat{H}$ is identically zero. On substitution of $\hat{\Phi}_{ij}$ into (3.8) and performing the integration (3.4), we readily obtain

$$\langle \boldsymbol{u} \times \boldsymbol{b} \rangle = -\tfrac{1}{2}\nabla\eta_T \times \boldsymbol{B} - \eta_T \operatorname{curl} \boldsymbol{B}\,, \tag{3.11}$$

where

$$\eta_T = \tfrac{2}{3}\iint \hat{\Phi}\Lambda^*(\omega - i\eta k^2)\,\mathrm{d}\boldsymbol{k}\,\mathrm{d}\omega\,, \tag{3.12}$$

the integrations being over all $\boldsymbol{k}$ and over ω from $-\infty$ to $+\infty$. These results were given by RÄDLER (1968a, eq. 48), who also included variability on the long (T) time-scale. In the low conductivity approximation in which (2.32a) holds, (3.12) may be written as

$$\eta_T = \frac{16\pi\eta}{3}\int_0^\infty\int_0^\infty \frac{k^4\hat{\Phi}(k,\omega)\,\mathrm{d}k\,\mathrm{d}\omega}{\omega^2 + \eta^2k^4}\,, \tag{3.13}$$

which is positive by (2.24): see KRAUSE and ROBERTS (1973, a, b). [The dependence of $\hat{\Phi}$ on $\boldsymbol{X}$ is suppressed in (3.13), and in other similar integrals encountered later.]

When we recall that $\boldsymbol{B}$ is governed by the induction equation

$$\partial\boldsymbol{B}/\partial t = \operatorname{curl}\langle \boldsymbol{u} \times \boldsymbol{b} \rangle + \eta\nabla^2\boldsymbol{B}\,, \tag{3.14}$$

we see that, according to (3.11), the mean field evolves as it would if it were subject to classical electromagnetic theory in a stationary medium, but one whose permeability and conductivity are both reduced by the same factor $[\eta/(\eta_T + \eta)]$; the resulting diffusivity is therefore increased from η to $\eta_T + \eta$.

It is interesting to note that these results can readily be generalized to the compressible case. If we add $C\delta_{ij}$ to the leading order expression for Φ_{ij}, as described in § 2, and compute from (3.7) the necessary modifications to $\mathscr{E}_i$, we again recover (3.11) but with (3.12) replaced by

$$\eta_T = \tfrac{1}{3}\iint [2\hat{\Phi} - k\partial\hat{C}/\partial k]\Lambda^*(\omega - i\eta k^2)\,\mathrm{d}\boldsymbol{k}\,\mathrm{d}\omega\,. \tag{3.15}$$

Evidently nothing said so far has aided regeneration of mean field. Our next application touches on mean field hydrodynamics and is essentially a rederivation of the basic result (47) of STEENBECK, KRAUSE and RÄDLER (1966) containing an α-effect. We suppose that the mean fluid velocity, $\boldsymbol{U}$, is non-zero, but that it is so small that its effect on the turbulence can be treated by perturbation methods. We first consider the effect of the rotational part $\partial U_i/\partial X_j - \partial U_j/\partial X_i$ of the rate of strain tensor $\partial U_i/\partial X_j$. We will suppose that this varies so slowly with $\boldsymbol{X}$, compared with the scales over which Φ_{ij}, and B_i vary, that the mean motion is essentially a solid body rotation, $\boldsymbol{\Omega}_*$, of the fluid.[1])

Transforming to the rotating frame, the governing equation for $\boldsymbol{u}$ is

$$\partial\boldsymbol{u}/\partial t + \boldsymbol{u}\cdot\nabla\boldsymbol{u} + 2\boldsymbol{\Omega}_* \times \boldsymbol{u} = -\nabla p + \nu\nabla^2\boldsymbol{u} + \boldsymbol{F}^{(0)}\,, \tag{3.16}$$

where p is the reduced pressure (i.e. pressure divided by density), ν is the kinematic viscosity, and $\boldsymbol{F}^{(0)}$ is the assumed stochastic body force driving the turbulence (see § 1). In the absence of rotation the mean motion would be $\boldsymbol{u}^{(0)}$, say. It would be governed by

$$\partial\boldsymbol{u}^{(0)}/\partial t + \boldsymbol{u}^{(0)}\cdot\nabla\boldsymbol{u}^{(0)} = -\nabla p^{(0)} + \nu\nabla^2\boldsymbol{u}^{(0)} + \boldsymbol{F}^{(0)}\,. \tag{3.17}$$

If the ROSSBY number, $V/\Omega_*\lambda$, defined by the microscale is sufficiently large, we may suppose that, in the presence of rotation,

$$\boldsymbol{u} = \boldsymbol{u}^{(0)} + \boldsymbol{u}^{(1)}\,, \tag{3.18}$$

where $|\boldsymbol{u}^{(1)}| \ll |\boldsymbol{u}^{(0)}|$. By subtracting (3.17) from (3.16) we obtain

$$\partial\boldsymbol{u}^{(1)}/\partial t + \boldsymbol{u}^{(1)}\cdot\nabla\boldsymbol{u}^{(0)} + \boldsymbol{u}^{(0)}\cdot\nabla\boldsymbol{u}^{(1)} + 2\boldsymbol{\Omega}_* \times \boldsymbol{u}^{(0)} = -\nabla p^{(1)} + \nu\nabla^2\boldsymbol{u}^{(1)}\,. \tag{3.19}$$

We have earlier remarked that the validity of our expression for $\mathscr{E}$ rests on first order smoothing theory, and that this is untenable unless V is small compared with λ/τ, η/λ or both. We will now assume that $V \ll \lambda/\tau$, so that the inertial terms on the left of (3.19) may be ignored. We also suppose that the

[1]) The asterisk is added to $\boldsymbol{\Omega}$ because the vorticity of the mean flow is $2\boldsymbol{\Omega}_*$ $\boldsymbol{\Omega}$.

EKMAN number, $\nu/\Omega\lambda^2$, of the microscale is small, so that the viscous term $\nu\nabla^2\boldsymbol{u}^{(1)}$ on the right of (3.19) may be neglected, to give

$$\partial\boldsymbol{u}^{(1)}/\partial t = -\nabla p^{(1)} - 2\boldsymbol{\Omega}_* \times \boldsymbol{u}^{(0)} . \tag{3.20}$$

Using (2.28) as initial condition for $\boldsymbol{u}^{(1)}$, we may solve (3.20) to obtain

$$\hat{u}_i^{(1)} = -2P_{ij}(\boldsymbol{k})\,\varepsilon_{jqm}\Omega_{*q}\hat{u}_m^{(0)}\,\Lambda(\omega) . \tag{3.21}$$

It is useful here to introduce the tensor ψ_{ij}, where

$$\tilde{\psi}_{ij} = \langle \hat{u}_i^{(1)}(\boldsymbol{k} + \tfrac{1}{2}\boldsymbol{K}, \omega)\,\hat{u}_j^{(0)}(-\boldsymbol{k} + \tfrac{1}{2}\boldsymbol{K}, -\omega)\rangle . \tag{3.22}$$

Once this is known, we can easily compute the tensor

$$\hat{\boldsymbol{\Phi}}_{ij}^{(1)} = \hat{\psi}_{ij} + \hat{\psi}_{ij}^* \tag{3.23}$$

which gives the increase, $\hat{\boldsymbol{\Phi}}_{ij}^{(1)}$, produced by CORIOLIS forces in the velocity correlation, $\hat{\boldsymbol{\Phi}}_{ij}^{(0)}$, of the unperturbed motion $\boldsymbol{u}^{(0)}$. According to (3.21), we have

$$\tilde{\psi}_{ij} = -2P_{ip}(\boldsymbol{k} + \tfrac{1}{2}\boldsymbol{K})\,\varepsilon_{pqm}\Omega_{*q}\tilde{\boldsymbol{\Phi}}_{mj}^{(0)}\,\Lambda(\omega) , \tag{3.24}$$

and, expanding to $O(\boldsymbol{K})$ in $\boldsymbol{K}$ and inverting, we obtain

$$\hat{\Psi}_{ij} = -\left[2P_{ip}\varepsilon_{pqm}\Omega_{*q}\hat{\boldsymbol{\Phi}}_{mj}^{(0)} + \frac{i}{k^2}(k_l P_{pl} + k_p P_{il})\,\varepsilon_{pqm}\Omega_{*q}\frac{\partial\hat{\boldsymbol{\Phi}}_{mj}^{(0)}}{\partial X_l}\right]\Lambda(\omega) . \tag{3.25}$$

We may take one of two equivalent attitudes here. We can regard $\boldsymbol{F}^{(0)}$ in (3.17) as being given as (say) a locally-isotropic mirror-symmetric body force, or we may regard the solution $\boldsymbol{u}^{(0)}$ of (3.17) as a known flow with those statistical properties. In either case, we may adopt (2.21), and obtain from (3.23) and (3.25), the helical correlation (2.23)

$$\boldsymbol{\Phi}_{ij}^{(1)} = \boldsymbol{\Phi}_{ij}^{H} , \tag{3.26}$$

where here the helicity spectrum is *non*-isotropic, and is given by

$$\hat{H}(\boldsymbol{X}, T; \boldsymbol{k}, \omega) = 4i(\boldsymbol{k}\cdot\boldsymbol{\Omega}_*)\,\hat{\boldsymbol{\Phi}}^{(0)}[\Lambda(\omega) - \Lambda^*(\omega)] + 2\Omega_{*q}P_{qs}\partial\hat{\boldsymbol{\Phi}}^{(0)}/\partial X_s\,[\Lambda(\omega) + \Lambda^*(\omega)] , \tag{3.27}$$

where, of course, terms involving more than one large spatial derivative have been neglected. Since we are only concerned with the ultimate motion, we may restrict attention to the limit $t \to \infty$, for which

$$\hat{H} \to -\frac{8}{\omega}(\boldsymbol{k}\cdot\boldsymbol{\Omega}_*)\,\hat{\boldsymbol{\Phi}}^{(0)} + 4\pi\delta(\omega)\,\Omega_{*q}P_{qs}\,\partial\hat{\boldsymbol{\Phi}}^{(0)}/\partial X_s . \tag{3.28}$$

It may be seen that the time symmetry of the leading order term of (3.28) is opposite to that of the original turbulence, while that of the correction term is the same.

The addition, $\mathcal{E}^{(1)}$, produced by the CORIOLIS forces, to the electromotive force, $\mathcal{E}^{(0)}$, generated by $\boldsymbol{u}^{(0)}$ may now be computed by substituting (3.26) into (3.8), and using (3.28) and (2.23). It is important to realize, however, that this procedure can be justified only when the conductivity is not "too large". When $\lambda^2/\eta \gg (T_B, T_V)$, the procedure would give rise as $t \to \infty$ to meaningless contributions to $\mathcal{E}$ such as terms proportional to $|\delta(\omega)|^2$. This failure would first require that the time integrations leading to $\mathcal{E}^{(1)}$ be more carefully treated, by returning to (3.2) with $\boldsymbol{u}$ given by (3.21), and adjusting the subsequent argument appropriately. When this is done, however, it is found that, although (3.26) and (3.28) give $\mathcal{E}$ correctly at moderate conductivities, they do not do so when $\lambda^2/\eta \gg (T_B, T_V)$. In that case, a secular term appears in $\mathcal{E}$, which increases linearly with time. This secularity is real enough, and reflects the unphysical nature of the perturbed velocity, and its 'memory' of the artificial initial state (2.28).

Because of these difficulties, we now explicitly exclude the limit (2.32b) of very high conductivities, and use (2.32a) in evaluating the limiting form of (3.8). On substituting (3.26) and (3.28) into (3.8) we find

$$\begin{aligned}\mathcal{E}^{(1)} = {} & [-4(\boldsymbol{\Omega}_*\cdot\nabla f)\,\boldsymbol{B} + (\boldsymbol{B}\cdot\nabla f)\,\boldsymbol{\Omega}_* + (\boldsymbol{\Omega}_*\cdot\boldsymbol{B})\,\nabla f] - [\beta_1\boldsymbol{\Omega}_*\cdot\nabla\boldsymbol{B} + \beta_2\nabla(\boldsymbol{\Omega}_*\cdot\boldsymbol{B})] - \\ & - \tfrac{1}{2}(\boldsymbol{\Omega}_*\cdot\nabla\beta_1)\,\boldsymbol{B} - (\boldsymbol{B}\cdot\nabla\beta_2)\,\boldsymbol{\Omega}_* - \nabla\{\tfrac{1}{2}\beta_1(\boldsymbol{\Omega}_*\cdot\boldsymbol{B})\} ,\end{aligned} \tag{3.29}$$

where

$$f = \frac{16\pi^2}{15\eta}\int_0^\infty \hat{\Phi}(k, 0)\,\mathrm{d}k , \tag{3.30}$$

$$\beta_1 = \frac{128\pi\eta^2}{15} \int_0^\infty \int_0^\infty \frac{k^6 \hat{\Phi}(k, \omega)}{(\omega^2 + \eta^2 k^4)^2} \, \mathrm{d}k \, \mathrm{d}\omega \,, \tag{3.31}$$

$$\beta_2 = -\frac{32\pi}{15} \int_0^\infty \int_0^\infty \frac{k^2(5\omega^2 + 3\eta^2 k^4) \, \hat{\Phi}(k, \omega)}{(\omega^2 + \eta^2 k^4)^2} \, \mathrm{d}k \, \mathrm{d}\omega \,. \tag{3.32}$$

We may note that, by (2.24),

$$f \geqq 0 \,, \qquad \beta_1 \geqq 0 \,, \qquad \beta_2 \leqq 0 \,, \tag{3.33}$$

and that (2.25) gives

$$\beta_1 \leqq 2f \,, \qquad \beta_2 \geqq -4f \,. \tag{3.34}$$

The first three terms in (3.29) between square brackets, are contained in the expression (47) given by STEENBECK, KRAUSE and RÄDLER (1966); the first term of their result does not appear in (3.29) but has already been obtained as the first term on the right-hand side of (3.11) above. The next two terms in square brackets in (3.29) lead to the $\boldsymbol{\Omega}_* \times \boldsymbol{J}$ effect of RÄDLER (1969), as may be seen by writing them in the form

$$\mu\beta_1 \boldsymbol{\Omega}_* \times \boldsymbol{J} + (\boldsymbol{\Omega}_* \cdot \boldsymbol{B}) \nabla(\beta_1 + \beta_2) - \nabla\{(\beta_1 + \beta_2)(\boldsymbol{\Omega}_* \cdot \boldsymbol{B})\} \,. \tag{3.35}$$

The last term of (3.35), like that of (3.29), is a gradient which makes no contribution to the generation of mean field. The second term of (3.35), like the two penultimate terms of (3.29), merely alter the ratios $-4:1:1$ of the three contributions located by STEENBECK, KRAUSE and RÄDLER (1966).

When the correlation time, τ, is small compared with λ^2/η, it is a good approximation to replace $\hat{\Phi}(k, \omega)$ in (3.31), (3.32) and indeed in (3.13), by $\hat{\Phi}(k, 0)$. Then $2\beta_1$ and $-\beta_2$ achieve approximate equality with $4f$ in (3.34) and all terms in (3.29) are of comparable magnitude. In the low conductivity limit $\tau \gg \lambda^2/\eta$ considered by STEENBECK, KRAUSE and RÄDLER (1966), however, $\hat{\Phi}(k, \omega)$ is essentially non-zero only in the region $\omega \ll \eta k^2$, and it follows that β_1 and $|\beta_2|$ are small compared with f. It was for this reason that these authors did not obtain the $\boldsymbol{\Omega}_* \times \boldsymbol{J}$ effect isolated by RÄDLER (1969). It may be shown that the constants β_1 and β_2 of (3.31) and (3.32) coincide with those given in equations (13) and (14) of RÄDLER's paper, bearing in mind that $\Phi_{ij}^{(1)}$ is here given by (3.26) and (3.28). It may, incidentally, be noticed that the scalars h, k and l, which he located by invariance arguments and which appear in his expression (10) for the velocity correlation, are given in the present theory by

$$\frac{1}{3} \overline{v'^2} h = -\frac{4}{r} \frac{\partial \psi}{\partial r} \,, \qquad \frac{1}{3} \overline{v'^2} k = -4r \frac{\partial}{\partial r} \left(\frac{1}{r} \frac{\partial \psi}{\partial r} \right) , \qquad l = 0 \,, \tag{3.36}$$

where $\partial(\nabla^2\psi)/\partial t$ is our correlation function $\Phi^{(\psi)}$.

A further point should be made. For obvious astrophysical reasons, STEENBECK, KRAUSE and RÄDLER (1966) sought to generalize their findings to stratified fluids, the mean density, ϱ, being supposed slowly varying in space, i.e. $\varrho = \varrho(\boldsymbol{X})$. Since the acoustic time-scales are short compared with those of prime astrophysical import they sought to 'filter-out' sound waves by adopting a simplified form of the continuity equation, replacing the equation $\operatorname{div} \boldsymbol{u} = 0$ by

$$\nabla \cdot (\varrho \boldsymbol{u}) = 0 \,. \tag{3.37}$$

Similarly motivated simplifications have been made in the study of convection in compressible atmospheres, and are generally described as 'anelastic approximations' (GOUGH, 1969). The assumption (3.37) has the particular merit of being easily incorporated into the present framework. For example, it is easy to show that (2.21) must be replaced by

$$\hat{\Phi}_{ij} = \hat{\Phi} P_{ij} + \frac{i}{2\varrho^2 k^2} (k_i \delta_{jp} - k_j \delta_{ip}) \frac{\partial}{\partial X_p} (\varrho^2 \hat{\Phi}) \,, \tag{3.38}$$

and (3.20) by

$$\partial(\varrho \boldsymbol{u}^{(1)})/\partial t = -\nabla p^{(1)} - 2\boldsymbol{\Omega}_* \times (\varrho \boldsymbol{u}^{(\alpha)}) \,, \tag{3.39}$$

where p is here not the reduced pressure but the actual pressure. The analysis given below (3.20) follows a similar course, and relationships such as (3.27), (3.28) and (3.29) are recovered with every gradient, $\partial\hat{\Phi}^{(\alpha)}/\partial X_i$, of $\hat{\Phi}^{(\alpha)}$ replaced by $\varrho^{-2}\partial(\varrho^2\hat{\Phi}^{(\alpha)})/\partial X_i$. In this way, the more general form given by STEENBECK, KRAUSE and RÄDLER (1966) for the first three terms of (3.29) is obtained.

Up to this point we have supposed that the mean flow $\boldsymbol{U}$ is zero or is purely a rotation that can be removed by choice of reference frame. We will now assume that $\boldsymbol{U}$ contains weak straining motions

on the long length scale, i.e. we will suppose that $\partial U_i/\partial X_j$ is non-zero but is so small that its effect on the microscale turbulence, $\boldsymbol{u}^{(0)}$, may be treated as a perturbation. Thus, we again expand $\boldsymbol{u}$ as in (3.18), but now regard $\boldsymbol{u}^{(1)}$ as being created from $\boldsymbol{u}^{(0)}$ by the mean rate of strain. Once more we assume that $\Gamma^{(0)} \ll \lambda/\tau$ and $V^{(0)} \ll \nu/\lambda$ and in the non-rotating frame we then obtain in place of (3.17) and (3.18)

$$\partial \boldsymbol{u}^{(0)}/\partial t + (\boldsymbol{U} + \boldsymbol{u}^{(0)}) \cdot \nabla \boldsymbol{u}^{(0)} = -\nabla p^{(0)} + \nu \nabla^2 \boldsymbol{u}^{(0)} + \boldsymbol{F}^{(0)}, \tag{3.40}$$

$$\partial \boldsymbol{u}^{(1)}/\partial t + \boldsymbol{U} \cdot \nabla \boldsymbol{u}^{(1)} + \boldsymbol{u}^{(0)} \cdot \nabla \boldsymbol{U} = -\nabla p^{(1)}. \tag{3.41}$$

We may neglect the term $\boldsymbol{U} \cdot \nabla \boldsymbol{u}^{(1)}$ in (3.41) since it gives rise to terms quadratic in $\boldsymbol{U}$, that are negligible to the order to which we will work. Thus (3.41) is replaced by

$$\partial \boldsymbol{u}^{(1)}/\partial t = -\nabla p^{(1)} - \boldsymbol{u}^{(0)} \cdot \nabla \boldsymbol{U}, \tag{3.42}$$

an equation which is similar to (3.20) and (3.39) above, and which will be treated similarly.

It may be seen that, while $\boldsymbol{u}^{(0)} \cdot \nabla \boldsymbol{U}$ is contained in (3.42), the related term $\boldsymbol{U} \cdot \nabla \boldsymbol{u}^{(0)}$ appears in the leading order equation (3.40), which defines the turbulence. This procedure allows us to appeal to GALILEAN invariance and to claim that $\boldsymbol{u}^{(0)}$ is isotropic in the frame moving with $\boldsymbol{U}$. For in the new frame the material derivative $\partial/\partial t + (\boldsymbol{U} + \boldsymbol{u}^{(0)}) \cdot \nabla$ in (3.40) reduces to $\partial/\partial t + \boldsymbol{u}^{(0)} \cdot \nabla$. It follows that, in the leading order term $\hat{\Phi} P_{ij}$ of (2.21),

$$\hat{\Phi} = \hat{\Phi}(\boldsymbol{X} - \boldsymbol{U}T; \quad k, \omega - \boldsymbol{k} \cdot \boldsymbol{U}). \tag{3.43}$$

Although, of course, the correction term in (2.21) will require modification, we will find that (3.43) suffices for our purposes, and indeed the DOPPLER shift in ω may be ignored in the order to which we will work. We will not consider variations in $\nabla \boldsymbol{U}$ over the macroscale.

As before, we obtain from (3.42)

$$\hat{u}_i^{(1)} = -P_{ij}(\boldsymbol{k}) (\partial U_j/\partial X_m) \hat{u}_m^{(0)} A(\omega), \tag{3.44}$$

and

$$\hat{\Phi}_{ij}^{(1)} = -\{P_{il}(\boldsymbol{k}) \hat{\Phi}_{mj} A(\omega) + P_{jl}(\boldsymbol{k}) \hat{\Phi}_{mi}^* A^*(\omega)\} \partial U_l/\partial X_m. \tag{3.45}$$

For isotropic turbulence, $\hat{\Phi}_{ij} = P_{ij}(\boldsymbol{k}) \hat{\Phi}$ and (3.45) gives, in the limit $t \to \infty$,

$$\hat{\Phi}_{ij}^{(1)} = -\pi P_{il} P_{jm} \hat{\Phi} \delta(\omega) \left(\frac{\partial U_l}{\partial X_m} + \frac{\partial U_m}{\partial X_l} \right) - \frac{i\hat{\Phi}}{\omega} P_{il} P_{jm} \left(\frac{\partial U_l}{\partial X_m} - \frac{\partial U_m}{\partial X_l} \right). \tag{3.46}$$

Recalling that the vorticity, $\boldsymbol{\Omega} = \operatorname{curl} \boldsymbol{U}$, is twice the local rate of rotation, we see that the last term in (3.46) is half the first term of (3.28). But we have previously assumed that the flow is isotropic in the frame rotating with the local mean angular velocity: we have made no such assumption here. Equation (3.46) shows that the mean REYNOLDS stress tensor is

$$\langle u_i^{(1)} u_j^{(0)} + u_i^{(0)} u_j^{(1)} \rangle = \Phi_{ij}^{(1)}(\boldsymbol{X}, T; 0, 0) = -\nu_T \left(\frac{\partial U_i}{\partial X_j} + \frac{\partial U_j}{\partial X_i} \right), \tag{3.47}$$

where

$$\nu_T = \frac{28\pi^2}{15} \int_0^\infty \hat{\Phi}(k, 0)\, k^2\, \mathrm{d}k \tag{3.48}$$

is the turbulent viscosity, which is positive by (2.24). It is interesting to note that, in the high conductivity limit in which $\tau \gg \lambda^2/\eta$, (3.13) and (3.48) show that $\nu_T \approx 7\eta_T/10$, i.e. are of the same order of magnitude. Our result (3.47) is that given by KRAUSE and RÜDIGER (1974a, b) by a different method in the limiting case $\nu = 0$. A difference between our value of ν_T and theirs has arisen because they have made no assumption about GALILEAN invariance.

We conclude this section with a few remarks about the form taken by the theory when the correlation time, τ, is small compared with all other relevant time scales, such as λ/V, T_B, L_B^2/η, and (particularly) λ^2/η. Approximation (2.32b) now applies and, as explained beneath that equation, we must pay heed to secularities. Since τ is small compared with λ/V, first order smoothing theory is again tenable. It should be emphasized that we need no longer suppose that $\lambda \ll L$, since we do not have to expand our expressions in powers of $\boldsymbol{K}$.

The form of the theory for small correlation times has been considered by a number of authors, and we do not wish to claim significant originality for the following remarks. We wish to point out, however, that two apparently sophisticated techniques have been applied to the situation, both of which eventually lead to the results (3.54) and (3.55) below. First, VAINSHTEIN (1972) has developed a series solution in powers of $V\tau/\lambda$ and, after averaging over the ensemble, has used diagram techniques to resum to all orders the resulting series for the mean fields of interest. Second, LERCHE (1973) has developed a mean field induction theory for isotropic flows using KRAICHNAN's direct interaction

approximation (e.g. KRAICHNAN, 1961), and has specialized the resulting equations to the case of small correlation times.

We will show below how VAINSHTEIN's results can be derived by direct methods that do not involve expansion and re-summation. Concerning LERCHE's application of KRAICHNAN's method, we wish to point out that the evolution of the mean magnetic field depends on that of a GREENS tensor, $G_{ij}(\boldsymbol{x}, t|\boldsymbol{x}', t')$, which according to the direct interaction approximation is governed by a certain non-linear integro-differential equation that is hard to solve in general. In the special case in which $\Phi_{ij} = \Phi_{ij}(\boldsymbol{x})\,\delta(t)$ however, the equation governing the mean magnetic field requires the value of G_{ij} for $t = t'$ only. By the definition of the GREEN's function, this is necessarily $\delta_{ij}\delta(\boldsymbol{x} - \boldsymbol{x}')$, and is completely independent of the method by which G_{ij} is evaluated for $t \neq t'$. It therefore gives the same final result as VAINSHTEIN's method, or the procedure to be described below.

We consider the problem which with this section opened: that of computing $\mathcal{E}$ for a homogeneous turbulence. No secularities arise, and we obtain (3.5) as before, but with $A^*(\omega - i\eta(\boldsymbol{k} + \frac{1}{2}\boldsymbol{K})^2)$ replaced by $A(-\omega)$. Since the turbulence is homogeneous, we may set

$$\langle \hat{u}_\alpha(\boldsymbol{k} + \tfrac{1}{2}\boldsymbol{K}, \omega)\, \hat{u}_l(-\boldsymbol{k} + \tfrac{1}{2}\boldsymbol{K} - \boldsymbol{K}', -\omega)\rangle = \delta(\boldsymbol{K} - \boldsymbol{K}')\, \hat{\Phi}_{\alpha l}(\boldsymbol{k} + \tfrac{1}{2}\boldsymbol{K}', \omega)\,, \tag{3.49}$$

in (3.5), to obtain

$$\tilde{\mathcal{E}}_i = i\varepsilon_{i\alpha\beta}\varepsilon_{\beta jk}\varepsilon_{klm}(k'_j - K_j)\, \hat{\Phi}_{\alpha l}(\boldsymbol{k}', \omega)\, B_m(\boldsymbol{K})\, A(-\omega)\,, \tag{3.50}$$

where $\boldsymbol{k}' = \boldsymbol{k} + \frac{1}{2}\boldsymbol{K}$. Integrating over all $\boldsymbol{k}'$ and ω, and inverting the result with respect to $\boldsymbol{K}$, we obtain the mean electromotive force in the form

$$\langle (\boldsymbol{u} \times \boldsymbol{b})_i\rangle = \varepsilon_{i\alpha\beta}\varepsilon_{\beta jk}\varepsilon_{klm}\{(\partial B_m/\partial X_j) \iint \hat{\Phi}_{\alpha l}(\boldsymbol{k}, \omega)\, A(-\omega)\, \mathrm{d}\boldsymbol{k}\, \mathrm{d}\omega - B_m \iint ik_j\, \hat{\Phi}_{\alpha l}(\boldsymbol{k}, \omega)\, A(-\omega)\, \mathrm{d}\boldsymbol{k}\, \mathrm{d}\omega\}\,, \tag{3.51}$$

a result which does not depend on expansion in $\boldsymbol{K}$ and is therefore exact.

If the turbulence is isotropic, its mirror-symmetric and helical parts will be given by the leading order parts of (2.21) and (2.23). Using (2.32b), we obtain from (3.51)

$$\langle \boldsymbol{u} \times \boldsymbol{b}\rangle = -\eta_T\, \mathrm{curl}\, \boldsymbol{B} + \alpha \boldsymbol{B}\,, \tag{3.52}$$

where

$$\eta_T = \frac{8\pi^2}{3}\int_0^\infty \hat{\Phi}(k, 0)\, k^2\, \mathrm{d}k\,, \qquad \alpha = -\frac{8\pi^2}{3}\int_0^\infty \hat{H}(k, 0)\, k^2\, \mathrm{d}k\,, \tag{3.53}$$

the first of which is consistent with (3.13), when it is recognized that $\hat{\Phi}(k, \omega)$ can be replaced by $\hat{\Phi}(k, 0)$ in the small correlation time limit.

4. Mean field magnetohydrodynamics

The objective of this section is to explore the back reaction of the perturbation magnetic fields and currents on the turbulent motions and hence on the mean electromotive force. Our method is again essentially one of perturbation, it now being assumed that the effect of the mean field on the microscale turbulence is small. We then follow the procedure adopted in § 3, now linearizing with respect to the LORENTZ forces rather than the CORIOLIS forces. The basic motion, again written as $\boldsymbol{u}^{(0)}$ rather than $\boldsymbol{u}$, gives rise to a field, $\boldsymbol{b}^{(0)}$, whose spectral components are shown in (3.2). The LORENTZ forces, $\boldsymbol{j}^{(0)} \times \boldsymbol{B}$ and $\boldsymbol{J} \times \boldsymbol{b}^{(0)}$, associated with this field generate a new motion, $\boldsymbol{u}^{(1)}$, which is added to $\boldsymbol{u}^{(0)}$, and which therefore modifies the mean electromotive force $\langle \boldsymbol{u} \times \boldsymbol{b}\rangle$.

Ignoring viscous and inertial forces as before, the equation governing $\boldsymbol{u}^{(1)}$ follows by an argument similar to that by which (3.19) was obtained. In the low conductivity limit $A \ll \eta/\lambda$, where $A\,(= B/\sqrt{\mu\varrho})$ is a typical ALFVÉN velocity based on a mean field strength, $|\boldsymbol{b}| \ll B$ and we may adopt first order smoothing in which $\boldsymbol{j}^{(0)} \times \boldsymbol{b}^{(0)}$ is negligible in comparison with $\boldsymbol{j}^{(0)} \times \boldsymbol{B}$ and $\boldsymbol{J} \times \boldsymbol{b}^{(0)}$. We then obtain[2])

$$\frac{\partial u_i^{(1)}}{\partial t} = -\frac{\partial}{\partial x_i}\left[p^{(1)} + \boldsymbol{B}\cdot\boldsymbol{b}^{(0)}\right] + \frac{\partial}{\partial x_j}\left[B_i b_j^{(0)} + B_j b_i^{(0)}\right]\,, \tag{4.1}$$

where

$$\boldsymbol{b}^{(0)}(\boldsymbol{x}, t) = \iint \hat{\boldsymbol{b}}(\boldsymbol{k}, \omega)\, \exp\left[i(\boldsymbol{k}\cdot\boldsymbol{x} - \omega t)\right]\, \mathrm{d}\boldsymbol{k}\, \mathrm{d}\omega\,, \tag{4.2}$$

and $\hat{\boldsymbol{b}} = \hat{\boldsymbol{b}}^{(0)}$ is given by (3.2).

Taking the FOURIER transform of (4.1), we have

$$\hat{u}_i^{(1)} = i\left[P_{ij}k_l + P_{il}k_j\right] \mathrm{e}^{i\omega t} \iint_0^t \hat{B}_j(\boldsymbol{K})\, \hat{b}_l(\boldsymbol{k} - \boldsymbol{K}, \omega)\, \mathrm{e}^{-i\omega t'}\, \mathrm{d}\boldsymbol{K}\, \mathrm{d}t'\,. \tag{4.3}$$

[2]) From here until the end of this section, we define $\boldsymbol{B}$ and $\boldsymbol{b}$, by dividing them by $\sqrt{(\mu\varrho)}$, to be ALFVÉN velocities.

Our first objective is that of calculating the tensor

$$\tilde{\psi}_{ij} \equiv \langle \hat{u}_i^{(1)}(\boldsymbol{k} + \tfrac{1}{2}\boldsymbol{K}, \omega)\, u_j^{(0)}(-\boldsymbol{k} + \tfrac{1}{2}\boldsymbol{K}, -\omega)\rangle\,, \tag{4.4}$$

which by (4.3) is

$$\begin{aligned}\tilde{\psi}_{ij} = &-\varepsilon_{ipq}\varepsilon_{qrs}\{(k_l + \tfrac{1}{2}K_l)\, P_{ik}(\boldsymbol{k} + \tfrac{1}{2}\boldsymbol{K}) + (k_k + \tfrac{1}{2}K_k)\, P_{il}(\boldsymbol{k} + \tfrac{1}{2}\boldsymbol{K})\} \times \\ &\times \int \hat{B}_p(\boldsymbol{K}')\,(k_p + \tfrac{1}{2}K_p - K'_p)\, \Lambda_1(\omega + i\eta\,|\boldsymbol{k} + \tfrac{1}{2}\boldsymbol{K} - \boldsymbol{K}'|^2) \times \\ &\times \int \tilde{\Phi}_{rj}^{(0)}(\boldsymbol{K} - \boldsymbol{K}' - \boldsymbol{K}''; \boldsymbol{k} - \tfrac{1}{2}\boldsymbol{K}' - \tfrac{1}{2}\boldsymbol{K}'', \omega)\, B_s(\boldsymbol{K}'')\, \mathrm{d}\boldsymbol{K}'\, \mathrm{d}\boldsymbol{K}''\,,\end{aligned} \tag{4.5}$$

where

$$\Lambda_1(\omega + i\eta k^2) \equiv \mathrm{e}^{i\omega t} \int_0^t \Lambda(\omega + i\eta k^2)\, \mathrm{e}^{-i\omega t}\, \mathrm{d}t\,. \tag{4.6}$$

It may be shown that

$$\Lambda_1(\omega + i\eta k^2) \to (\eta k^2 - i\omega)^{-1} \lceil \pi\delta(\omega) + i/\omega \rceil\,, \quad \text{as} \quad t \to \infty\,. \tag{4.7}$$

We expand all terms in (4.5) and retain those linear in $\boldsymbol{K}$, $\boldsymbol{K}'$ and $\boldsymbol{K}''$. For example, we write

$$\begin{aligned}&\tilde{\Phi}_{rj}^{(0)}(\boldsymbol{K} - \boldsymbol{K}' - \boldsymbol{K}''; \boldsymbol{k} - \tfrac{1}{2}\boldsymbol{K}' - \tfrac{1}{2}\boldsymbol{K}'', \omega) = \\ &= \tilde{\Phi}_{rj}^{(0)}(\boldsymbol{K} - \boldsymbol{K}' - \boldsymbol{K}''; \boldsymbol{k}, \omega) - \tfrac{1}{2}(K'_s + K''_s)\, \partial\tilde{\Phi}_{rj}^{(0)}(\boldsymbol{K} - \boldsymbol{K}' - \boldsymbol{K}''; \boldsymbol{k}, \omega)/\partial k_s\,.\end{aligned} \tag{4.8}$$

Further simplifications follow from the incompressibility conditions, in the form (3.10), and after inversion we find that (4.5) becomes

$$\begin{aligned}\hat{\psi}_{ij} = &-\Lambda_1(\boldsymbol{k}\cdot\boldsymbol{B})^2\, \hat{\Phi}_{ij}^{(0)} + \tfrac{1}{2} i \left\{ \frac{\partial}{\partial k_s}[\Lambda_1(\boldsymbol{k}\cdot\boldsymbol{B})^2] \frac{\partial \hat{\Phi}_{ij}^{(0)}}{\partial X_s} - \left[\frac{\partial \hat{\Phi}_{ij}}{\partial k_s} + \frac{k_i}{k^2}\hat{\Phi}_{sj}^{(0)} + \frac{k_j}{k^2}\hat{\Phi}_{is}^{(0)}\right] \frac{\partial}{\partial X_s}[\Lambda_1(\boldsymbol{k}\cdot\boldsymbol{B})^2] - \right. \\ &\left. - \frac{1}{k^2}[k_i\hat{\Phi}_{lj}^{(0)} - k_j\hat{\Phi}_{il}^{(0)}]\, \Lambda_1 \frac{\partial}{\partial X_l}(\boldsymbol{k}\cdot\boldsymbol{B})^2 \right\} + \cdots\,,\end{aligned} \tag{4.9}$$

where $\Lambda_1 = \Lambda_1(\omega + i\eta k^2)$.

Suppose that the basic turbulence is locally isotropic and mirror-symmetric, i. e. is given by (2.21). On substitution into (4.9), we obtain

$$\begin{aligned}\hat{\Psi}_{ij} = &-\Lambda_1(\boldsymbol{k}\cdot\boldsymbol{B})^2\, P_{ij}\hat{\Phi}^{(0)} + \frac{1}{2} i \left\{ P_{ij}\frac{\partial}{\partial k_s}[\Lambda_1(\boldsymbol{k}\cdot\boldsymbol{B})^2] \frac{\partial \hat{\Phi}^{(0)}}{\partial X_s} - \right. \\ &\left. - P_{ij}\frac{\partial \hat{\Phi}^{(0)}}{\partial k_s}\frac{\partial}{\partial X_s}[\Lambda_1(\boldsymbol{k}\cdot\boldsymbol{B})^2] - \frac{1}{k^2}[k_i P_{lj} - k_j P_{il}]\, \hat{\Phi}^{(0)} \frac{\partial}{\partial X_l}[\Lambda_1(\boldsymbol{k}\cdot\boldsymbol{B})^2] \right\} + \cdots\,,\end{aligned} \tag{4.10}$$

where again $\Lambda_1 = \Lambda_1(\omega + i\eta k^2)$. By (2.16), $\hat{H}$ is zero: the magnetic field does not, to this order, create helicity in a turbulence (2.21) which lacks it.

The induced electromotive force may be evaluated from (3.8), to which only the leading order term in (4.10) contributes. The correction $\hat{\mathcal{E}}^{(1)}$ to $\hat{\mathcal{E}}^{(0)}$ is found to be

$$\begin{aligned}\mathcal{E}_i^{(1)} = &\left\{ \varepsilon_{mpq} P_{ip} P_{jq} \frac{\partial}{\partial X_j}[(\boldsymbol{k}\cdot\boldsymbol{B})^2\, \hat{\Phi}^{(0)}]\, B_m + \varepsilon_{iqm} P_{jq}(\boldsymbol{k}\cdot\boldsymbol{B})^2\, \hat{\Phi}^{(0)} \frac{\partial B_m}{\partial X_j} \right\} \times \\ &\times \Lambda^*(\omega - i\eta k^2)\,[\Lambda_1(\omega + i\eta k^2) + \Lambda_1^*(\omega - i\eta k^2)]\,.\end{aligned} \tag{4.11}$$

On integrating this expression with respect to $\boldsymbol{k}$ and ω as in (3.4), we see that the change, $\langle \boldsymbol{u}\times\boldsymbol{b}\rangle^{(1)}$, in $\langle \boldsymbol{u}\times\boldsymbol{b}\rangle$ is

$$\langle \boldsymbol{u}\times\boldsymbol{b}\rangle^{(1)} = \tfrac{1}{2}\nabla\eta_B \times \boldsymbol{B} + \eta_B \boldsymbol{J} - \tfrac{1}{3}\eta_B\,[\boldsymbol{B}\times(\boldsymbol{J}\times\boldsymbol{B})]/B^2\,, \tag{4.12}$$

where

$$\begin{aligned}\eta_B = &\frac{16\pi B^2}{5}\int_0^\infty\int_0^\infty k^4 \hat{\Phi}^{(0)}(k,\omega)\, \Lambda^*(\omega - i\eta k^2)\,[\Lambda_1(\omega + i\eta k^2) + \Lambda_1^*(\omega - i\eta k^2)]\, \mathrm{d}k\, \mathrm{d}\omega = \\ = &\frac{32\pi B^2}{5\eta^2}\int_0^\infty\int_0^\infty \frac{\eta^2 k^4}{\eta^2 k^4 + \omega^2}\left[\frac{\hat{\Phi}^{(0)}(k,0)}{\eta k^2} - \frac{\eta k^2 \hat{\Phi}^{(0)}(k,\omega)}{\eta^2 k^4 + \omega^2}\right] \mathrm{d}k\, \mathrm{d}\omega\,.\end{aligned} \tag{4.13}$$

According to (2.25), η_B is positive. The first two terms of (4.12) should be compared with (3.11). Clearly they have the effect of reducing the turbulent diffusivity η_T given by (3.12) or (3.13) by the amount η_B. The suppression of the turbulent intensity by the back-reaction of the mean field decreases

the rate at which that field diffuses. The last term in (4.12) apparently has the opposite effect. It may be noted that it contains a term proportional to $(\boldsymbol{B}\cdot\boldsymbol{J})\,\boldsymbol{B}$ with a positive coefficient, as predicted by ROBERTS (1971) from a general argument.

It may be noted that the microscale fields contribute to the mean LORENTZ forces. The dominant contribution is, in fact, obtained from $\boldsymbol{b}^{(0)}$ itself. In the simplest theory, of globally homogeneous turbulence in a uniform field, (2.30) holds, and we at once obtain

$$\langle \hat{b}_i \hat{b}_j^* \rangle = (\boldsymbol{k}\cdot\boldsymbol{B})^2\, \Lambda(\omega + i\eta k^2)\, \Lambda^*(\omega - i\eta k^2)\, \hat{\Phi}_{ij}\,, \tag{4.14}$$

so that, in the particular case of isotropic turbulence, we have

$$\langle b_i b_j - \tfrac{1}{2} b^2 \delta_{ij}\rangle = -\,(\eta_T/\eta)\,(B_i B_j + \tfrac{1}{2} B^2 \delta_{ij})\,, \tag{4.15}$$

where η_T is given by (3.13). On using (4.15) to compute the mean energy density (by setting $i = j$), we see that the mean energy density is increased by the microscale fields by the factor $1 + 3\eta_T/\eta$. An expression for $\langle \boldsymbol{b}^2\rangle$ related to (4.15) has been given by BRÄUER and KRAUSE (1973). Since terms involving $B^2\delta_{ij}$ in the mean MAXWELL stress tensor (4.15) can be absorbed into the mean pressure, the only significant parts of (4.15) arise from the B_iB_j term. This effectively reduces the mean MAXWELL stress tensor by the factor $(\eta - \eta_T)/\eta$. Of course, (4.15) has been obtained on the assumption that $\boldsymbol{B}$ is a constant, and therefore it strictly gives rise to zero mean LORENTZ forces. However when $\boldsymbol{B}$ is slowly varying, (4.15) still continues to give the dominant part of the mean LORENTZ force created by the microscale.

To this point we have concentrated on mirror-symmetric flows. Let us suppose now instead that the turbulence has the helical structure (2.23). It is found from (4.9) that the modification $\hat{\Phi}_{ij}^{(1)}$ created by a $\hat{\Phi}_{ij}^{(0)}$ of the form (2.23) is *also* of the form (2.23). The associated helicity spectrum, $\hat{H}^{(1)}$, which must be added to $\hat{H}^{(0)}$ is

$$\begin{aligned}\hat{H}^{(1)} = &-(\boldsymbol{k}\cdot\boldsymbol{B})^2\,[\Lambda_1(\omega + i\eta k^2) + \Lambda_1^*(\omega - i\eta k^2)]\,\hat{H}^{(0)} + \\ &+\frac{1}{2}\,i\left[\frac{\partial}{\partial k_s}\{[\Lambda_1(\omega + i\eta k^2) - \Lambda_1^*(\omega - i\eta k^2)]\,(\boldsymbol{k}\cdot\boldsymbol{B})^2\}\,\frac{\partial \hat{H}^{(0)}}{\partial X_s} - \right. \\ &\left. - k\,\frac{\partial}{\partial k_s}\left(\frac{\hat{H}^{(0)}}{k}\right)\frac{\partial}{\partial X_s}\{[\Lambda_1(\omega + i\eta k^2) - \Lambda_1^*(\omega - i\eta k^2)]\,(\boldsymbol{k}\cdot\boldsymbol{B})^2\}\right].\end{aligned} \tag{4.16}$$

The similarity of this expression with the change $\hat{\Phi}^{(1)}$ in $\hat{\Phi}^{(0)}$ given by (4.10) is of some interest.

We now apply the result (4.16) to find how the α-effect computed by STEENBECK, KRAUSE and RÄDLER (1966) is reduced by the LORENTZ forces. We restrict attention to the low conductivity limit $\tau \gg \lambda^2/\eta$, in which the generation of helicity by the CORIOLIS forces is dominated by the last term of (3.28). To overcome convergence difficulties, we introduce a small kinematic viscosity, ν, so that, for $t \to \infty$, (3.28) is replaced by

$$\hat{H}^{(0)} \to 4\nu k^2 (\omega^2 + \nu^2 k^4)^{-1}\, \Omega_{*q} P_{qs}\, \partial \hat{\Phi}^{(0)}/\partial X_s\,. \tag{4.17}$$

Corresponding modifications to (4.16) show that the helicity associated with the LORENTZ forces is

$$\hat{H}_1^{(1)} = -\,\frac{8\nu k^2(\eta\nu k^4 - \omega^2)}{(\omega^2 + \eta^2 k^4)\,(\omega^2 + \nu^2 k^4)^2}\,(\boldsymbol{k}\cdot\boldsymbol{B})^2\, \Omega_{*q} P_{qs}\, \frac{\partial \hat{\Phi}^{(0)}}{\partial X_s}\,, \tag{4.18}$$

(since, when $\nu \neq 0$, (4.7) is replaced by

$$\Lambda_1(\omega + i\eta k^2) \to (-i\omega + \eta k^2)^{-1}\,(-i\omega + \nu k^2)^{-1}, \qquad t \to \infty\,.) \tag{4.19}$$

We must also include the reduction of the intensity in the prescribed mirror-symmetric turbulence due to the LORENTZ forces. According to (4.10), the change in $\hat{\Phi}^{(0)}$ is

$$-\,\frac{2(\eta\nu k^4 - \omega^2)}{(\omega^2 + \eta^2 k^4)\,(\omega^2 + \nu^2 k^4)}\,(\boldsymbol{k}\cdot\boldsymbol{B})^2\,\hat{\Phi}^{(0)}\,. \tag{4.20}$$

This, through the CORIOLIS forces, leads to a reduction in $\hat{H}$. In fact according to (3.28) we must add to $\hat{H}^{(0)}$

$$\hat{H}_2^{(1)} = -\,\frac{8\nu k^2(\eta\nu k^4 - \omega^2)}{(\omega^2 + \eta^2 k^4)\,(\omega^2 + \nu^2 k^4)^2}\,\Omega_{*q} P_{qs}\,\frac{\partial}{\partial X_s}[(\boldsymbol{k}\cdot\boldsymbol{B})^2\,\hat{\Phi}^{(0)}]\,. \tag{4.21}$$

We now use (3.8) to determine the contribution $\mathscr{E}^{(1)}$ to $\mathscr{E}$. We obtain

$$\hat{\mathscr{E}}_i^{(1)} = \frac{8\nu(\eta\nu k^4 - \omega^2)\,\Omega_{*q}P_{qs}}{(\omega^2 + \eta^2 k^4)\,(\omega^2 + \nu^2 k^4)^2}\left\{2(\boldsymbol{k}\cdot\boldsymbol{B})^2\,\frac{\partial \hat{\Phi}^{(0)}}{\partial X_s} + \hat{\Phi}^{(0)}\,\frac{\partial}{\partial X_s}(\boldsymbol{k}\cdot\boldsymbol{B})^2\right\}\frac{k_i k_m}{i\omega + \eta k^2}\,B_m\,. \tag{4.22}$$

We can now compute the change, $\langle \boldsymbol{u} \times \boldsymbol{b} \rangle^{(1)}$, in the mean electromotive force by integrating (4.22) over all $\boldsymbol{k}$ and ω, and using (3.4). We find

$$\langle \boldsymbol{u} \times \boldsymbol{b} \rangle^{(1)} = B^2 \{4(\boldsymbol{\Omega}_* \cdot \nabla F)\, \boldsymbol{B} - (\boldsymbol{B} \cdot \boldsymbol{\Omega}_*)\, \nabla F - (\boldsymbol{B} \cdot \nabla F)\, \boldsymbol{\Omega}_*\} + 2(\boldsymbol{\Omega}_* \times \boldsymbol{B}) \cdot (\nabla F \times \boldsymbol{B})\, \boldsymbol{B} + \\ + \{3(\boldsymbol{\Omega}_* \cdot \nabla)\, (FB^2 \boldsymbol{B}) - (\boldsymbol{B} \cdot \nabla)\, (F \boldsymbol{B} \cdot \boldsymbol{\Omega}_* \boldsymbol{B}) - \tfrac{1}{2} \nabla (FB^2 \boldsymbol{\Omega}_* \cdot \boldsymbol{B}) - \tfrac{1}{2} \boldsymbol{B} \cdot \nabla (FB^2)\, \boldsymbol{\Omega}_*\}\,, \tag{4.23}$$

where

$$F = \frac{256\pi\nu\eta}{105} \int_0^\infty \int_0^\infty \frac{k^8(\nu\eta k^4 - \omega^2)}{(\omega^2 + \eta^2 k^4)^2\, (\omega^2 + \nu^2 k^4)^2}\, \hat{\Phi}^{(0)}(k, \omega)\, \mathrm{d}k\, \mathrm{d}\omega\,. \tag{4.24}$$

The first group of three terms in (4.23) is almost identical to that discovered by STEENBECK, KRAUSE and RÄDLER (1966); see the first bracket on the right-hand side of (3.29). It is clear that the larger the magnetic field, the greater the modification to their α-effect and, on physical grounds, we would expect that α is, in fact, reduced. Evidently F is unchanged under exchange of ν and η; its value will depend on the magnetic PRANDTL number ν/η. Convergence of the integral (4.24) requires that $\hat{\Phi}^{(0)}(k, \omega)$ is $O(k^2)$ for $k \to 0$. This point will be taken up in the next section.

It is worth recalling that (4.23) and (4.24) have been obtained assuming that

$$(T_B, T_V) \gg \lambda^2/\nu \gg \tau \gg \lambda^2/\eta\,. \tag{4.25}$$

To lowest order the inequalities $\lambda^2/\nu \gg \tau \gg \lambda^2/\eta$ serve to extract the results of STEENBECK, KRAUSE and RÄDLER (1966) from (3.29), so isolating the effects of primary interest with a minimum of algebra. The further restriction $(T_B, T_V) \gg \lambda^2/\nu$ was required to justify (4.19). The distinction between (4.19) and (4.7) is essentially one of the order in which the limits $\nu \to 0$ and $t \to \infty$ are taken. In both cases $\mathcal{E}^{(1)}$, defined by (4.22), tends to infinity; in the case (4.19), $\mathcal{E}^{(1)}$ increases as ν^{-1} as $\nu \to 0$, in the second it is secular in t.

The assumption (4.25) leads to further simplifications, and to lowest order

$$F = \frac{64\pi^2}{105\nu\eta^2} \int_0^\infty \frac{\hat{\Phi}(k, 0)\, \mathrm{d}k}{k^2}\,, \tag{4.26}$$

which is indeed positive by (2.24).

5. Summary and discussion

Many of the results of this paper are not new, but the method used to obtain them is novel, and more general than previous derivations. It provides a unified means of attacking problems of this genre. It is based on a simple idea. Since we believe that quantities such as the mean magnetic field vary slowly spatially, and since small wave-numbers ($\boldsymbol{K}$) correspond to large length-scales ($\boldsymbol{X}$), we suppose that the FOURIER transforms of these quantities are significant for small wave-numbers $\boldsymbol{K}$, only. We therefore adopt the following procedure: (i) transform the governing equations to FOURIER space, solve them formally, and use the solutions so obtained to evaluate formally average quantities of interest (for example, $\mathcal{E}$); (ii) Expand these mean quantities formally in powers of $\boldsymbol{K}$; (iii) Return to physical space, re-expressing any convolution integrals that may have been generated in step (i) back to product form; (iv) Evaluate turbulent transport coefficients that have arisen, such as the turbulent diffusivity, by integration over the relative wave-numbers $\boldsymbol{k}$ and ω, or alternatively seek to put general bounds on these quantities by applying physical reality conditions such as BOCHNER's theorem.

It is perhaps worth recalling that, although the values of turbulent transport coefficients depend crucially on the dependence of two-point two-time correlation function such as (2.1) on the relative coordinates $\boldsymbol{x}$ and t, the ultimate objective is a system of equations for mean fields that involve only the single-point single-time mean coordinates $\boldsymbol{X}$ and T.

Our results (3.11) and (3.29) include expressions of the turbulent permeability, turbulent diffusivity, α-effect and $\boldsymbol{\Omega} \times \boldsymbol{J}$ effect, contained in the fundamental papers of STEENBECK, KRAUSE and RÄDLER (1966) and RÄDLER (1969); they also involve some additional terms which slightly modify their results. At the same time they confer a generality on their results which had not been apparent from such seemingly special choices as equation (16) of STEENBECK, KRAUSE and RÄDLER (1966). In the course of our analysis, we obtained expressions such as (3.26) to (3.28) which succinctly show how the rotation and the gradient of turbulent intensity together excite helicity in an otherwise mirror-symmetric flow.

It is well-known that a general shearing motion in the neighbourhood of a point may be analysed into a local rotation and a pure straining motion. When the mean field is so divided, we find that the

Table 1. Orders of magnitude of transport coefficients
$R_m \equiv V\lambda/\eta$, $Q \equiv V\tau/\lambda$ and $A = B/\sqrt{(\mu\varrho)}$

Quantity	Equation reference	Low Conductivity Limit $\lambda^2 \ll \eta\tau$	High Conductivity Limit $\lambda^2 \gg \eta\tau$
η_T/η	(3.13)	R_m^2	$Q^2(\lambda^2/\eta\tau)$
$(\boldsymbol{\Omega}\cdot\nabla f)\, L/\eta$	(3.30)	$\Omega\tau R_m^2$	$\Omega\tau Q^2(\lambda^2/\eta\tau)$
$(\boldsymbol{\Omega}\cdot\nabla\beta_1)\, L/\eta$	(3.31)	$\Omega\tau R_m^2(\lambda^2/\eta\tau)$	$\Omega\tau Q^2(\lambda^2/\eta\tau)^2$
ν_T/ν	(3.48)	———	$Q^2(\lambda^2/\nu\tau)$ (For $\lambda^2 \gg \nu\tau$)
η_Π/η	(4.13)	$(A\tau/\lambda)^2 R_m^2(\lambda^2/\eta\tau)$	$(A\tau/\lambda)^2 Q^2(\lambda^2/\eta\tau)^3$
$A^2 l^2/f$	(4.24)	$(A\tau/\lambda)^2(\lambda^2/\eta\tau)\,(\lambda^2/\nu\tau)$ (For $\lambda^2 \gg \nu\tau$)	———

rotation is associated with helicity in the local turbulence spectrum, and the straining motion with a mirror-symmetric correction that in the first approximation adds, to the kinematic viscosity diffusing the mean flow, a positive turbulent viscosity (3.48).

Finally, we derived some results in mean field magnetohydrodynamics, first showing how the back-reaction of the induced field on mirror-symmetric flow has the effect of reducing the turbulent permeability and turbulent diffusivity, and of adding a new term (perpendicular to $\boldsymbol{B}$ and $\boldsymbol{J}\times\boldsymbol{B}$) to the mean electromotive force; see (4.12). It does not, however, create helicity in a flow where none existed previously. The back-reaction of the induced fields on a non-mirror-symmetric flow does however, tend to reduce its helicity, an effect which can be computed from (4.16). We illustrated this by, reconsidering the classic calculation of STEENBECK, KRAUSE and RÄDLER (1966) with the back-reaction of induced fields added. A result (4.23) closely similar to their expression for the mean electromotive force was derived; it can be interpreted roughly to mean that α is fractionally reduced by an amount of the order of the square of the HARTMANN number defined by the mean field, i.e. $\lambda^2 B^2/\nu\eta\mu\varrho$. [Compare (3.30) and (4.26).] As a by-product of this computation, we also evaluated the modifications, (4.18) and (4.21), created by the LORENTZ forces to the helicity of the STEENBECK, KRAUSE and RÄDLER process.

We conclude the paper by reviewing the circumstances in which our method can be validly expected to yield precise results. Unfortunately these do not always include the cases of prime astrophysical interest.

It has become clear in the course of our analysis that our expansion procedure can fail at small k or ω unless the spectrum $\hat{\Phi}(\boldsymbol{X}, T; k, \omega)$ tends to zero sufficiently rapidly with k and ω. This difficulty is highlighted by the asymptotic representations of $\Delta(\omega + i\eta k^2)$ given in (2.32) which are evidently singular at $k = \omega = 0$. In the extreme case (2.32b), the delta function $\delta(\omega)$ makes no physical sense except when interpreted as a function having a non-zero (but large) value only in an interval in ω-space whose width is considerably greater than $1/T_B$ and $1/T_V$. In other words, it represents a function whose transform, on the short τ time-scale, resembles a δ-function but which on the large T time-scale does not. The reason for the non-uniformity in the expansion procedure as $\omega \to 0$ is now seen to be due to a blurring of the two-scale concept that must inevitably occur unless the spectrum function, by falling to zero sufficiently rapidly as $\omega \to 0$, allows an unequivocal separation of scales. Similar remarks apply to the non-uniformity that can arise as $k \to 0$.

The non-uniformities just described lead to increasingly singular integrals as our expansion scheme develops. It is for this reason that the integrands defining the turbulent transport coefficients required special consideration in the vicinity of $k = \omega = 0$. In the most extreme case, arising in (4.24), it was necessary to postulate that $\hat{\Phi}^{(0)}(k, \omega) = O(k^2)$ as $k \to 0$ (or alternatively that $\hat{\Phi}^{(0)}(k, \omega) = O(\omega^2)$ as $\omega \to 0$). It is worth noting however that this behaviour for $k \to 0$ has usually been accepted as an inevitable consequence of the dynamical theory of turbulence (BATCHELOR and PROUDMAN, 1956; but see SAFFMAN, 1967).

Finally, we make some estimates of the sizes of the turbulent transport coefficients we have computed, and give ranges in which they may be expected to be reliable. First order smoothing of the magnetic induction equation is justified in *the low conducticity limit* ($\lambda^2/\eta \ll \tau$, denoted by LCL) provided that $R_m (\equiv V\lambda/\eta) \ll 1$; in *the high conductivity limit* ($\lambda^2/\eta \gg \tau$, denoted by HCL) it is valid provided $Q \equiv V\tau/\lambda \ll 1$ and provided the breaking up of the scales of the magnetic field (e.g. VAINSHTEIN and ZEL'DOVICH, 1972) does not take place.

Further restrictions on the parameters are imposed by the dynamical equations. The neglect of CORIOLIS forces in setting up the initial mirror-symmetric flow visualized in § 3 requires that

$$Q = \frac{V\tau}{\lambda} \gg \Omega\tau\,. \tag{5.1}$$

The neglect of LORENTZ forces in that analysis cannot be justified unless

$$R_m \ll 1 , \tag{5.2 LCL}$$

$$R_m \ll Q . \tag{5.2 HCL}$$

First order smoothing in the dynamical equations requires that terms as small as $\boldsymbol{J} \times \boldsymbol{b}'$ are nevertheless large compared with $\boldsymbol{j}' \times \boldsymbol{b}'$ and this requires that

$$R_m \ll \lambda/L , \tag{5.3 LCL}$$

$$Q \ll \lambda/L . \tag{5.3 HCL}$$

There are further restrictions imposed by the condition that perturbation velocities be small compared with the initial turbulence. Applied to the CORIOLIS force this requires that

$$\Omega\tau \ll 1 . \tag{5.4}$$

Applied to the LORENTZ force, it demands

$$R_m Q \ll 1 , \tag{5.5 LCL}$$

$$Q^2 \ll 1 . \tag{5.5 HCL}$$

The parameter range over which all the results of this paper can be validly applied is clearly most restricted. (For examply, $\Omega\tau R_m^2$ is necessarily small, so that the α-effect is unable to maintain a large scale mean field against ohmic decay in the low conductivity limit.) Nevertheless it appears to us that the computations have some validity as tendency calculations that have isolated certain effects which, hopefully, are significant outside the parameter range in which they are strictly correct. It may also be hoped that the positivity of some of the turbulent transport coefficients we have calculated will hold outside the parameter range in which our computed expressions for them are accurate.

References

BATCHELOR, G. K., and PROUDMAN, I.: 1956, Phil. Trans. Roy Soc. Lond. A **248**, 369.
BENNEY, D. J., and NEWELL, A. C.: 1969, Studies in Appl. Math. **48**, 29.
BRÄUER, H., and KRAUSE, F.: 1973, Astron. Nachr. **294**, 179.
GOUGH, D. O.: 1969, J. Atmosph. Sci. **26**, 448.
KRAICHNAN, R. H.: 1961, J. Math. Phys. **2**, 124.
KRAUSE, F.: 1967, Habilitationsschrift, Jena.[1])
KRAUSE, F.: 1968, ZAMM, **48**, 333.
KRAUSE, F.: 1973, ZAMM, **53**, 475.
KRAUSE, F.: 1969, Monats. Dt. Akad. Wiss., Berlin **11**, 188.[1])
KRAUSE, F., and RÄDLER, K.-H.: 1971, in Ergebnisse der Plasmaphysik und der Gaselektronik **2** (Edited by R. ROMPE and M. STEENBECK) Akademie-Verlag. Berlin.
KRAUSE, F., and ROBERTS, P. H.: 1973a, Astrophys. J. **181**, 977.
KRAUSE, F., and ROBERTS, P. H.: 1973b, Mathematika, **20**, 24.
KRAUSE, F., and RÜDIGER, G.: 1974a, Astron. Nachr. **295**, 93.
KRAUSE, F., and RÜDIGER, G.: 1974b, Astron. Nachr. **295**, 185.
LERCHE, I.: 1973, J. Math. Phys. **14**, 1588.
MOFFATT, H. K.: 1970a, J. Fluid Mech. **41**, 435.
MOFFATT, H. K.: 1970b, J. Fluid Mech. **44**, 705.
MOFFATT, H. K.: 1972, J. Fluid Mech. **53**, 385.
MOFFATT, H. K.: 1974, J. Fluid Mech. **65**, 1.
PARKER, E. N.: 1955, Astrophys. J. **122**, 293.
RÄDLER, K.-H.: 1968a, Z. Naturforsch., **23a**, 1841.[1])
RÄDLER, K.-H.: 1968b, Z. Naturforsch., **23a**, 1851.[1])
RÄDLER, K.-H.: 1969, Monats. Dt. Akad. Wiss., Berlin **11**, 194.[1])
RÄDLER, K.-H.: 1974, to appear.
ROBERTS, P. H.: 1971, Lectures in Applied Mathematics **14** (Edited by W. H. REID). Am. Math. Soc. Providence, R. I., U.S.A.
ROBERTS, P. H., and SOWARD, A. M.: 1975, J. Math. Phys. (to appear).
ROBERTS, P. H., and STIX, M.: 1971, Technical Note NCAR/TN/1A-60 from National Center for Atmospheric Research, Boulder, Colorado, U.S.A.
RÜDIGER, G.: 1973, Astron. Nach. **294**, 183.
SAFFMAN, P. G.: 1967, J. Fluid Mech. **27**, 581.
SOWARD, A. M.: 1975, J. Fluid Mech. (to appear).
STEENBECK, M., KRAUSE, F., and RÄDLER, K.-H.: 1966, Z. Naturforsch. **21a**, 369.[1])
VAINSHTEIN, S. I.: 1970, Soviet Physics, J.E.T.P. **31**, 87.
VAINSHTEIN, S. I.: 1972, Soviet Physics, J.E.T.P. **34**, 327.
VAINSHTEIN, S. I., and ZEL'DOVICH, YA. B.: 1972, Soviet Physics, Uspekhi, **15**, 159.

Address of the authors:
P. H. ROBERTS, A. M. SOWARD
University of Newcastle upon Tyne
School of Mathematics
Newcastle upon Tyne
England

[1]) These papers have been translated into English by ROBERTS and STIX (1971).

MAC WAVES

P. H. Roberts*
Department of Physics, University of Oregon
Eugene, Oregon 97403

Magnetohydrodynamics (MHD) is that branch of continuum mechanics which studies the flow of electrically conducting fluid in a magnetic field. Though it finds its greatest applications in astrophysics and planetary physics, most stars and planets are — when measured by the appropriate non-dimensional numbers — highly rotating. And because of the Coriolis force, the MHD of rotating fluids differs drastically from the MHD of non-rotating fluids.

A good example is what happens to the Alfvén wave. In it simplest form, the Alfvén wave is a transverse dispersionless disturbance carried along the lines of force of a uniform field, $\mathbf{B}_0$, permeating an inviscid, perfectly conducting and otherwise stationary fluid of uniform density, ρ_0. Its velocity is $A = B_0/\sqrt{(\mu\rho_0)}$, where μ is the permeability in S.I. units (assumed constant). In a highly rotating fluid, defined as one whose angular velocity, Ω, greatly exceeds $\tau = A/d$ [d being a typical (wave) length], the Alfvén wave disappears. There are then two highly dispersive disturbances: a 'fast wave' with frequency, ω, close to 2Ω and a 'slow wave' with $\omega \approx \tau^2/2\Omega$ (Lehnert, 1954, 1955). The first wave closely resembles an inertial oscillation (e.g., Greenspan, 1968), but inertia has negligible influence on the slow wave, which is therefore sometimes called an MC wave signifying that the only effective forces are the **M**agnetic (or Lorentz) force and the **C**oriolis force (and, of course, the pressure gradient).

It is popularly believed that many astrophysical bodies contain strong internal fields that are hidden from direct observation. It is only through the sunspot pairs that occur when such a toroidal field erupts through the solar surface that its existence can be directly inferred in the Sun. The mantle prevents similar eruptions from the Earth's core to the Earth's surface, but it is often held that the dominant magnetic field of the Earth is not the one we observe at the surface, but is a field wrapped round lines of latitude in the core and having a strength perhaps a hundred times greater, i.e., several hundred Gauss. Even assuming — as we shall do here — that this view is correct, Ω greatly exceeds τ for disturbances of planetary scale d, and the waves will correspondingly be divided as above between the fast and slow families. The fast will be shielded from our observation by mantle conduction, small though it is. The slow are often identified with the observed secular westward drift of the geomagnetic field. Moreover, such waves could in principle be the working components of a geodynamo, so explaining the longevity of the Earth's magnetic field (Braginsky, 1964).

Because of dissipative effects ignored so far in this discussion, the fast and slow waves would ultimately disappear unless fed energy. The most probable energy source is buoyancy, the force which **A**rchimedes is often credited with discovering. This will modify particularly the slow waves, which are then often called 'MAC waves', the addition of A (rather than B for buoyancy) making the acronym pronouncable. These waves ride on the basic steady state

$$\mathbf{B} = s\tau(s,z)\mathbf{1}_\phi, \qquad \mathbf{v} = s\zeta(s,z)\mathbf{1}_\phi, \qquad C = C(s,z), \tag{1}$$

* School of Mathematics, University of Newcastle upon Tyne, Newcastle upon Tyne, England NE1 7RU.

where (s, ϕ, z) are cylindrical polar coordinates with Oz as geographical axis, and $\mathbf{1}_\phi$ is the unit vector in the direction of increasing longitude, ϕ. In (1), $\mathbf{B}$ is magnetic field, $\mathbf{v}$ is fluid velocity and $C = (\rho - \rho_0)/\rho_0$ is the fractional density excess on which the gravitational field, $\mathbf{g}$, acts to give buoyancy.

If C is 'large enough', the state (1) must be unstable. The linear stability analysis of (1) has been elegantly formulated by Braginsky (1967), but only for the case where all dissipative processes are assumed negligible. The first aim of his approach is to determine how great the top-heavy density contrast must be for instability to occur. It is implied that, in the real situation of finite dissipative loss, a slightly larger density difference would maintain waves of steady amplitude, and a secondary aim of the theory is to understand the resulting finite amplitude waves (e.g. Roberts and Soward, 1972, §5). Evidently, the tacit presumption of this approach is that diffusive effects stabilize the system against dynamic (i.e. dissipationless) instabilities.

Dynamic instabilities of (1) occur typically when a modified Rayleigh number,

$$R_1 = g\alpha\beta/\tau^2, \tag{2}$$

reaches a critical value of order unity. Here $\rho_0\alpha\beta$ is the upward density gradient; this could be produced thermally by a downward temperature gradient β, if the material has a coefficient of volume expansion of α. When finite kinematic viscosity (ν), thermal diffusivity (κ) and magnetic diffusivity ($\eta = 1/\mu\sigma$, where σ is electrical conductivity) act, it is found that in the ranges of primary interest $[A^2 = O(\eta\Omega), q \equiv \kappa/\eta \ll 1]$ instability occurs when $R_1 = O(q) \ll 1$. Thus diffusive effects strongly destabilize state (1). The system can be decidedly sub-critical over the short dynamic time-scales during which diffusive effects are negligible, but simultaneously wildly super-critical over the longer dissipative time-scales.

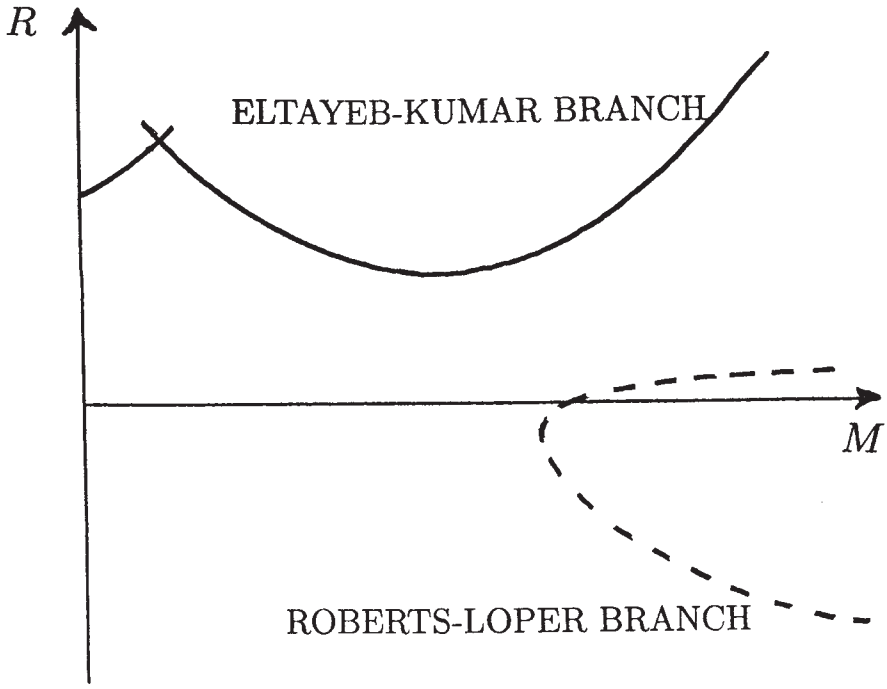
This has been strongly brought out in a recently published study by Eltayeb and Kumar (1977; see also Eltayeb and Kumar, 1974) of a spherical model (1) with $\zeta = 0$, $\tau =$ constant and both $-\mathbf{g}$ and ∇C proportional to the radius vector $\mathbf{r}$. As well as conforming the results just stated, their analysis dramatically demonstrated how magnetic field destabilizes the system. Introducing the usual Rayleigh, Taylor and Hartmann numbers

$$R = \frac{g\alpha\beta d^4}{\nu k}, \qquad T = \left(\frac{2\Omega d^2}{\nu}\right)^2, \qquad M = B_0 d\left(\frac{\sigma}{\rho\nu}\right)^{1/2}, \tag{3}$$

which are dimensionless measures of β, ω and B_0, they showed that, although $R = O(T^{2/3})$ as $T \to \infty$ when magnetic field is absent ($M = 0$), R is only $O(T^{1/2})$ if $M = O(T^{1/4})$. Their stability diagram for fixed large T and $q \ll 1$ (the relevant case for the Earth) is sketched below.

The drastic effects of diffusion have been further emphasized in a recent study by Roberts and Loper (1978). In the case where the spherical boundary ($r = d$) of the Eltayeb-Kumar model is a perfect electrical conductor, Roberts and Loper showed that (1) is unstable *even in the absence of buoyancy*, for all sufficiently small η. Moreover, they show that a top-heavy density distribution is *stabilizing* and a bottom-heavy density distribution is *destabilizing*, contrary to intuition. The analytic tool employed by Roberts and Loper is not strong enough to determine this branch of the stability diagram with precision, but it is presumably as shown dotted in the figure below. Similar in fact is the stability curve located by Soward (1978) for a simpler model, constructed by him in an attempt to elucidate these unexpected discoveries by Roberts and Loper.

Clearly these recent developments in the theory of resistive instabilities of rotating systems evoke fascinating questions for the future, whose answers may well cast light on the still obscure workings of fluid dynamos in geophysics and astrophysics.



REFERENCES
Braginsky, S. I., 1964, *Geomag. & Aeron.* **4**, 698.
Braginsky, S. I., 1967, *Geomag. & Aeron.* **7**, 851.
Eltayeb, I. A. & Kumar, S., 1974, Complete version of Eltayeb & Kumar (1977). Available from British Library, Boston Spa, Yorkshire, U.K., ref: SUP 10021.
Eltayeb, I.A., & Kumar, S., 1977, *Proc. Roy. Soc. Lond. A* **353**, 145.
Greenspan, H.P., 1968, *The Theory of Rotating Fluids*, Cambridge: University Press.
Lehnert, B., 1954, *Astrophys. J.* **119**, 647.
Lehnert, B., 1955, *Astrophys. J.* **121**, 481.
Roberts, P.H. and Loper, D.E., 1978 to be submitted.
Roberts, P.H. and Soward, A.M., 1972, *Ann. Rev. Fl. Mech.* **4**, 117.
Soward, A.M., 1978 to be submitted.

SAACM - Vol. 1, n. 3, 1991

On the motion of a fluid that is incompressible in a generalized sense and its relationship to the Boussinesq Approximation

R.N. HILLS, P.H. ROBERTS

Mathematics Department, Heriot-Watt University, Edinburgh, U.K.
Department of Mathematics, University of California
Los Angeles, CA 90024, U.S.A.

Abstract. *This paper presents a fresh approach to the derivation of the Oberbeck-Boussinesq approximation, an approximation that is widely used in the theory of stratified fluids and thermal convection. Here it is exhibited as a particular case of a new continuum model, the generalized incompressible fluid.*

1. It has long been appreciated that the theories of fluid dynamics describing incompressible materials are generally more likely to be tractable than their compressible counterparts. In the literature two attitudes to incompressibility are common. The first approach is perhaps the more mathematical and stipulates, by definition, that an incompressible material can only undertake isochoric, that is volume preserving, motions. The velocity field $\mathbf{v}$ is thereby constrained to be solenoidal. Then, from the equation of continuity, it follows that $\dot{\rho} = 0$ where ρ is the density and a superposed dot denotes the motional derivative. The pressure, p, that appears in the equation of motion becomes, in effect, a Lagrangian multiplier arising from the constraint $\nabla \cdot \mathbf{v} = 0$. In the second approach the term "incompressible" is more literally interpreted to mean that the density of the fluid cannot be changed by compression. Thus $(\partial\rho/\partial p)_T = (\partial\rho/\partial p)_S = 0$, where T is the temperature and S is the specific entropy. There remains though the possibility of changing ρ by varying T or S and so violating the central isochoric axiom of the first approach. To avoid confusion we shall describe a material in the less restricted case as being "incompressible in the generalized sense" and regard "incompressible" by itself as indicating strictly that no change in volume occurs.

Although the two outlined approaches generally differ, the final results are sometimes identical. To understand why, we must look more closely at the physical reasoning that

underpins the second view. There we are concerned with situations in which changes in field quantities take place so slowly compared with the time taken by sound to cross the system that the velocity of sound, $a = \sqrt{(\partial p/\partial \rho)_S}$ can be assumed infinite, i.e. $(\partial \rho/\partial p)_S = 0$, so that $\dot{\rho} = (\partial \rho/\partial S)_p \dot{S}$. Suppose now that the conduction of heat is so inefficient that the entropy of each fluid element changes negligibly. Then $\dot{S} = 0$, so that $\dot{\rho} = 0$ and hence from the continuity equation, $\nabla \cdot \mathbf{v} = 0$. The pressure appearing in the momentum balance equation then represents the deviation in pressure from the initial state rather than a thermodynamic pressure. In both approaches the energy equation is abandoned since it is decoupled from the remaining equations. If these remaining equations are solved, the temperature field can always be determined, if necessary, by returning to the discarded energy equation.

One of the objectives of the present note is to provide a continuum theory for the second approach, and in this way obtain the continuum theory governing a generalized incompressible fluid. We may reasonably expect that such a theory will be a natural environment in which to discuss the celebrated Oberbeck-Boussinesq (OB-) approximation which undoubtedly is the most widely used model in the theory of stratified fluids and thermal convection. Some regard it as merely a simplifying device that recognises the key physical ingredient (buoyancy) but which can have no strict physical justification. Others more ambitiously have sought to deduce the OB-approximation from the full thermo-mechanical equations and have attempted to define strict limits for its validity [see for example, Jeffreys (1930), Spiegel and Veronis (1960), Veronis (1962), Mihaljan (1962), Roberts (1967), Cordon and Valarde (1975), Velarde and Cordon (1976), Roberts and Stewartson (1977), Spiegel and Weiss (1982)]. This has sometimes led to intricate analyses, and occasionally to obscurities. In this paper we attempt a fresh approach which, we feel, has the merit of clarity. We display the OB-approximation as a particular case of a new continuum model, the generalized incompressible fluid. By taking this model as a starting point, we do not have to consider the limit of large sound speeds that has typified earlier derivations, and we avoid reliance not only on physical illumination, as provided for example by Spiegel and Weiss (1982) but also on a systematic mathematical expansion, as for example that of Roberts and Stewartson (1977). Our analysis constitutes a clarification and correction of the approach adopted in chapter 7 of Roberts (1967).

2. In suffix notation, and with the summation convention, the conservation laws for mass, linear momentum and energy are

$$\dot{\rho} + \rho v_{i,i} = 0, \tag{1}$$

$$\rho \dot{v}_i = \rho F_i + \sigma_{ji,j}, \tag{2}$$

$$\rho Q - \rho\dot{U} - q_{k,k} + \sigma_{ki}d_{ik} = 0\,. \tag{3}$$

Here F_i is the applied body force per unit mass, σ_{ij} the (symmetric) stress tensor, Q the heat supply per unit mass per unit time, U the specific internal energy, q_k the heat flux vector and $d_{ik} = (v_{i,k} + v_{k,i})/2$ the rate of deformation tensor. The entropy production inequality takes the form

$$\rho(T\dot{S} - \dot{U}) + \sigma_{ki}d_{ik} - q_kT_{,k}/T \geq 0\,, \tag{4}$$

where S is the specific entropy and T the temperature.

We are interested in materials whose density can be changed by variations in the temperature, T, but not in the pressure P. This suggests that we formulate the constitutive theory in terms of P and T and then the natural thermodynamic potential is the Gibbs energy

$$G = U - ST + P/\rho. \tag{5}$$

In terms of G the inequality (4) becomes

$$-\rho[\dot{G} + S\dot{T}] + \dot{P} + [\sigma_{ki} + P\delta_{ki}]d_{ik} - q_kT_{,k}/T \geq 0\,. \tag{6}$$

The simplest constitutive fluid model that embraces viscosity and heat conduction, and which eliminates the pressure dependence of ρ, is

$$\begin{aligned} &G = G(T,P)\,, \quad S = S(T,P)\,, \quad \rho = \rho(T)\,,\\ &\sigma_{ij} = -p\delta_{ij} + \lambda d_{uu}\delta_{ij} + 2\mu d_{ij}\,, \quad q_k = -\kappa T_{,k}\,, \end{aligned} \tag{7}$$

where the mechanical pressure, p, and the material coefficients λ, μ, κ are general functions of P and T. The inequality (6) becomes

$$\begin{aligned} -\rho\left(\frac{\partial G}{\partial T} + S\right)\dot{T} - \rho\left(\frac{\partial G}{\partial P} - \frac{1}{\rho}\right)\dot{P} - (p - P)d_{ii}\\ + \lambda d_{uu}d_{ss} + 2\mu d_{ij}d_{ij} + \kappa T_{,i}T_{,i}/T \geq 0\,, \end{aligned} \tag{8}$$

and this must hold for all motions and thermal states that also satisfy the reduced form of (1), *viz.*

$$\alpha\dot{T} = v_{i,i}, \tag{9}$$

where α $(= -\rho^{-1}d\rho/dT)$ is the thermal expansion coefficient of the material. To account for the constraint (9) we introduce the Lagrangian multiplier Γ and then in the usual way we deduce

$$S = -\left(\frac{\partial G}{\partial T} + \frac{\Gamma\alpha}{\rho}\right), \quad \frac{\partial G}{\partial P} = \frac{1}{\rho}, \quad p = P + \Gamma\,, \quad \Gamma = \Gamma(T,P)\,, \tag{10}$$

and

$$\lambda + \frac{2}{3}\mu \geq 0, \quad \mu \geq 0, \quad \kappa \geq 0. \tag{11}$$

If we introduce the Gibbs energy $\hat{G} = G + \Gamma/\rho$ we find from (10) that we may eliminate G, P, Γ in favour of $\hat{G}, p$ and

$$\hat{G} = \hat{G}(T,p) = G_0(T) + \frac{p}{\rho}, \quad \frac{\partial \hat{G}}{\partial T} = -S, \quad \frac{\partial \hat{G}}{\partial p} = \frac{1}{\rho}. \tag{12}$$

In addition, the material parameters are now assumed functions of p and T. Using these results, we find that the governing equations for a generalized incompressible linear viscous fluid are

$$\alpha \dot{T} = v_{i,i}, \tag{13}$$

$$\rho \dot{v}_i = \rho F_i - p_{,i} + [\lambda v_{k,k}\delta_{ij} + 2\mu d_{ij}]_{,j}, \tag{14}$$

$$\rho c_p \dot{T} - \alpha T \dot{p} = \rho Q + (\kappa T_{,i})_{,i} + \lambda d_{uu} d_{vv} + 2\mu d_{ij} d_{ij}, \tag{15}$$

where $c_p = T(\partial S/\partial T)_p$ is the specific heat at constant pressure.

At a boundary of the fluid the usual thermal conditions apply, that is continuity of the temperature and the normal component of the heat flux. At a stationary, non-slip surface we have $\mathbf{v} = 0$ while at a free surface, $\mathscr{S}$, at an ambient pressure π the stress vector is continuous. Using $(7)_4$, we may express this condition by

$$-\pi = -p + \lambda \nabla \cdot \mathbf{v} + 2\mu(\mathbf{n} \cdot \nabla)\mathbf{v}, \quad \mathbf{n} \times \boldsymbol{\omega} = 0, \quad \text{on} \quad \mathscr{S}, \tag{16}$$

where $\mathbf{n}$ is the unit normal to $\mathscr{S}$ and $\boldsymbol{\omega}$ the vorticity vector. Condition (16) must be supplemented by the kinematic condition that

$$\mathbf{n} \cdot \mathbf{v} = U, \quad \text{on} \quad \mathscr{S}, \tag{17}$$

where U is the velocity of $\mathscr{S}$ along its normal $\mathbf{n}$.

3. To discuss the derivation of the OB-approximation we select a simple system: a horizontal layer of fluid contained between a fixed lower boundary $x_3 = L$ and a free top surface $x_3 = f(x_1, x_2, t)$: condition (17) may then be written as

$$v_3 = \dot{f} \quad \text{on} \quad x_3 = f. \tag{18}$$

We also have $\mathbf{F} = \mathbf{g}$ where $\mathbf{g}$, the uniform acceleration due to gravity, is parallel to the x_3-axis. We choose a convenient reference state (T_r, p_r) and will shortly expand the thermodynamic variables and material parameters about that state using the notation $\alpha_r = \alpha(T_r), \kappa_r = \kappa(p_r, T_r)$, etc. The expansion for the density about T_r has the form

$$\rho(T) = \rho_r[1 - \alpha_r(T - T_r) + \ldots]. \tag{19}$$

The central physical idea is that typical accelerations promoted in the fluid by variations in the density are always much less than the acceleration of gravity. As observed by Roberts (1967), the OB-equations result from taking the simultaneous limits, $g \to \infty, \alpha_r \to 0$ but with the restriction that $g\alpha_r$ remains finite. As we shall see this last requirement is essential because otherwise, buoyancy forces are lost. Since $g\alpha_r$ is dimensional, it is more precise to stipulate that the Rayleigh number

$$R = g\alpha_r\beta L^4/\nu_r k_r, \tag{20}$$

is $O(1)$ as the double limit is taken. Here $k = \kappa/\rho c_p$ is the thermal diffusivity, $\nu = \mu/\rho$ is the kinematic viscosity, L is the vertical scale of the system and β is typical of the temperature gradient applied across the system.

Accordingly, we assume that the pressure field has an expression of the form

$$p = p^0 g + p^1 + p^2/g + \ldots, \tag{21}$$

and that for the fields $\mathbf{v}, T - T_r$ we have

$$\chi = \chi^1 + \chi^2/g + \ldots \tag{22}$$

These expansions can equally be regarded as in rising powers of α_r, since $g\alpha_r = O(1)$ by hypothesis. A corresponding expansion for the free surface assumes

$$f = f^2/g + \ldots \tag{23}$$

expressing the fact that in the limit $g \to \infty$, gravity holds a free surface to an equipotential and that for the layer, to leading order, the free boundary becomes planar and horizontal.

Taking the reference state at $x_3 = 0$, the dominant terms of equation (14) give

$$0 = \rho_r\delta_{3i} - p^0_{,i}, \tag{24}$$

so that

$$p^0 g = \rho_r g x_3 + p_r, \tag{25}$$

and to leading order $\dot{p} = \rho_r g v_3^1$. To leading order the condition $(16)_1$ yields $\pi = p_r$, while (18) gives $v_3^1 = 0$ on $x_3 = f$. Combining this with the remaining condition (16) we have

$$v_{1,3}^1 = v_{2,3}^1 = v_3^1 = 0, \quad \text{on} \quad x_3 = 0. \tag{26}$$

The dominant term in (13) gives

$$v_{i,i}^1 = 0, \tag{27}$$

and we determine the leading order terms of the energy equation as

$$(\rho c_p)_r \dot{T}^1 - \rho_r g \alpha_r (T_r + T^1) v_3^1 = \rho_r Q + \kappa_r \nabla^2 T^1 + 2\mu_r d_{ij}^1 d_{ij}^1. \tag{28}$$

The second term on the left-hand side arises from the adiabatic temperature gradient [Jeffreys (1930)]. The next order terms from linear momentum conservation are

$$\rho_r \dot{v}_i^1 = -p_{,i}^1 - \rho_r g \alpha_r T^1 \delta_{i3} + \mu_r v_{i,jj}^1. \tag{29}$$

Equations (27)-(29), together with appropriate conditions [e.g. specified T^1 together with (25) and $\mathbf{v} = 0$ on $x_3 = L$], constitute a closed system of equations determining p^1, T^1 and v_i^1. It contains the OB-approximation as we shall see, but it is more general. We shall call it the "generalized OB-system". The point here is that, because $g \to \infty$ in our limit, (27)-(29) still cling to some vestiges of compressibility. Note particularly that the adiabatic gradient,

$$\beta_{ad} = g(\alpha T / c_p)_r, \tag{30}$$

may be large or small compared with β.

4. To develop the OB-equations, we must assume that the layer is thin, in the sense that $\beta_{ad} \ll \beta$. Equivalently

$$\varepsilon \to 0, \tag{31}$$

where ε, the dissipation parameter, is given by

$$\varepsilon = g\alpha_r L / (c_p)_r. \tag{32}$$

In short, in our double limit $g \to \infty, \alpha_r \to 0$, we must assume that $g\alpha_r \ll (c_p)_r / L$ even though $g\alpha_r = O(\nu_r k_r / T_r L^3)$; see (20). Thus $(c_p)_r \gg \nu_r k_r / T_r L^2$. This point was overlooked by Roberts (1967).

To develop consequences of (31) it is perhaps simplest to introduce dimensionless variables by

$$x_i \to L x_i, \quad t \to (L^2/\nu_r)t, \quad v_i^1 \to (\nu_r/L)v_i, \quad p^1 \to (\rho_r \nu_r^2/L^2)p$$
$$T_r \to \beta L \sigma_r T_r, \quad T^1 \to \beta L \sigma_r T, \quad Q \to [\beta(\nu c_p)_r/L]Q, \tag{33}$$

where $\sigma = \nu/k$ is the Prandtl number. Equation (27) is unchanged while (28) and (29) become

$$\sigma_r[\dot{T} - \varepsilon(T_r + T)v_3^1] = Q + T_{,ii} + 2(\sigma_r \varepsilon/R)d_{ij}d_{ij}, \tag{34}$$

$$\dot{v}_i = -RT\delta_{i3} - p_{,i} + v_{i,jj}. \tag{35}$$

In the limit (31), the last effect of compressibility on the left-hand side of (34) disappears. Simultaneously, the viscous regeneration of heat, the final term in (34), no longer affects the energy budget. Returning to the dimensional form, we recover the Boussinesq system (dropping the superscript 1):

$$\nabla \cdot \mathbf{v} = 0, \tag{36}$$

$$\dot{T} = k\nabla^2 T + q, \tag{37}$$

$$\dot{\mathbf{v}} = -\nabla \varpi - \alpha_r T \mathbf{g} + \nu \nabla^2 \mathbf{v}, \tag{38}$$

where $\varpi = p/\rho$ and $q = Q/c_p$.

ACKNOWLEDGEMENT

The research of one of the authors (PHR) is sponsored by the U.S. Office of Naval Research under contract N00014-86-K-0691 with the University of California, Los Angeles.

REFERENCES

Cordon, R.P., Velarde, M.C., 1975 *On the (non-linear) foundations of Boussinesq approximation applicable to a thin layer of fluid*, Journal of Physique **36**, 591.

Jeffreys, H., 1930 *The instability of a compressible fluid heated below*, Proc. Camb. Phil. Soc. **26**, 170.

Mihaljan, J.M., 1962 *A rigorous exposition of the Boussinesq approximation applicable to a thin layer of fluid*, Astrophys. J. **136**, 1126.

Roberts, P.H., 1967 *An Introduction to Magnetohydrodynamics*, Longmans, London.

Roberts, P.H., Stewartson, K. 1977 *The effect of finite electrical and thermal conductivities on magnetic buoyancy in a rotating gas*, Astron. Nachr. **298**, 311.

Spiegel E.A., Veronis G., 1960 *On the Boussinesq approximation for a compressible fluid*, Astrophys. J. **131**, 442.

Spiegel E.A., Weiss N.O., 1982 *Magnetic buoyancy and the Boussinesq approximation*, Geophys. & Astrophys. Fluid Dyn. **22**, 219.

Velarde M.G., Cordon R.P., 1976 *On the (non-linear) foundations of Boussinesq approximation applicable to a thin layer of fluid (II). Viscous dissipation and large cell gap effects*, Journal de Physique **37**, 177.

Veronis G., 1962 *The magnitude of the dissipation terms in the Boussinesq approximation*, Astrophys. J. **135**, 655.

Manuscript received: August 10, 1990.

Nonlinear Theory of Localized Standing Waves

Bruce Denardo,[1] Andrés Larraza,[1] Seth Putterman,[2] and Paul Roberts[3]
[1]*Department of Physics, Naval Postgraduate School, Monterey, California 93943*
[2]*Department of Physics, University of California, Los Angeles, California 90024*
[3]*Department of Mathematics, University of California, Los Angeles, California 90024*
(Received 20 May 1992)

An investigation of the nonlinear dispersive equations of continuum mechanics reveals localized standing-wave solutions that are domain walls between regions of different wave number. These states can appear even when the dispersion law is a single-valued function of the wave number. In addition, we calculate solutions for kinks in cutoff and noncutoff modes, as well as cutoff breather solitons.

PACS numbers: 43.25.+y, 03.40.Kf

Experiments on continuous media have demonstrated the existence of propagating envelope solitons [1,2], Scott Russell–type solitons [3], as well as standing solitons such as breathers [4] and kinks in cutoff modes [5]. These localized modes are described by equations that, in leading order, are closely allied to the nonlinear Schrödinger (NLS) and Korteweg–de Vries (KdV) equations [6]. The recent observation of domain walls [7] has motivated us to investigate the general theory of modulated standing waves in a continuous medium. We are especially interested in the fact that the domain wall breaks parity as well as translational invariance. We report below the finding that the domain wall and noncutoff kinks are described by a new set of modulational equations (i.e., not the NLS, KdV, Toda, or sine-Gordon equations). Our calculations are limited to nontopological soliton, localized states.

We consider the general class of nonlinear dispersive continuum mechanical systems that can be cast in the form

$$\rho'_{tt}+\hat{\omega}_0^2\rho'=\hat{G}\rho'^3, \tag{1}$$

where $\hat{\omega}_0^2$ and $\hat{G}$ are linear isotropic differential operators in ∇, i.e., are functions of ∇^2; the subscript t denotes differentiation with respect to time. The equations of motion for acoustic and optical phonons, the flexing modes of plates and shells, the continuum limit of the pendulum lattice, as well as many other systems can, to a reasonable approximation, be written in form (1). Our investigations will be limited to one space dimension plus time. Even with this restriction, the results may have applications to information storage and to fiber optic communication [8], and hopefully may also point the way to generalizations in higher dimensions.

We seek solutions to (1) in a form that allows for the modulations of both amplitude and wave number $k=\theta_x$:

$$\rho'(x,t)=\psi(x,t)\cos\theta(x,t)e^{-i\omega t}+\text{c.c.}, \tag{2}$$

where ω is a constant (the frequency of the standing wave being modulated), ψ and θ are complex, and c.c. denotes the complex conjugate of the preceding expression.

To derive approximate modulational equations of motions for θ and ψ, we assume that $\psi=O(\epsilon)$, where $0<\epsilon\ll 1$, and assume that time derivatives are small compared with space derivatives. By substituting (2) into (1) and equating the coefficients of $\sin\theta$ and $\cos\theta$ separately to zero (this is permissible when $k\sim 1$), we obtain

$$2i\omega\psi_t=[\omega_0^2(k)-\omega^2-\tfrac{9}{4}G|\psi|^2]\psi-\tfrac{1}{2}g_k\psi_{xx}-g_{kk}[\tfrac{1}{2}k_x\psi_x+\tfrac{1}{6}k_{xx}\psi]-\tfrac{1}{8}g_{kkk}k_x^2\psi, \tag{3}$$

$$2i\omega\psi\theta_t+g^{1/2}(g^{1/2}\psi)_x=0, \tag{4}$$

where the subscript x stands for differentiation with respect to x, $\omega_0(k)$ is the dispersion relationship governing infinitesimal waves of constant k, and $g(k)=d\omega_0^2/dk$. We shall assume that ω_0 is a monotonically increasing function of k, so that $g\geq 0$. There is no significant loss of generality here, for the issue we face is that of understanding how two modes of different wave number can be in equilibrium, even when ω_0 is a single-valued function of k. We suppose that $\hat{G}$ is a constant (G); progress can also be made, at a cost of greater complexity, for more general $\hat{G}$.

Equations (3) and (4) provide a description of how smooth modulations of standing waves develop in space and time. For a stationary state they imply

$$\psi\equiv\psi_1[g_1/g]^{1/2}, \tag{5}$$

$$a+bk_{xx}+\tfrac{1}{4}\frac{db}{dk}k_x^2=0, \tag{6}$$

where ψ_1 and g_1 are real constants and

$$a(k)=\frac{3\{[\omega^2-\omega_0^2(k)]g(k)+9Gg_1|\psi_1|^2/4\}}{[g(k)]^2}, \tag{7}$$

$$b(k)=-g^{1/2}\frac{d^2}{dk^2}g^{-1/2}. \tag{8}$$

Consider now solutions in which k asymptotes to $k_{+\infty}$ for $x\to+\infty$, and in which $k\to k_{-\infty}$ for $x\to-\infty$. From (6), we see that $a(k_{+\infty})=a(k_{-\infty})=0$. For the kink $k_{+\infty}=k_{-\infty}$, but for the domain wall $k_{+\infty}\neq k_{-\infty}$ [so that in this case $a(k)=0$ must have multiple solutions]. The restriction $a(k_{\pm\infty})=0$ is a unification of two physical conditions that are (1) at finite amplitude, the resonant frequency shifts to ω, where $\omega^2=\omega_0^2(k)-\frac{9}{4}G|\psi|^2$; and

(2) the amplitude and wave number must adjust so that the flux of the adiabatic invariant $(U/\omega)d\omega/dk$, where U is the energy density, vanishes in the steady state (5).

In the case of the domain wall, a direct integration of (6) yields the key connection relation

$$\int_{k_-}^{k_+} a(k)b^{1/2}(k)dk = 0. \tag{9}$$

A simple interpretation of (9) can be obtained by the transformation $(k,x) \to (x,t)$ which converts the solution to (6) into the motion of a zero-energy particle of unit mass in the potential

$$V(k,k_-) = \frac{1}{b^{3/2}(k)}\left\{\int_{k_-}^{k} a(\kappa)b^{1/2}(\kappa)d\kappa - \tfrac{1}{2}b^{3/2}(k_-)k_x^2(k_-)\right\}, \tag{10}$$

where k_- is a reference value of k. Should this potential have a maximum at some value of k (say, k_-), then the constant of integration can be chosen so that $k_x(k_-) = 0$. Since $\partial k_x^2/\partial k = 2k_{xx}$, it follows that $k_{xx}(k_-) = a(k_-) = 0$, so that $k \to k_-$ as $x \to -\infty$. When the potential has a second maximum (at k_+, say), then a particle starting from rest at k_- in the potential (10) will just reach k_+ if $V(k_+,k_-) = V(k_-,k_-)$, i.e., if (9) holds. This corresponds to a domain wall with $k_+ = k_{+\infty}$. If $V(k_+,k_-) > V(k_-,k_-)$, the particle returns to k_-. This corresponds to a kink. In order for V to have a maximum, at $k = k_-$, say, it is necessary that $g^2 + \frac{9}{4}G|\psi|^2 g_k > 0$. The above interpretation applies for $b > 0$; a similar analysis can be carried out if $b < 0$ but, if b passes through zero, this simple picture breaks down.

The domain-wall profile can be calculated analytically when the net change in k and ω^2 are both small, i.e., when $k' \equiv k - k_1 = O(\delta) \ll 1$ and $g = O(\epsilon^2)$, so that

$$\omega_0^2 = \omega_1^2[1 + \epsilon^2(\gamma_1 k' + \gamma_2 k'^2 + \gamma_3 k'^3 + \gamma_3 k'^4 + \cdots)], \tag{11}$$

where k_1 and $\omega_1 = \omega_0(k_1)$ are constants. In this case, a domain-wall solution exists in the form

$$k = k_1 + \Delta k' \tanh[D\Delta k'(x - x_0)], \tag{12}$$

where $D = 5 - \gamma_1^2\gamma_4/\gamma_3^3$, provided that the dispersion satisfies the constraint $2\gamma_2^2 = \gamma_1\gamma_3$, or equivalently

$$\left[\frac{d^2}{dk^2}g^{-2}\right]_{k=k_1} = 0. \tag{13}$$

Substitution of (11) into (5) and (2) yields the complete solution

$$\rho' = \frac{2}{3}\left[\frac{2\omega_1(\omega_1-\omega)}{G}\right]^{1/2}\left\{1 - \frac{\gamma_2}{\gamma_1}\Delta k' \tanh[D\Delta k'(x-x_0)]\right\}\sin\left\{k_1(x-x_0) + \frac{1}{D}\ln\cosh[D\Delta k'(x-x_0)]\right\}. \tag{14}$$

Figure 1 shows an example of a domain-wall solution given by this theory.

The domain-wall solution to the nonlinear standing-wave equations has been obtained by requiring that the expansion of a/b as a power series in k' be $O(\delta^3)$. This restriction is imposed by the connection relation (9) and can only be met when (13) is satisfied along with

$$\omega - \omega_1 = \tfrac{1}{2}\epsilon^2\left\{\frac{\gamma_1^2\omega_1}{2\gamma_2} + \gamma_2\omega_1 D\Delta k'^2\right\}, \qquad |\psi_1| = \frac{2\sqrt{2}}{3}\left[\frac{\omega_1(\omega_1-\omega)}{G}\right]^{1/2}. \tag{15}$$

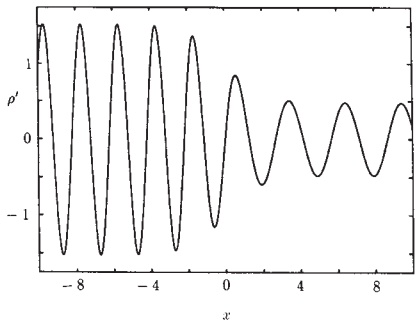


FIG. 1. Snapshot of a standing domain wall between modes of wavelength 2 for $x \to -\infty$ and waves of wavelength 3 for $x \to \infty$. This is a plot of ρ' given by (2), (5), and (11) for $x_0 = 0$, $\gamma_1 = \gamma_2$, $D = 1$, $\Delta k' = \pi/6$, and $k_1 = 5\pi/6$. The ordinate is ρ' in units of $\frac{2}{3}(-2/G\gamma_2)^{1/2}\epsilon\omega_1\gamma_1$.

The strong restriction (13) is not applicable to the kinks. In this case one need only require that $a/b = O(\delta^2)$. Setting now

$$\omega - \omega_1 = \tfrac{1}{2}\epsilon^2\left\{\frac{\gamma_1^2\omega_1}{2\gamma_2} + \gamma_1\left[1 - \frac{\gamma_1\gamma_3}{2\gamma_2^2}\right]\omega_1\bar{k}'\right\}, \tag{16}$$

we obtain the stationary solution

$$\theta = \theta(x_0) + k_1(x - x_0) + \left[\frac{\bar{k}'}{K}\right]^{1/2}\tanh[(K\bar{k}')^{1/2}(x - x_0)], \tag{17}$$

where

$$K = \frac{\gamma_1(\gamma_1\gamma_3 - 2\gamma_2^2)}{4\gamma_2(\gamma_1\gamma_3 - \gamma_2^2)}. \tag{18}$$

The spatial phase shift between $x = -\infty$ and $x = +\infty$ is $2(\bar{k}'/K)^{1/2}$. When changes in k are $O(1)$, it may be

necessary to restore the term $-\theta_l^2\psi$ on the right-hand side of (3).

In the two solutions just obtained, $\Delta k'$ and $\bar{k}'$ are free parameters that characterize the discontinuity of the domain wall and phase shift of the kink. Also, the scaling of ψ (i.e., ϵ) is independent of δ, which scales the spatial variations. Figure 2 shows a kink soliton given by this theory.

The algebraic details of the solutions for a kink and for a domain wall are given elsewhere [9], where time-dependent solutions are also considered.

Turning now to possible physical realizations of noncutoff kinks, we first consider the pendulum lattice [7] and optical phonons for which the dispersion law is of the form $\omega_\delta^2=\omega_{\delta,0}^2+(\omega_D^2-\omega_{\delta,0}^2)\sin^2(ka/2)$, where $\omega_D \equiv\omega_0(ka=\pi)$. According to (8), $b<0$ for this dispersion law, and a necessary condition for a kink solution is $g^2+\frac{9}{4}G|\psi|^2g_k<0$. This clearly excludes low-amplitude kink solutions when g is $O(1)$, and these systems cannot possess standing noncutoff kinks described by (3) and (4) unless $|\omega_D^2-\omega_{\delta,0}^2|/\omega_{\delta,0}^2\ll 1$. Similarly, $b<0$ for flexing modes, but the dispersion law for flexing shells [1] has a minimum at finite k. Near this wave number, g is small, and again such a system may have localized states.

In the presence of damping and parametric drive, (1) becomes

$$\frac{\partial^2\rho'}{\partial t'^2}+\hat{\omega}_0^2\rho'+2\gamma\rho'\cos 2\omega t+\beta\frac{\partial\rho'}{\partial t}=\hat{G}\rho'^3. \tag{19}$$

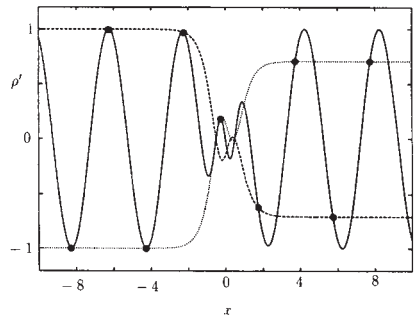


FIG. 2. Snapshot of a kink in a mode of wavelength 4. This is a plot of ρ' as given by (2), (5), and (15), for $x_0=0$, $\gamma_1=\gamma_2$, $K=1$, and $\bar{k}'=9\pi^2/64$. To make the kink profiles more apparent, two sets of points separated by half-wavelength intervals are indicated, and the curves of constant phase joining these points are shown dashed.

In the steady state (with θ real), the right-hand side of (3) is supplemented by the additional term $\gamma\psi^*-i\beta\omega\psi$. The steady state is now characterized by a phase lag, S, relative to the drive, where

$$\sin 2S=-\omega\beta/\gamma, \tag{20}$$

and $a(k)$ in (7) must be replaced by

$$\hat{a}(k)=\frac{3[\{\omega^2-\omega_\delta^2(k)+(\gamma^2+\beta^2\omega^2)^{1/2}\}g(k)+9Gg_1|\psi_1|^2/4]}{[g(k)]^2}. \tag{21}$$

When the dispersion law possesses a finite cutoff at long wavelengths, a solution of the form (2) exists for $\theta=0$, so that only the amplitude is modulated. In this case (1) gives

$$2i\omega\psi_t=[\omega_\delta^2(0)-\omega^2-3G|\psi|^2]\psi-c^2\psi_{xx}, \tag{22}$$

where we have set

$$\omega_\delta^2=\omega_\delta^2(0)+c^2k^2+\cdots. \tag{23}$$

Equation (22) is a well-known NLS equation. It possesses a nonpropagating breather soliton for $G>0$,

$$\psi=\left[\frac{2[\omega_\delta^2(0)-\omega^2]}{3G}\right]^{1/2}\operatorname{sech}\left\{\left[\frac{\omega_\delta^2(0)-\omega^2}{c^2}\right]^{1/2}(x-x_0)\right\}, \tag{24}$$

and a nonpropagating lower cutoff kink for $G<0$,

$$\psi=\left[\frac{\omega^2-\omega_\delta^2(0)}{-3G}\right]^{1/2}\tanh\left\{\left[\frac{\omega^2-\omega_\delta^2(0)}{2c^2}\right]^{1/2}(x-x_0)\right\}. \tag{25}$$

Our analysis of the modulational equations for standing waves leads us to the conclusion that such systems can display a wealth of localized states. The solutions presented here have the advantage of being analytic together with the disadvantage of not being the most general solutions to (3) and (4). Indeed, the observations of domain walls which prompted this research dealt with a system for which (13) does not apply. We conclude that the parameter space for these new localized states must be richer than has emerged from our leading-order solutions to (3) and (4).

One of us (S.J.P.) thanks the Division of Engineering and Geophysics of the Office of Basic Energy Science of

U.S. DOE for support. We wish to acknowledge a valuable discussion with Joseph Rudnick. Two of us (B.D. and A.L.) were supported from the NPS Direct-Funded Research Program. We are grateful to Maureen Roberts for help in preparing this manuscript.

[1] J. Wu, J. Wheatley, S. Putterman, and I. Rudnick, Phys. Rev. Lett. **59**, 2744 (1987).

[2] H. C. Yuen and B. M. Lake, Phys. Fluids **18**, 956 (1975).

[3] J. Scott Russell, Br. Assoc. Adv. Sci. Rep. **14**, 311 (1845).

[4] J. Wu, R. Keolian, and I. Rudnick, Phys. Rev. Lett. **52**, 1421 (1984).

[5] B. Denardo, W. Wright, S. Putterman, and A. Larraza, Phys. Rev. Lett. **64**, 1518 (1990).

[6] V. Zakharov and A. Shabat, Zh. Eksp. Teor. Fiz. **64**, 1627 (1973) [Sov. Phys. JETP **37**, 823 (1973)]; M. J. Ablowitz and H. Segur, *Solitons and the Inverse Scattering Transform* (SIAM, Philadelphia, 1981); J. Miles, J. Fluid Mech. **148**, 450 (1984).

[7] B. Denardo, B. Galvin, A. Greenfield, A. Larraza, S. Putterman, and W. Wright, Phys. Rev. Lett. **68**, 1730 (1992).

[8] L. F. Mollenauer, J. P. Gordon, and S. G. Evangelides, Laser Focus World **27**, No. 11, 159 (1991); L. F. Mollenauer, Opt. News, May issue, 42 (1986).

[9] S. J. Putterman and P. H. Roberts (to be published).

Geophys. Astrophys. Fluid Dynamics, Vol. 69, pp. 133–160
Reprints available directly from the publisher
Photocopying permitted by license only

IDEAL INSTABILITIES IN RAPIDLY ROTATING MHD SYSTEMS THAT HAVE CRITICAL LAYERS[1]

SIKUN LAN

Department of Mathematics, University of California, Los Angeles, California, 90024

WEIJIA KUANG[2] and PAUL H. ROBERTS[3]

Institute of Geophysics and Planetary Physics, University of California, Los Angeles, California, 90024

(*Received 7 May, 1992*)

The study of MHD instabilities in Earth's fluid core is simplified by making the usual assumptions that the fluid is uniform, inviscid, incompressible, and electrically-conducting (conductivity, σ). It is also supposed that it is in a state of rapid, and nearly solid-body rotation about the polar axis, Oz; in that frame, the prevailing velocity, $\mathbf{V}_0$, is slow and, like the magnetic field, $\mathbf{B}_0$, is nearly azimuthal. In an ideal fluid (defined as one in which σ is infinite), an important class of instabilities arise, namely Acheson's *field gradient* modes. In the case considered here, in which $V_0 = 0$, these arise when the s-gradient of $(B_0/s)^2$, where s is the distance from the axis of rotation, is sufficiently large. Commonly (and here), these instabilities are studied using a simple model in which B_0 is a function of s alone, and in which the fluid fills a cylindrical annulus $s_i \leqslant s \leqslant s_o$.

When such cylindrical models have been studied in the past, it has usually been supposed that the field has no "critical levels", i.e. that B_0 has no zeros in $s_i < s < s_o$. Here, however, the explicit aim is to understand better what happens when critical levels are present. Acheson's analysis is extended to this case, and a procedure is developed that generalizes his stability criterion. To the same qualitative accuracy as his method, the growth rate, p, of the instability is estimated. A "double-turning point" technique is presented that provides quantitatively accurate p. This technique also explains why and how the eigenfunctions are highly localized in s.

The results of numerical integrations for finite value σ are also presented.

KEY WORDS: Earth's core, ideal instabilities, rotating MHD.

1. INTRODUCTION

In a notable paper, Braginsky (1964a) proposed that the α-effect required to maintain the axisymmetric part of the geomagnetic field, which cannot alone act as a self-excited dynamo (Cowling's theorem), is created by the asymmetric instabilities of that axisymmetric part. He investigated a Cartesian model of the instabilities which was

[1] Contribution no. 92-05 of the Center for Earth and Planetary Interiors.

[2] Now at Department of Earth and Planetary Sciences, Harvard University, 20 Oxford Street, Cambridge, MA 02138.

[3] Also at Department of Mathematics, University of California, Los Angeles, California, 90024.

simple enough for all diffusive processes to be included. Later (Braginsky, 1967), when he considered the geophysically more relevant case of spherical geometry, the exigencies of the analysis forced him to concentrate on ideal systems, i.e. ones in which all forms of dissipation are absent. Since the magnetic field, **B**, and the fluid velocity, **V**, in the Earth's core are often thought to be predominantly zonal, he supposed for simplicity that they are precisely zonal, i.e. he took $\mathbf{B}=B_0\hat{\boldsymbol{\phi}}$, $\mathbf{V}=V_0\hat{\boldsymbol{\phi}}$, where $\hat{\boldsymbol{\phi}}$ is the unit vector in the direction of increasing longitude, and he tested the stability of this state in the presence of an axisymmetric, top-heavy density distribution. The resulting instabilities move zonally as planetary waves. Braginsky christened them "MAC Waves", an acronym that appropriately emphasizes the dominating role of *M*agnetic, *A*rchimedean (buoyancy) and *C*oriolis forces in their dynamics.

The neglect of viscosity (except in very thin boundary layers) in Braginsky's (1967) paper appears to be justified by the very small values of the Ekman number estimated for the core. (The Ekman number measures the ratio of viscous forces to Coriolis forces.) The abandonment of density diffusion is more questionable since it excludes persistent overturning of the core, of the type necessary to maintain the geodynamo over geological time; it introduces a Rayleigh-Taylor type mechanism where a Bénard type instability would have been more appropriate.

The neglect of the resistive diffusivity, η, is conceptually serious since, in its absence, the α-effect of the waves is lost, even though it was the regenerative properties of the waves that provided the main motivation for their study! Moreover, the geodynamo is usually believed to be "slow", i.e. it replenishes the magnetic energy it loses to ohmic diffusion on the same "diffusive time scale" as that over which the magnetic field would disappear if the α-source were absent. It would be an embarrassment if the instability operated on the much faster dynamic time scale, but this is the only time scale that enters the ideal theory! (It is the so-called "slow time scale", which is shorter than the diffusive time scale; see Section 2.) Not surprisingly resistivity was re-introduced into MAC wave theory by Braginsky and Roberts (1975), Roberts and Loper (1979), by Braginsky (1980), and by Fearn and his associates in many papers.

Brakinsky's 1967 theory contained a second strongly simplifying assumption. He supposed that the waves are of infinitesimal amplitude, i.e. he undertook a linear stability analysis. The very small values of the Rossby number estimated for the core suggest that this was a reasonable first step. Even though there is then no nonlinear feedback through the α-effect to equilibrate the geomagnetic field, a feedback still exists through the Lorentz force that the axisymmetric field exerts on the zonal flow, the so-called Malkus-Proctor mechanism. (The Rossby number is the ratio of the inertial forces, in the reference frame rotating with the angular velocity, Ω, of the Earth, to the Coriolis forces.)

The stress laid by Braginsky on the central role of asymmetric instabilities in the generation mechanism of the geodynamo sparked off considerable interest in the theory of rapidly rotating MHD systems, and significant advances were made, amongst others, by Acheson, Fearn and their associates. One such advance was the isolation by Acheson (1972) of a particular class of ideal instabilities which was subsequently christened "field-gradient instabilities", since when $V_0=0$ they rely on nonzero gradients in B_0, or more precisely in $(B_0/s)^2$, where s is distance from the

rotation axis. [Which $V_0 \neq 0$, the instability depends instead on the gradient of $(B_0/s)^2 - 2\Omega\mu\rho V_0/s$, where ρ is core density and μ is magnetic permeability; see Appendix B below.] Acheson gave a simple criterion for the occurrence of a field gradient instability; see Section 4 below. This cannot be applied to a system in which a critical surface is present, i.e. a surface on which $B_0 = 0$.

In his well-known study of the nearly axisymmetric kinematic dynamo, Braginsky (1964b) noted that the asymmetric magnetic field, **b**, induced by the (weak) asymmetric motions, **v**, might be particularly large near "resonant surfaces", where the wave speed of the asymmetric field coincided with V_0. He demonstrated that these were surrounded by resonant layers, $O(\eta^{1/3})$ in thickness, of "concentrated generation" where the α-effect is so strong that it provides an almost δ-function source; see also Braginsky and Roberts (1975). This suggested that these might play a key role in geodynamo theory. His discussion was based on the induction equation alone, i.e. both V_0 and **v** were given. The full MHD equations, governing the propagation of infinitesimal **b** and **v**, show however that resonant surfaces are not in general singularities of the MAC wave equations; the meridional components of **v** vanish on the resonant surfaces. Nevertheless, as Fearn has emphasized, when instabilities are resistive, i.e. when they owe their existence to finite ohmic resistance and would not occur in an ideal fluid, the critical surfaces, which would be singularities in idel theory, are replaced by critical layers having the same thickness, $O(\eta^{1/3})$, as Braginsky's resonant layers. Fearn (1984) found that the growth rate, p, of these resistive instabilities tends to zero, $p \propto \eta^{1/3}$, as $\eta \to 0$, in contrast to ideal instabilities, for which p tends to a constant as $\eta \to 0$.

In an effort to simplify the mathematics without sacrificing the essential (geo)physics, a number of studies have adopted one-dimensional models that pose eigenvalue problems involving ordinary, rather than partial, differential equations; for example, Kuang and Roberts (1990, 1991) analyzed a Cartesian model, the sheet pinch, and Acheson and Fearn frequently studied, as we do below, instabilities in a cylindrical annulus, $s_i \leqslant s \leqslant s_o$. Through this simplification, Fearn could more readily check Acheson's criterion for field-gradient instabilities, and could more easily get to the heart of the resistive instabilities.

In Section 2, we define a simple axisymmetric magnetostatic state in a cylinder, and we set up the basic equations and boundary conditions that govern perturbations that may ride on that state. These are simplified equations in two principal respects: (a) viscosity is abandoned (see below), and (b) the fluid inertia in the rotating frame, which is generally thought to be insignificant on secular variation time scales, is ignored. As a result, the dynamics is governed by the magnetostrophic equations, and the important dynamical time scale is the so-called "slow time scale", τ_s. The induction equation introduces a second key time scale, the "electromagnetic diffusion time", τ_η. Their ratio $\Lambda = \tau_\eta/\tau_s$ is the "Elsasser number", which is the important dimensionless number of the theory. We are mainly interested in lightly damped systems in which $\Lambda \to \infty$. [We also assume that the kinematic viscosity, ν, simultaneously tends to zero. In this way we avoid the complications, elucidated by Fearn and Weiglhofer (1992), that arise when ν is fixed as $\Lambda \to \infty$.]

Neither the theory of Braginsky (1967) nor that of Acheson (1972) could clarify

the nature of ideal instabilities, especially when one or more critical layers exist, although Braginsky (1980) did examine critical layer cases numerically, laying particular emphasis on situations in which the zonal shear is large. One of the main motivations of the present paper is to study ideal stabilities in greater detail, and especially local instabilities, both by generalizing the approach of Acheson (1972) and by developing an asymptotic theory capable, in the same range of parameters as that in which his criterion was derived, a quantitatively more accurate asymptotic theory. Since little is known about the structure of the toroidal field in the core, which is undetectable by magnetic measurements, we have felt at liberty to choose B_0 almost at random, although we have insisted that it possesses critical levels within $s_i < s < s_o$. Since the modes we examine are concentrated in intervals away from the boundaries, our choice of boundary conditions has little influence on the final outcome.

There is little overlap with the work of Fearn who, motivated by the geophysical application, has mainly (though not exclusively—see Section 4) concentrated on resistive instabilities in the presence of critical levels, by (for example) choosing a constant B_0/s that excluded the ideal instabilities, or by adding sufficient differential rotation V_0 that suppressed them. There is, however, a close connection between our work and that of Soward and Jones (1983), who initiated a double-turning point technique intimately related to Section 5 below, a technique that was subsequently adopted by Huerre and Monkowitz (1990).

In Section 3, we present our numerical results, which set the challenges of Sections 4 and 5. This led us to reconsider Acheson's celebrated criterion for field gradient instability, and in Section 4 we present his ideas in a form in which they can readily be applied to configurations that have critical levels. The analysis of this Section lead to qualitatively reliable results for both the occurrence and the growth rate of ideal instabilities. To obtain quantitatively reliable results, we present a new asymptotic analysis in Section 5 (which can be applied whether critical levels are present or not), and verify its accuracy by comparisons with integrations of the full equations. We summarize our findings in Section 6. The Appendices present arguments that justify our procedure in Section 5, and generalize Section 4 to include nonzero V_0.

2. GOVERNING EQUATIONS

The model we study is a cylindrical annulus with perfectly conducting walls. Its flat ends are $z = \pm h$; its outer curved boundary is $s = s_o$, and its inner curved boundary is $s = s_i$, where (s, ϕ, z) are cylindrical coordinates. It is filled with inviscid, incompressible fluid of constant electrical conductivity σ, magnetic permeability μ, and density ρ. Motivated by geophysical and astrophysical situations, we shall assume that the annulus is rapidly rotating about its axis, Oz, with a uniform angular velocity $\mathbf{\Omega}$, where by "rapidly rotating" we mean that $\Omega \gg v_A/s_o$, where $v_A = B_M/\sqrt{(\mu\rho)}$ is the Alfvén velocity based on the characteristic strength, B_M, of the prevailing magnetic field $\mathbf{B}$. As a result, we may replace the equation of motion by the simpler magnetostrophic equation in which the co-rotating reference frame is used and the inertial forces in that frame are neglected. The basic equations governing the system

are therefore

$$2\boldsymbol{\Omega}\times\mathbf{V}=-\nabla(P/\rho)+(\mu\rho)^{-1}(\nabla\times\mathbf{B})\times\mathbf{B},$$

$$(\partial_t-\eta\nabla^2)\mathbf{B}=\nabla\times(\mathbf{V}\times\mathbf{B}), \tag{1}$$

$$\nabla\cdot\mathbf{V}=\nabla\cdot\mathbf{B}=0,$$

where $\mathbf{V}$ is the fluid velocity and P is the pressure in which the centrifugal force has been absorbed; $\eta=1/\mu\sigma$ is the magnetic diffusivity of the fluid, ∂_t is the Eulerian derivative and t is time. The viscosity of the fluid is ignored for geophysical reasons; see Section 1.

Four time scales arise: $\tau_\Omega=1/2\Omega$, which is the rotation period or "fast" time scale, $\tau_A=s_o/v_A$, which is the Alfvén time scale, $\tau_s=\tau_A^2/\tau_\Omega$, which is the "slow" time scale, and $\tau_\eta=s_o^2/\eta$, which is the diffusion time scale. The last two of these are centrally important in what follows. Their ratio defines the Elsasser number, $\Lambda=\tau_\eta/\tau_s$, which measures (inversely) the importance of the magnetic diffusivity and (directly) that of the Lorentz forces. We use τ_s in defining our non-dimensional variables:

$$t\to\tau_s t, \qquad \mathbf{r}\to s_o\mathbf{r}, \qquad \mathbf{B}\to B_M\mathbf{B}, \qquad \mathbf{V}\to\frac{s_o}{\tau_s}\mathbf{V}. \tag{2}$$

The inner and outer radius of the annulus become $s_{ib}=s_i/s_o$ and 1, respectively; its height becomes 2α, where $\alpha=h/s_o$. Then (1) become

$$\hat{\boldsymbol{\Omega}}\times\mathbf{V}=-\nabla\Pi+(\nabla\times\mathbf{B})\times\mathbf{B},$$

$$(\partial_t-\Lambda^{-1}\nabla^2)\mathbf{B}=\nabla\times(\mathbf{V}\times\mathbf{B}), \tag{3}$$

$$\nabla\cdot\mathbf{V}=\nabla\cdot\mathbf{B}=0,$$

where $\hat{\boldsymbol{\Omega}}$ is the unit vector in the direction of $\boldsymbol{\Omega}$, and $\Pi=P/\rho v_A^2$. From now onwards we shall work with non-dimensional variables.

Suppose the system has achieved a magnetostatic equilibrium: $\mathbf{B}=\mathbf{B}_0$, $\mathbf{V}=\mathbf{0}$ and $\Pi=\Pi_0$, where $\nabla\Pi_0=(\nabla\times\mathbf{B}_0)\times\mathbf{B}_0$. We study the linear stability of this state by perturbing it infinitesimally:

$$\mathbf{B}=\mathbf{B}_0+\mathbf{b},$$

$$\mathbf{V}=\mathbf{0}+\mathbf{v}, \tag{4}$$

$$\Pi=\Pi_0+\varpi.$$

Substituting (4) into (3), we obtain the linearized equations for $\mathbf{v}$, $\mathbf{b}$ and ϖ:

$$\hat{\mathbf{\Omega}} \times \mathbf{v} = -\nabla\varpi + (\nabla \times \mathbf{B}_0) \times \mathbf{b} + (\nabla \times \mathbf{b}) \times \mathbf{B}_0,$$

$$(\partial_t - \Lambda^{-1}\nabla^2)\mathbf{b} = \nabla \times (\mathbf{v} \times \mathbf{B}_0), \tag{5}$$

$$\nabla \cdot \mathbf{v} = \nabla \cdot \mathbf{b} = 0.$$

The boundary conditions that solutions of (5) should be required to obey is a matter of some subtlety. At first sight it appears that, since the fluid is inviscid, we must apply

$$\mathbf{n} \cdot \mathbf{v} = 0, \qquad \mathbf{n} \cdot \mathbf{b} = 0, \qquad \mathbf{n} \times \mathbf{e} = 0, \tag{6}$$

where $\mathbf{n}$ is the normal to the boundary and $\mathbf{e} = \sigma^{-1}\mathbf{j} - \mathbf{v} \times \mathbf{B}_0 = \eta\nabla \times \mathbf{b} - \mathbf{v} \times \mathbf{B}_0$ is the perturbed electric field. (We have temporarily reverted to dimensional units.) Since the second of (6) is implied by the last of (6), this provides three independent scalar conditions on the boundary. The order in s of the differential system is, however, only four, so allowing only two conditions to be applied at $s = s_{ib}$ and at $s = 1$. This difficulty was also encountered by Fearn (1983a) who, following Condi (1978), restored a nonzero kinematic viscosity ν, so increasing the differential order in s of the system to ten, and adding two further scalar boundary conditions to (6), namely $\mathbf{n} \times \mathbf{v} = 0$ in the case of rigid walls. Both he and Condi analyzed the limit $\nu \to 0$ and the resulting boundary layers at $s = s_{ib}$ and $s = 1$. They showed that these could accommodate three of the five scalar conditions at $s = s_{ib}$ and $s = 1$, leaving the outer (or "mainstream") solution to satisfy the remaining two. These are the two that we apply in (15) below and, when they have been met and the necessary boundary layers have been constructed, conditions (6) are satisfied. There are no similar difficulties at the top and bottom boundaries, $z = \pm\alpha$.

The magnetic field in the basic state is taken to be toroidal and to depend only on the radius s: $\mathbf{B}_0 = B_0(s)\hat{\boldsymbol{\phi}}$.[4] The components of $\mathbf{v}$ and $\mathbf{b}$ are denoted by (v_s, v_ϕ, v_z) and (b_s, b_ϕ, b_z), respectively. After writing $\mathbf{\Omega} = \Omega\hat{\mathbf{z}}$ and eliminating ϖ, v_ϕ and b_ϕ, we obtain the following equations for v_s, v_z, b_s and b_z:

$$\frac{\partial v_s}{\partial z} + \left(B_0 \frac{\partial}{\partial s} - B_0'\right)\frac{\partial b_s}{\partial z} + B_0\left(\frac{1}{s^2}\frac{\partial^2}{\partial \phi^2} + \frac{\partial^2}{\partial z^2}\right) b_z = 0, \tag{7}$$

$$\frac{\partial v_z}{\partial z} + \left(-B_0 \frac{\partial^2}{\partial s^2} - \frac{3B_0}{s}\frac{\partial}{\partial s} + B_0'' + \frac{B_0'}{s} - \frac{2B_0}{s^2} - \frac{B_0}{s^2}\frac{\partial^2}{\partial \phi^2}\right) b_s$$
$$- \left(B_0 \frac{\partial}{\partial s} + B_0' + \frac{3B_0}{s}\right)\frac{\partial b_z}{\partial z} = 0, \tag{8}$$

[4] Note: We need *not* demand that $\mathbf{B}_0$ obeys Ohm's law or insist that the tangential components of the electric field $\mathbf{E}_0 = \eta\nabla \times \mathbf{B}_0$ in the basic state vanish on the walls. If we did so, $\mathbf{B}_0$ would decay to zero on the τ_η time scale. Since this time is long compared with the time scales of the instabilities, we are entitled, when studying the instabilities, to ignore the diffusive evolution of $\mathbf{B}_0$ and to regard $\mathbf{B}_0$ as given.

$$-\frac{B_0}{s}\frac{\partial v_s}{\partial \phi}+\left(\partial_t-\frac{1}{\Lambda}\nabla^2\right)b_s-\frac{1}{\Lambda}\left(\frac{2}{s}\frac{\partial b_s}{\partial s}+\frac{b_s}{s^2}+\frac{2}{s}\frac{\partial b_z}{\partial z}\right)=0, \tag{9}$$

$$-\frac{B_0}{s}\frac{\partial v_z}{\partial \phi}+\left(\partial_t-\frac{1}{\Lambda}\nabla^2\right)b_z=0. \tag{10}$$

The coefficients in these equations depend only on s and we may therefore seek solutions of normal mode form:

$$u(s,\phi,z,t)=u(s,z)e^{im\phi+pt}+c.c., \tag{11}$$

where the complex-valued function u might be v_s, v_z, b_s or b_z; "$c.c.$" stands for the complex conjugate of the expression preceding it, m is an integer, and $p=p_r+ip_i$ is the complex growth rate. Solutions of (7)–(10) with $p_r>0$ are unstable modes; those with $p_r<0$ are linearly stable. (We shall call p_r the "growth rate" and p_i the "frequency".) We confine attention to the asymmetric modes ($m\neq 0$). The axisymmetric modes are harder to excite in highly rotating systems, and there is also less geophysical motivation in studying them (see Section 1).

Because of the simplicity of our boundary conditions (6) on the flat walls of the annulus, it is also easy to separate out the z-dependence of the solutions. These conditions show that $b_z=\partial b_s/\partial z=0$ on $z=\pm\alpha$. By differentiating $\nabla\cdot\mathbf{b}=0$ with respect to z, we also find that $\partial^2 b_z/\partial z^2=0$ on $z=\pm\alpha$. We may therefore write

$$\begin{pmatrix}v_s(s,z)\\ b_s(s,z)\end{pmatrix}=\begin{pmatrix}v_s(s)\\ b_s(s)\end{pmatrix}\cos\frac{k\pi}{2\alpha}(z+\alpha),$$
$$\begin{pmatrix}v_z(s,z)\\ b_z(s,z)\end{pmatrix}=\begin{pmatrix}v_z(s)\\ b_z(s)\end{pmatrix}\sin\frac{k\pi}{2\alpha}(z+\alpha) \qquad k=1,2,\ldots \tag{12}$$

Substituting (11) and (12) into (7)–(10), and eliminating v_s and v_z, we obtain the system of equations that we solved numerically:

$$\frac{1}{\Lambda}\left[\left(D^2+\frac{3}{s}D-n^2-\frac{m^2-1}{s^2}\right)b_s+\frac{2n}{s}b_z\right]$$
$$-\frac{imB_0}{s}\left[(B_0D-B_0')b_s+B_0\left(n+\frac{m^2}{ns^2}\right)b_z\right]=pb_s, \tag{13}$$

$$\frac{1}{\Lambda}\left(D^2+\frac{1}{s}D-n^2-\frac{m^2}{s^2}\right)b_z+\frac{imB_0}{s}\left[\left(B_0D+B_0'+\frac{3B_0}{s}\right)b_z\right.$$
$$\left.+\frac{1}{n}\left(B_0D^2+\frac{3B_0}{s}D-B_0''-\frac{B_0'}{s}-\frac{(m^2-2)B_0}{s^2}\right)b_s\right]=pb_z. \tag{14}$$

where D and $'$ denote d/ds and n is the wave number in z direction: $n=k\pi/2\alpha$; these coincide with equations (2.20) and (2.21) of Fearn and Weiglhofer (1992). As boundary conditions, we shall require (Fearn 1983a) that

$$\begin{aligned} &b_s=0, \qquad B_0^2 Db_s+\frac{B_0^2}{n}\left(n^2+\frac{m^2}{s_{ib}^2}\right)b_z-\frac{1}{\Lambda}\left(\frac{ins_{ib}}{m}+1\right)Db_z=0, \quad \text{at} \quad s=s_{ib}, \\ &b_s=0, \qquad B_0^2 Db_s+\frac{B_0^2}{n}(n^2+m^2)b_z-\frac{1}{\Lambda}\left(\frac{in}{m}-1\right)Db_z=0, \qquad \text{at} \quad s=1. \end{aligned} \tag{15}$$

Equations (13)–(15) define a fourth order eigenvalue problem for the complex eigenvalue p and the complex-valued eigenfunctions $b_s(s)$ and $b_z(s)$. The results of numerical integrations will be presented in Section 3. The remaining Sections and Appendix A will describe its analytic properties.

By setting $\Lambda=\infty$ in (13)–(14), we may obtain the corresponding *ideal equations*:

$$-\frac{imB_0}{s}\left[(B_0D-B_0')b_s+B_0\left(n+\frac{m^2}{ns^2}\right)b_z\right]=pb_s, \tag{16}$$

$$\begin{aligned} \frac{imB_0}{s}\Bigg[&\left(B_0D+B_0'+\frac{3B_0}{s}\right)b_z \\ &+\frac{1}{n}\left(B_0D^2+\frac{3B_0}{s}D-B_0''-\frac{B_0'}{s}-\frac{(m^2-2)B_0}{s^2}\right)b_s\Bigg]=pb_z. \end{aligned} \tag{17}$$

Because these are of second order, we may apply only one condition at each end point, and the correct demands are

$$b_s=0, \qquad \text{at} \quad s=s_{ib}, 1. \tag{18}$$

We selected as the magnetic field in the basic state

$$B_0=\cos\left(\lambda\pi\frac{s-s_{ib}}{1-s_{ib}}\right). \tag{19}$$

The parameter λ, when taken to be a positive integer, is exactly the number of zeros of B_0 in $s_{ib}\leqslant s\leqslant 1$. The sign of B_0 is unimportant in (13)–(15) and in (16)–(18).

3. NUMERICAL RESULTS

The numerical procedures adopted to solve the complex eigenvalue problem were the following: The interval $s_{ib}\leqslant s\leqslant 1$ was divided into N equal parts, and stable finite

difference approximations were introduced to transform (13)–(15) into a complex matrix eigenvalue problem for p (see Kreiss, 1972); the resulting algebraic eigensystem was then solved either by applying the QR algorithm or by using inverse iteration (see Golub and van Loan, 1989). In all cases, we took $s_{ib}=0.35$.

The QR algorithm gives all the eigenvalues of an algebraic system, but is of practical use only for small systems: $N \leqslant 50$. Given an initial estimate of a single eigenvaue, inverse iteration converges to the eigenvalue closest to it in the complex p-plane. Because only sparse banded matrices are involved, inverse iteration can handle much larger systems (even with $N \sim 1000$) in an acceptable time, and this generally provides sufficient numerical resolution.

Only the largest growth rate is of physical interest and, given λ, m, Λ and n, we use the QR algorithm with a small N to pick out the eigenvalue with the largest p_r. Then, to obtain the growth rate to higher precision, we use the value of p obtained from the QR algorithm as the initial estimate in a subsequent inverse iteration with a much larger N. Similar methods were repeatedly used by Kuang and Roberts (1990, 1991). Because our main interest lies in asymmetric ($|m| \geqslant 1$) instabilities of fields having at least one critical level, computations were restricted to $\lambda=1$ or 2. We considered only $m=1$ or 2 because beyond $m=2$ the system is less prone to instability (see Section 4). Because we have left the height α of the annulus unspecified, we may, as for Λ, select any positive n that we wish. We found that, qualitatively, the dependence of p on Λ and n is virtually the same for all four of our chosen m and λ.

Figure 1 shows, in a log-log plot, contours of equal growth rate p_r in the Λn-plane for $m=1$ and $\lambda=2$. The labels on the curves specify the values of p_r on them, but no contours are shown in the blank, top-left area, where the corresponding modes are stable: $p_r<0$. The contour $p_r=0$ defines, for the given m and λ, the critical states, $\Lambda=\Lambda_c(n)$, separating stable and unstable states. Except for that contour, which was so accurately determined that the density of the crosses would have obscured the curve, crosses show the results of the numerical computations that guided the construction of the contours.

Because of mode-crossing (see below), there is a cusp on the Λ_c curve and on some of the other curves. We connect these in Figure 1 by a dashed curve. Below this curve, the values of the growth rate indicate that the modes are of resistive type; above it, they have the character of ideal modes.

Fearn (1984) showed that, if a critical level exists in the interval $s_{ib} \leqslant s \leqslant 1$ then, for the resistive mode, $p \propto \Lambda^{-1/3}$, asymptotically as $\Lambda \to \infty$. It is apparent from Figure 2 that, for small fixed n, the growth rate decreases to zero as Λ tends to infinity in precisely the required way. Figure 3 shows that the eigenfunction is concentrated around the critical points, so confirming the basic tenet of the critical layer analysis that led to the $p \propto \Lambda^{-1/3}$ asymptotic law.

Above the dashed line in Figure 1, the asymptotic behavior of the growth rate as $\Lambda \to \infty$ is completely different. Figure 4 provides an example. It shows that p tends to a constant as Λ tends to infinity. This is typical of ideal instabilities. The limiting value of the growth rate depends on n and tends to a constant as $n \to \infty$. The eigenfunction (Figure 5) is significantly large only far away from the critical levels.

Another limit is also of considerable later interest. Figure 6 shows that the Λ_c curve

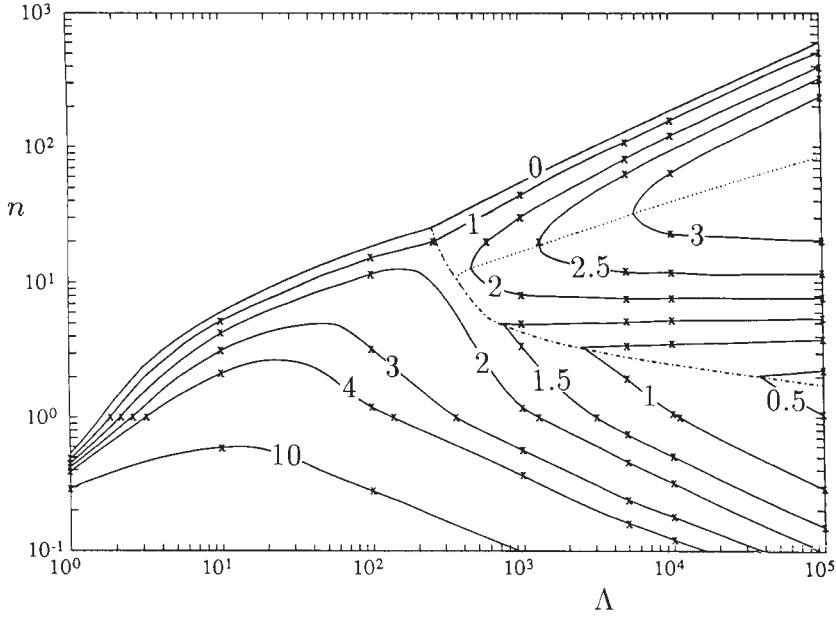


Figure 1 Contours of p_r in the $(\log \Lambda, \log n)$-plane. the modes above the dashed curve have the character of ideal instabilities while the curves beneath it are of resistive type. The labels on the contours specify the value of the growth rates p_r. The curve $p_r = 0$, and also the dashed and dotted curves, were calculated numerically with high precision, while the remaining curves were constructed from the numerical results marked with crosses.

approaches the upper right corner of Figure 1 in the direction of $n^2 \propto \Lambda$. This suggests that an important limit is $\Lambda, n \to \infty$ with $K \equiv n^2/\Lambda$ held fixed. In this limit, which will be studied analytically in detail in Sections 4 and 5, p tends to a constant.

The situation in the neighborhood of the dashed line in Figure 1 is interesting. The dashed line is composed of those (Λ, n) pairs for which the growth rate is the same for both ideal and resistive instabilities. Ideal stabilities exist for certain (Λ, n) pairs below the dashed line but have smaller growth rates than those of the resistive modes. Similarly, resistive instabilities may exist above the dashed line but they are dominated by the ideal modes. But, for any given (Λ, n) pair, we are generally interested in the mode having the maximum growth rate, p_r, and this is the only one shown in Figure 1. In the vicinity of the dashed line, mode crossings are expected. As a result, the numerical calculations are difficult and must be performed with great care. Figure 7 gives an example how the mode crossing curve behaves as $\Lambda \to \infty$. The dashed curve was obtained by solving (16)–(18) numerically. For each λ and m, there is a threshold, $n = n_c$ (say), such that the ideal equations possess discrete unstable

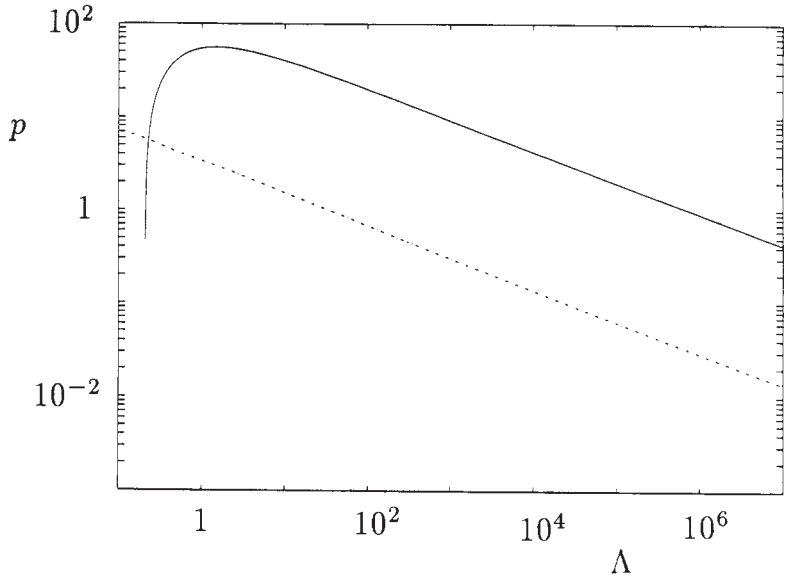


Figure 2 A log-log plot of the growth rate p_r (solid curve) and the frequency p_i (dashed curve) as a function of Λ for $n=0.1$, $m=1$ and $\lambda=2$. This figure demonstrates that $p \propto \Lambda^{-1/3}$ as $\Lambda \to \infty$, so showing that the instability is of resistive type.

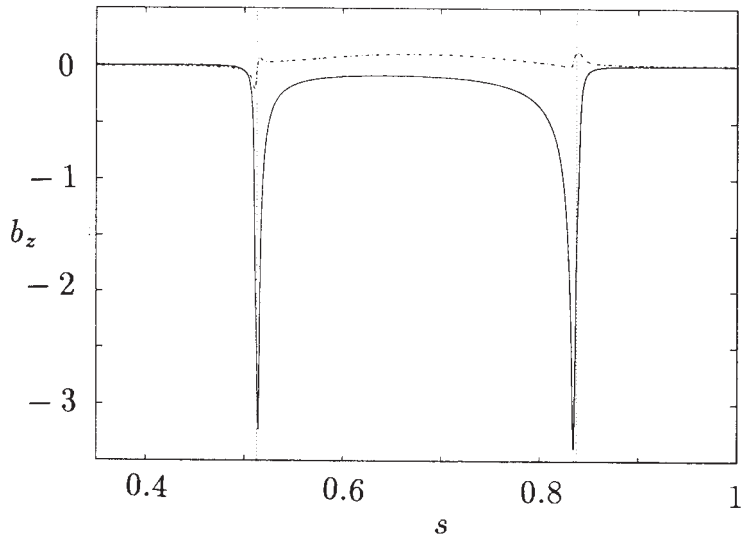


Figure 3 The eigenfunction $b_z(s)$ for $\Lambda=10^5$, $n=0.1$, $m=1$ and $\lambda=2$. The horizontal axis extends from $s=0.35$ to $s=1$. The solid curve shows $\mathrm{Re}(b_z)$; the dashed one displays $\mathrm{Im}(b_z)$. The two critical points are indicated the dotted lines. It may be particularly noted that the eigenfunction is significant only near the critical levels.

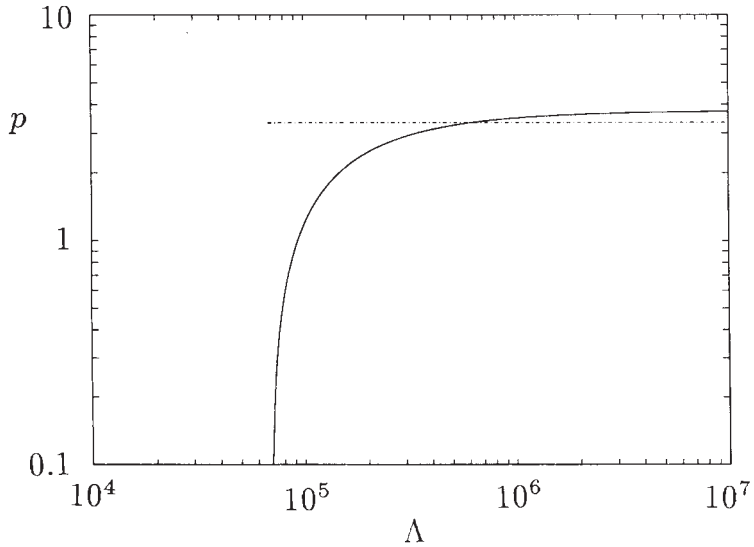


Figure 4 A log-log plot of the growth rate p_r (solid curve) and the frequency p_i (dashed curve) as a function of Λ for $n=500$, $m=1$ and $\lambda=2$. This figure demonstrates that $p\to$constant as $\Lambda\to\infty$, so showing that the instability is of ideal type.

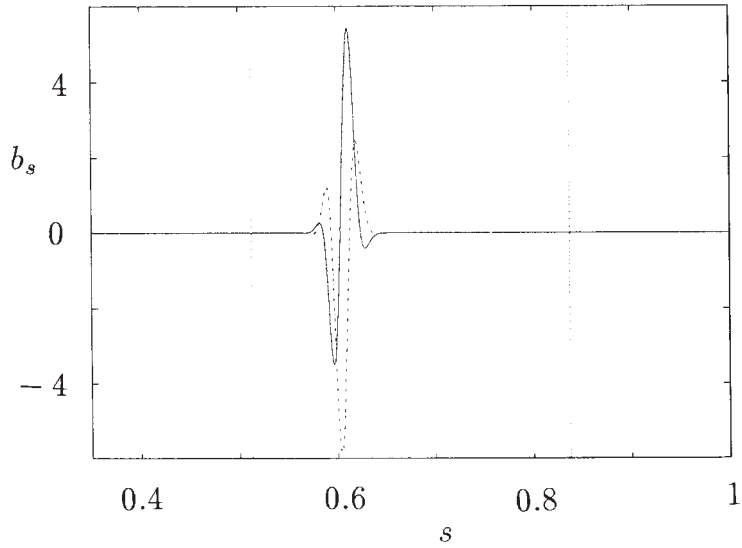


Figure 5 The eigenfunction $b_s(s)$ for $\Lambda=10^5$, $n=500$, $m=1$ and $\lambda=2$. The horizontal axis extends from $s=0.35$ to $s=1$. The solid curve shows $\mathrm{Re}(b_s)$; the dashed one displays $\mathrm{Im}(b_s)$. The two critical points are indicated the dotted lines. It may be particularly noted that the eigenfunction is significant only far from the critical levels.

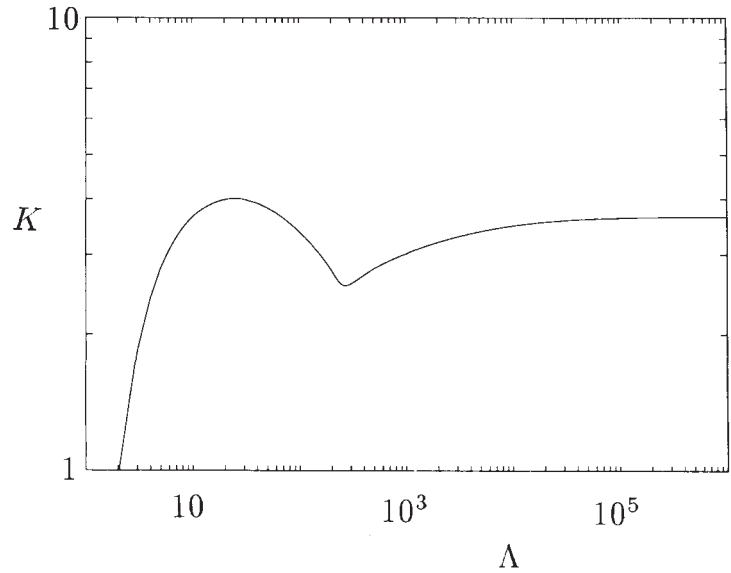


Figure 6 The Λ_c curve displayed as a log-log plot of $K \equiv n^2/\Lambda$ versus Λ for $m=1$ and $\lambda=2$. This shows that $K \to$ constant as $\Lambda \to \infty$. The cusp that separates the ideal instabilities from the resistive instability is clearly seen near the middle of the Figure.

modes for $n > n_c$, but no discrete modes whatever for $n < n_c$. The value of n_c for the case $m=\lambda=2$ is approximately 5.4 (see Figure 7); for $m=1$ and $\lambda=2$, it is approximately 0.3 (see Figure 1). It is interesting to observe (Figure 7) how the solutions to the full equations approach the n-axis along $0 < n < n_c$ as Λ increases. The curves displayed in Figure 7 are segments of those shown in Figure 8, which spans a greater range of n, and shows that, for a fixed finite Λ, a large value of n does not promote instability. For $\Lambda = \infty$ however, instability exists for any n.

Concerning the direction of propagation of the disturbances (i.e. the sign of $p_i m$), we found that ideal instabilities always move westward ($p_i m > 0$), in agreement with the general result of Acheson (1972). The situation is more complicated for the resistive modes. Fearn (1988) concluded, in the case of insulating boundaries, that $p_i m$ always becomes negative as $n \to 0$, corresponding to eastward drift. In our case of infinitely conducting boundaries, we sometimes obtained eastward drift (e.g. for $\lambda=2$, $m=2$), and sometimes westward drift, perhaps because the values of n that we chose, which in no case was less than 0.01, were not sufficiently small.

So far little asymptotic analysis has been performed for the ideal instability. The next two Sections will focus on the study of the ideal instability and interpret the corresponding numerical results presented in this Section.

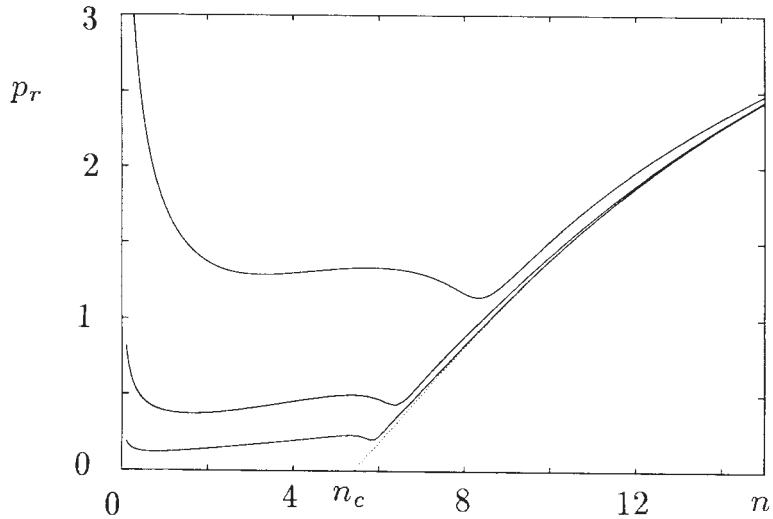


Figure 7 The growth rate, p_r, as a function of n for $m=2$, $\lambda=2$, and (from top to bottom) $\Lambda=5000$, 10^5, 10^6. The dotted line is the result of solving numerically the ideal eigensystem (16)–(18). Because of numerical difficulties in certain range of n, the dotted curve is less accurate than the solid curves.

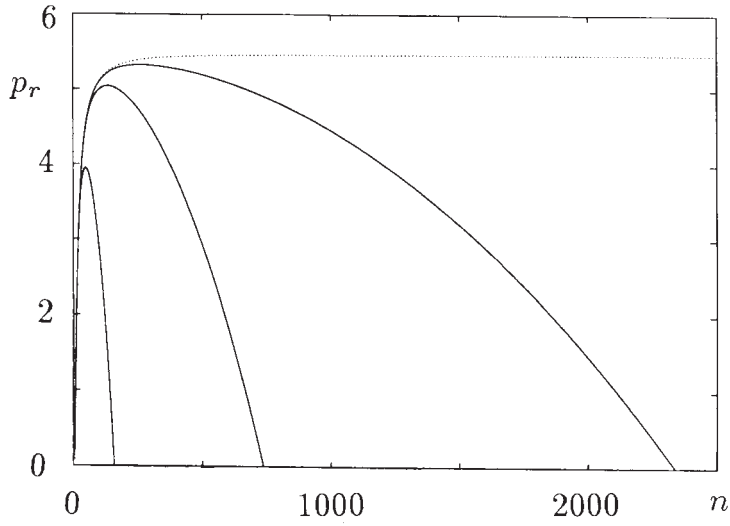


Figure 8 The same case as the previous figure but for a larger range of n. Here the curves refer (from top to bottom) to $\Lambda=\infty$, 10^6, 10^5, 5000. This figure shows that, except for $\Lambda=\infty$ (dotted), there exists for each Λ, a maximum n for which instability can occur.

4. LOCAL ANALYSIS OF IDEAL INSTABILITY

Ideal instability is usually so-named because it arises in an ideal fluid, i.e. one in which all diffusion processes are absent. The expectation is therefore that it can be located by an analysis of (16)–(18), which arose by formally setting $\Lambda = \infty$ in (13)–(15). This apparently simplifying step may, however, introduce serious mathematical complications, associated with the loss of discrete eigenmodes and the appearance of a continuous spectrum; see Lifschitz (1989).

Suppose that there is one critical level, at $s = s_c$ say. The system (16) and (17) then has an irregular singularity at $s = s_c$, and its general solution has an essential singularity at that point. Writing $x = s - s_c$, we find that

$$b_s \propto x\, e^{\pm q/x}, \qquad b_z \propto x^{-1}\, e^{\pm q/x}, \tag{20}$$

where $q = pn[s_c/mB_0'(s_c)]^2$. Provided $p_r \neq 0$, one of solutions (20) becomes infinite as $x \to 0+$ while the other tends to zero. By stipulating that only the latter be included, we may develop a solution that is nontrivial in $s_c < s < 1$, and identically zero in $s_{ib} < s < s_c$, so satisfying the boundary condition at $s = s_{ib}$. It will not in general satisfy the other boundary condition at $s = 1$, but will do so for the special eigenvalues of p. Similarly, by selecting the other solution in (20), we may be able to find a nontrivial solution that vanishes identically only in $s_c < s < 1$ and which therefore trivially obeys the boundary condition at $s = 1$.

More generally, if there are λ critical levels, we may construct solutions that are vanishingly small everywhere in $s_{ib} \leqslant s \leqslant 1$, except in one of the $\lambda + 1$ intervals, (s_{ib}, s_1), $(s_1, s_2), \ldots, (s_{\lambda-1}, s_\lambda)$, and $(s_\lambda, 1)$, into which the critical points $s_1, s_2, \ldots$ and s_λ divide $s_{ib} \leqslant s \leqslant 1$.

It was in this sense that we were able to solve (16)–(18) and to produce the dashed curves shown on Figures 7 and 8, which are obviously approached by solutions of the full equations as $\Lambda \to \infty$. It became increasingly difficult, however, to find ideally unstable solutions of (16)–(18) as $p_r \to 0$, and the reason is not far to seek: if $p_r = 0$ but $p_i \neq 0$, neither of the solutions (20) are acceptable.

Because of the difficulties introduced by the ideal equations, we find it helpful to approach the ideal modes through the limit $\Lambda \to \infty$. This also affords us the opportunity of studying ideal modes in the presence of resistivity. In what follows we regard Λ as finite, though very large.

To study ideal instability, Acheson (1972) introduced an analytic technique which he called "local analysis". With its help he proffered the following criterion for ideal instability (Acheson, 1983):

$$\Delta > m^2, \tag{21}$$

where

$$\Delta = \frac{s^3}{B_0^2} \frac{d}{ds}\left(\frac{B_0}{s}\right)^2. \tag{22}$$

To derive this, he assumed (Acheson, 1978) that $\Lambda \gg 1$, that $n \gg 1$, and that the instability is localized (i.e. that the amplitude of the eigenfunction is significant only in an interval of s much smaller than $1 - s_{ib}$, and which does not contain a boundary).

Fearn (1983b) tested the criterion (21) numerically for a variety of fields B_0 and concluded that it provides a good qualitative description of ideal instability. Because (21) requires that, for instability, the field gradient should be nonzero (or, more precisely, that the gradient of B_0/s should be nonzero), the instability is also often called "field gradient instability".

Acheson's analysis supposed that B_0 is nowhere zero and, if we attempted to apply (21) to our field (19), we would find that it predicts instability at a critical point, despite the fact that the eigenmode, such as that shown in Figure 5, is large only far away from the zero of B_0. At first sight therefore, (21) cannot be applied when critical points occur. In this Section, we perform the local analysis anew, and deduce a qualitative description of ideal instability in a field having one or more critical points.

It may be noted from the basic equations (13)–(15) that the eigenvalue p always appears in combination with the added term n^2/Λ which we denote by K. We shall write $p_K \equiv p + K \equiv p + n^2/\Lambda$, and we shall be interested only in cases in which $K \lesssim 1$, since when K is large there is no instability. Note particularly that $K \lesssim 1$ includes the case $K \ll 1$.

Following Acheson, we assume that $\Lambda \gg 1$ and $n \gg 1$; these inequalities are indeed essential for the existence of localized solutions (see Section 5). We also suppose that m, s, K, B_0 and the derivatives of B_0 are of O(1). Under these assumptions, we may, after considerable algebraic labors, eliminate b_z from (13)–(14) to obtain

$$(C_4 D^4 + C_3 D^3 + C_2 D^2 + C_1 D + C_0) b_s = 0, \tag{23}$$

where, to leading order,

$$\begin{aligned}
C_4 &= \Lambda^{-2}, \\
C_3 &= \mathrm{O}(\Lambda^{-2}), \\
C_2 &= \left(-2p_K + i\frac{4mB_0^2}{s^2} \right) \Lambda^{-1} - \frac{m^4 B_0^4}{n^2 s^4} = \mathrm{O}(d^4), \\
C_2 &= -\frac{3m^2 B_0}{n^2 s^3} + \mathrm{O}(\Lambda^{-1}) = \mathrm{O}(d^4), \\
C_0 &= p_K^2 - i\frac{4mB_0^2}{s^2} p_K - \frac{2m^2 B_0^3 B_0'}{s^3} + \frac{(m^2 - 2)m^2 B_0^4}{s^4},
\end{aligned} \tag{24}$$

and $d \equiv n^{-1/2}$. The O symbol in (24) refers to the limit $d \to 0$, but we also have in mind that $\Lambda \to \infty$ because $K \lesssim 1$.

If $D \gg 1$, the third order and first order terms may be neglected compared respectively with the fourth and second order terms. Even if $D \sim 1$, the third and first order terms may still be neglected compared with C_0 because $\Lambda \gg 1$ and $n \gg 1$. The asymptotic equation then becomes

$$(C_4 D^4 + C_2 D^2 + C_0) b_s = 0 \tag{25}$$

which agrees with Eq. (2.14) of Acheson (1983).

The local analysis assumes that the solution is localized in a small interval of the real s-axis in which the coefficients of (23) can be approximated by constants and in which $D \gg 1$. One imagines that the solution in this interval is a "wave packet" composed of a linear combination of solutions to (23) of the form

$$b_s(s) \sim e^{ils}, \tag{26}$$

where $l \gg 1$. It is visualized that interference between different modes in the packet obliterates the disturbance outside the interval under consideration.

Substituting (26) into (25), one can solve for p_K to obtain

$$p_K = \frac{mB_0^2}{s^2}\left\{2i \pm \sqrt{\left[\Delta - m^2\left(1 + \frac{l^2}{n^2}\right)\right]}\right\} - \frac{l^2}{\Lambda}. \tag{27}$$

Supposing that the radical in (27) is real, choosing the upper sign which gives the larger p_r, and taking the real and imaginary parts of (27), we obtain

$$p_r + K = \frac{mB_0^2}{s^2}\sqrt{\left[\Delta - m^2\left(1 + \frac{l^2}{n^2}\right)\right]} - \frac{l^2}{\Lambda}, \tag{28}$$

$$\leqslant \frac{mB_0^2}{s^2}\sqrt{(\Delta - m^2)} \equiv L_r, \tag{29}$$

$$p_i = -\frac{2mB_0^2}{s^2} \equiv L_i. \tag{30}$$

From (29) one can see that (21) is a necessary condition for p_r to be positive. But when B_0 possesses one or more zeros, (21) is misleading although (29) is still available, and it can be used to seek the maximum value of $L_r - K$ as a function of s; for instability, it is necessary that that maximum is positive. The function $L = L_r + iL_i$ depends on λ, m and s alone. It is shown in Figure 9 for $m = 1$ and $\lambda = 2$. One can see that neither of the maxima of L_r are situated near the critical levels where ideal instabilities might have been expected from (21). It may be anticipated that the eigenfunctions will be large only for values of s near that of the higher peak. That this is exactly the case may be confirmed by reference to Figure 5. Also, it may be anticipated that the corresponding eigenvalue will be numerically close to (but less

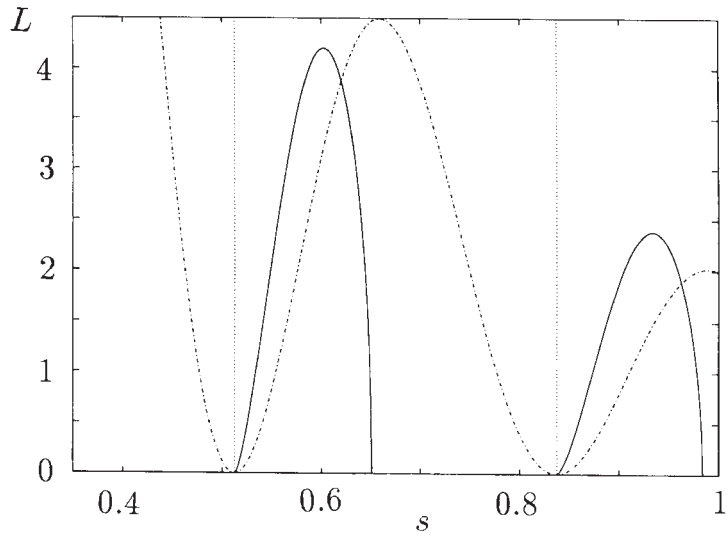


Figure 9 The real (solid curve) and imaginary (dashed curve) parts of $L(s)$ for $m=1$ and $\lambda=2$. The horizontal axis extends from $s=0.35$ to $s=1$. The Figure shows that the critical levels, indicated by the dotted lines are not in the region of instability. The maximum growth rate is attained near the maximum of L_r.

than) the value of L_r at the higher peak, and this also is true. Further, because L is independent of Λ and n, it may be anticipated that the eigenvalue will approach a constant not only as $\Lambda\to\infty$ with fixed n but also as Λ and n tend to ∞ for fixed $K\equiv\Lambda^{-1}n^2$. This too is confirmed by Figures 4 and 6. Similarly, the behavior of p_r versus n for fixed Λ, as shown in Figure 8, is explained by (29).

When no critical layers are present, all modes of sufficiently large m are stable according to (21), but (29) indicates that, when there are one or more critical levels, ideal instability may exist for any m. This is because $\Delta\to\infty$ as a critical level, $s=s_c$, is approached; therefore, for large m, maxima of L_r-K exist in the neighborhood of every critical level and move towards those levels as m increases. To determine the maximum near $s=s_c$ for $m\gg1$, we may write $B_0\approx B_0'(s_c)(s-s_c)$. In this approximation, the maximum of L_r occurs at $s=s_c+3s_c/2m^2$, and is $L_r=3\sqrt{3[B_0'(s_c)/2m]^2}$. In the case of model (19), this gives correctly the $m=6$ entry in Table 1, which lists to one decimal place accuracy the maximum value of L_r as a function of m in the case $\lambda=1$. It may also be noticed that, as is not uncommon, the $m=2$ mode is potentially more unstable than the $m=1$ mode.

Turning to quantitative comparisons with the numerical integrations, Figure 9 suggests that instability will occur for K as large as 4.2, but Figure 6 shows numerically that the marginal K is in reality about 3.6. The eigenvalue corresponding to the eigenfunctions in Figure 5 is $1.19+3.33i$ (for which $K=2.5$) whereas Figure 9 predicts

Table 1 Estimated growth rate versus m

m	1	2	3	4	5	6
L_r	1.6	2.3	2.0	1.5	1.0	0.8

$1.70+3.24i$. The reason why (29) and (30) do not provide better quantitative agreement is because (i) we set $l=0$ in obtaining (29) from (28), though l is unknown and should depend on Λ and n; (ii) the approximations on which the local analysis rests are not accurate and therefore (26) does not provide a strict mathematical solution.

In the next Section, an asymptotic description of ideal instability is devised that provides more accurate quantitative predictions.

5. DOUBLE TURNING POINT ANALYSIS FOR IDEAL INSTABILITY

The aim of this Section is to develop a method that improves on the approach of the last Section by providing quantitatively accurate solutions under the same assumptions as those underpinning the last Section, i.e. $\Lambda \gg 1$, $n \gg 1$, while m, s, p, K, B_0 and the derivatives of B_0 are O(1). We therefore obtain the same asymptotic equation (25) as before, but develop a different technique to analyze it. We believe our approach is conceptually significant since (i) it generally leads to an improved understanding of the behavior of the solutions, and (ii) most numerical methods rely on finite difference approximations that become increasingly inaccurate as some of the defining parameters become very large or very small i.e., in precisely those cases in which the asymptotic method should perform well. We suppose at the outset that ideal instabilities are zero at critical levels to leading order, and that there are therefore no boundary layers at the critical levels. This excludes resistive instabilities from consideration.

Numerical experiments show that the fourth order term in (25) does not affect the eigenvalue p significantly. Asymptotic analysis (see Appendix A) also shows that the eigenvalue is mainly determined by the C_2 and C_0 terms in (25). Note however that the C_4 term may be important on the real s-axis, and that therefore it was retained in the local analysis of Section 4. For the remainder of this Section, we therefore study the following eigenvalue problem for p:

$$(C_2 D^2 + C_0) b_s = 0, \tag{31}$$

where $b_s = 0$ at $s = s_{ib}$, 1.

Recall that $C_2 = O(d^4)$, which is very small when our assumptions apply. The WKBJ method can therefore be used. It is well known that the WKBJ solution of (31) breaks down at the zeros of C_0 and these play a crucial role in what follows.

As an example of an ideal instability, we consider the case $m=1$, $\lambda=1$, $\Lambda=2.5\times10^5$ and $n=500$. Numerical calculation gives $p=0.367+1.553i$. Substituting these values into the definition of $C_0(s)$, we find $C_0(s)\neq0$ for $s_{ib}\leqslant s\leqslant1$ but, if we continue (31) into the complex s-plane, we see that $C_0(s)$ possesses two *complex* zeros, s_1 and s_2 (see Figure 10). These zeros are usually called "turning points" in WKBJ theory. The anti-Stokes lines pass through the turning points and are defined by (Heading, 1962, Ch. IV)

$$\mathrm{Re}\left(i\int_{s_i}^{s}\left(\frac{C_0}{C_2}\right)^{1/2}ds\right)=0,\qquad i=1,2. \tag{32}$$

They divide the complex s-plane into four sectors: M_1, M_2, M_3 and M_4 (see Figure 10). It may be seen that the anti-Stokes lines are almost straight (except near the turning points, where they have no significance for the WKBJ method, and very far from the turning points; see Appendix A). For general values of p, a solution that is exponentially small in sector M_2 is exponentially large in sector M_4, and conversely. It is only for special values of p (the eigenvalues sought) that a solution exists that is exponentially small in both these sectors. It may also be noted that such a solution will necessarily be asymptotically large in sectors M_1 and M_3.

Figure 10 shows that, in the special case we are studying, the boundary points $s=s_{ib}$ and $s=1$ lie in the sectors M_4 and M_2 respectively. Since the boundary conditions (15) demand that b_s vanishes at $s=s_{ib}$ and $s=1$, we require precisely the special values of p referred to above for which solutions to (31) are asymptotically small in sectors M_2 and M_4. In confirmation, we note that the exact solution of (13)–(15) shown on Figure 10 lies very precisely within the sector M_1, where the solution is supposed to be large according to the asymptotic theory.

These observations show that the eigenvalues and eigenfunctions of ideal instability theory are determined by the double turning points in the complex s-plane. In the neighborhood of these points the WKBJ solutions are invalid, but another "turning point equation" applies whose solutions are intimately connected to the WKBJ solutions. To obtain this, we note that the turning points are close to one another in the complex s-plane. We may therefore in their neighborhood approximate $C_0(s)/C_2(s)$ by a second degree polynomial in $s-s_0$:

$$Q(s)\equiv\frac{C_0(s)}{C_2(s)}\approx Q(s_0)+Q'(s_0)(s-s_0)+\frac{1}{2}Q''(s_0)(s-s_0)^2. \tag{33}$$

We may choose s_0 to be the midpoint between the turning points: $s_0=\frac{1}{2}(s_1+s_2)$. Because $Q(s_1)=Q(s_2)=0$, we have

$$Q'(s_0)=0. \tag{34}$$

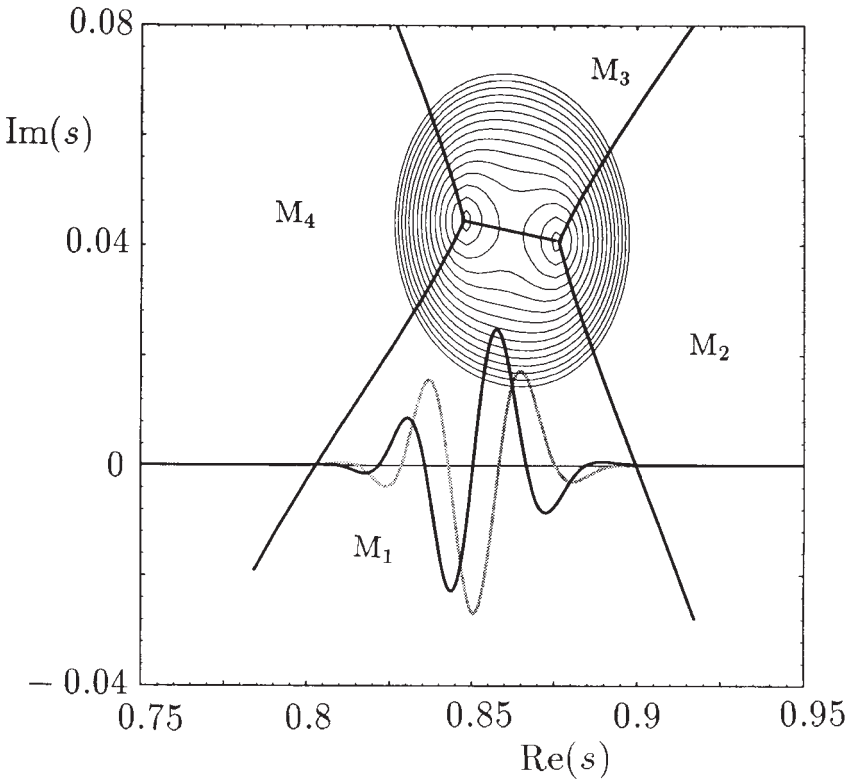


Figure 10 Contours of $|C_0(z)|$ are plotted for $m=1$, $\lambda=1$, $\Lambda=2.5\times10^5$, $n=500$ and $p=0.367+1.553i$ for which there are two turning points, z_1 and z_2. The horizontal axis is Re(s) from 0.75 to 0.95; the vertical axis is Im(s) from -0.04 to 0.08. The anti-Stokes lines divide the complex s-plane into four sectors, M_i, $i=1-4$. The critical level ($s=0.675$) and the physical boundaries ($s=0.35, 1$) are located in the sectors M_2 and M_4 where the eigenfunctions are exponentially small. The eigenfunction $b_s(s)$ on the real axis is also shown where now the ordinate gives the values of $\mathrm{Re}[b_s(s)]$ (black curve) and $\mathrm{Im}[b_s(s)]$ (grey curve). They are completely confined to the sector M_1.

With these approximations, equation (31) becomes (in a small neighborhood surrounding $s=s_1$ and $s=s_2$)

$$D^2 b_s + \left[Q(s_0) + \frac{1}{2} Q''(s_0)(s-s_0)^2 \right] b_s = 0. \tag{35}$$

The requirement that the WKBJ solutions be small in sectors M_2 and M_4 translates into the same demand on the solutions of the turning point equation (35). This is the condition that singles out the eigenvalues, p.

Equation (35) is a scaled version of the parabolic cylinder equation, the standard form of which is

$$\frac{d^2y}{dz^2}+\left(a-\frac{1}{4}z^2\right)y=0. \tag{36}$$

The crucial property of (36) that we require is that solutions exist that are exponentially small both for $z\to+\infty$ and $z\to-\infty$ only if $a=k+\frac{1}{2}$, where k is a non-negative integer, the solutions then being

$$y=H_k(z)\exp\left(-\frac{1}{2}z^2\right), \qquad k=0,1,2,\ldots, \tag{37}$$

where $H_k(z)$ are the Hermite polynomials. [See, for example, Section 16.7 of Whittaker and Watson (1962).] We shall be principally concerned with the case $k=0$ for which $H_0=1$ and $a=\frac{1}{2}$. Its anti-Stokes lines are straight, except near the turning points $z=\pm 2\sqrt{a}$; see, for example, Olver (1959).

We see that, to make the solutions of (35) asymptotically small in sectors M_2 and M_4 and therefore to obtain the required WKBJ solutions, we must have

$$\frac{Q(s_0)}{\sqrt{[-2Q''(s_0)]}}=\left(k+\frac{1}{2}\right), \qquad k=0,1,2,\ldots, \tag{38}$$

where the square root with positive real part must be chosen.

Our asymptotic method obtains p by iteration. We first make an estimate of s_0, and we then solve for p from (34). By (24) and the definition of Q, this simply involves finding the root with larger real part of a quadratic in p. Third, we use condition (38) to improve the estimate of s_0. The second and third steps are then repeated until s_0 has converged. Table 2 lists a few examples of results obtained from the double turning point analysis, and compares them with the numerical solutions of the full equations. Quantitative agreement is found in every case in which the assumptions stated at the beginning of the Section are satisfied. Because of the particular dependence of p on $Q(s)$, the eigenvalues with the largest real part always correspond to $k=0$ in (38).

From (24), (33), (34) and (38) with $k=0$, one finds that $|s_1-s_2|=O(d)$. This is another reason why Λ, $n\gg 1$ are such important assumptions. They ensure that the turning points are so close to each other ($d\ll 1$) that (33) is a good approximation in their neighborhood, and they also ensure that the part of real s-axis within the sector M_1 is so short that the solution is localized, a basic requirement for the validity of the local analysis (see Section 4).

Table 2 Eigenvalues obtained by asymptotic methods

λ	m	Λ	n	K	p	*Turning points*
1	1	10^4	10^2	1	**0.2875 + 1.5233*i***	0.8293 + 0.04684*i*
					0.2823 + 1.5218*i*	0.8929 + 0.03833*i*
1	1	10^6	10^3	1	**0.3779 + 1.5571*i***	0.8519 + 0.04396*i*
					0.3779 + 1.5571*i*	0.8722 + 0.04137*i*
1	1	10^6	10^2	10^{-2}	**1.3369 + 1.5347*i***	0.8383 + 0.04332*i*
					1.3365 + 1.5344*i*	0.8811 + 0.04237*i*
1	1	10^8	10^2	10^{-4}	**1.3478 + 1.5347*i***	0.8384 + 0.04323*i*
					1.3474 + 1.5344*i*	0.8809 + 0.04253*i*
1	2	10^4	10^2	1	**0.6300 + 2.2329*i***	0.7908 + 0.04345*i*
					0.6219 + 2.2255*i*	0.8528 + 0.03956*i*
1	2	10^6	10^3	1	**0.7919 + 2.3451*i***	0.8162 + 0.04318*i*
					0.7916 + 2.3453*i*	0.8361 + 0.04195*i*
1	2	10^6	10^2	10^{-2}	**1.6703 + 2.2317*i***	0.7926 + 0.04153*i*
					1.6672 + 2.2284*i*	0.8482 + 0.04287*i*
1	2	10^8	10^2	10^{-4}	**1.6808 + 2.2315*i***	0.7927 + 0.04150*i*
					1.6777 + 2.2284*i*	0.8482 + 0.04292*i*
2	1	10^4	10^2	1	**2.4697 + 3.2553*i***	0.5904 + 0.02106*i*
					2.4439 + 3.2481*i*	0.6279 + 0.01733*i*
2	1	10^6	10^3	1	**2.7602 + 3.3535*i***	0.6042 + 0.01983*i*
					2.7598 + 3.3537*i*	0.6162 + 0.01868*i*
2	1	10^6	10^2	10^{-2}	**3.6122 + 3.2741*i***	0.5942 + 0.01959*i*
					3.6089 + 3.2722*i*	0.6220 + 0.01928*i*
2	1	10^8	10^2	10^{-4}	**3.6245 + 3.2739*i***	0.5943 + 0.01956*i*
					3.6214 + 3.2722*i*	0.6219 + 0.01934*i*
2	2	10^4	10^2	1	**4.0183 + 5.1845*i***	0.5731 + 0.01964*i*
					3.9718 + 5.1448*i*	0.6125 + 0.01864*i*
2	2	10^6	10^3	1	**4.6612 + 5.5547*i***	0.5905 + 0.01981*i*
					4.6607 + 5.5547*i*	0.6032 + 0.01946*i*
2	2	10^6	10^2	10^{-2}	**5.1177 + 5.1660*i***	0.5737 + 0.01910*i*
					5.0912 + 5.1471*i*	0.6104 + 0.01997*i*
2	2	10^8	10^2	10^{-4}	**5.1288 + 5.1657*i***	0.5737 + 0.01910*i*
					5.1025 + 5.1471*i*	0.6104 + 0.01999*i*

Note: The highlighted values of p were obtained from the numerical solution of the full equations in which N was successively taken to be 250, 500 and 1000, the resulting eigenvalues being then Richardson extrapolated. The values beneath those highlighted were derived from the double turning point analysis.

The eigenvalue problem posed by the parabolic cylinder equation is well known in physics from the analysis of the quantum harmonic oscillator. A plot such as Figure 11 is similar to those commonly seen in quantum mechanics for a particle in a potential well. Analogously, the relationship between the double turning point analysis and the local analysis for the ideal instability problem is much the same as that between quantum mechanics and classical mechanics for the harmonic oscillator problem.

Although critical levels are present in the example analyzed in this Section, the method developed here applies with equal force to configurations that have no critical levels.

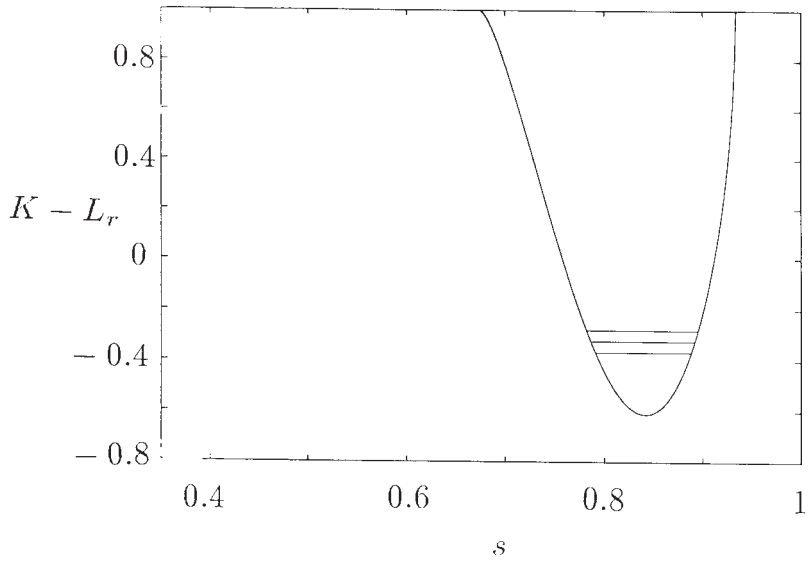


Figure 11 Curve showing $-p_r = K - L_r(s)$ as a function of s as given by (29) in the case $m = 1$, $\lambda = 1$, and $K = 1$. The three horizontal lines indicate the growth rates obtained from the double turning point analysis for (from bottom to top) $k = 0, 1, 2$, when $n = 500$ and $\Lambda = 2.5 \times 10^5$. The horizontal axis extends from $s = 0.35$ to $s = 1$. This Figure displays the analogy between our double turning point analysis and the solution to Schrödinger's equation for the harmonic oscillator in quantum mechanics.

6. CONCLUSIONS

We have investigated the linear stability of the simple zonal field (19) in a rapidly rotating (cylindrical) system. Although we have, in four cases ($m = 1, 2$ and $\lambda = 1, 2$), explored numerically all ranges of the Elsasser number, Λ, and the wave number, n, parallel to the rotation axis, we have mainly focused on the lightly damped case of large Λ. We have confirmed the existence of the resistive modes studied by Fearn, which dominate the ideal modes for small n, and have derived the same asymptotic law as him for the growth rate: $p \propto \Lambda^{-1/3}$ for $\Lambda \to \infty$. We have seen, as he and Braginsky saw, that the solutions are concentrated at critical levels. Our main focus has however, been on the ideal modes. We have found, as Braginsky (1980) did, that the ideal modes are highly localized when Λ and n are large. This leads us to hope that the usefulness of our findings will transcend the one dimensional model from which they were derived. We hope that they will eventually prove to be a useful guide in more complicated, but geophysically more realistic, situations in which the container is spherical and the zonal field depends on both s and z.

Using the same basic approximation as Acheson, namely $\Lambda \gg 1$ and $n \gg 1$, we have succeeded in extending his criterion to situations in which critical layers are present, and in Section 4 we have presented a method of generating simply a necessary

condition for stability. Although this technique leads to qualitatively correct answers, it is possible, as we showed in Section 5, to deduce quantitatively much better results, ones that in principle give precisely the correct growth rate, p, for $\Lambda \to \infty$. The method we use, which is closely related to the one developed by Soward and Jones (1983), expresses the eigenfunction in WKBJ form on the real s-axis, where it relates to the real physical problem, and connects that to two turning points in the complex s-plane. Associated with those turning points is a simple equation, which is a scaled and rotated version of the parabolic cylinder equation. From this euation, which is valid in only a small domain in the complex s-plane in our problem, the growth rate could be deduced with comparative simplicity and accuracy.

Braginsky (1980) investigated the ideal modes numerically. He too found that they have highly localized structures, which he called "wave packets". The location of a packet was not very obviously connected to the spatial structure of the field, but (as with our solutions above) the mode was vanishingly small outside the packet itself. Although he studied an equation related to the parabolic cylinder equation, which is the key equation of our study, he did not emphasize the role of the complex s-plane, the turning points or the anti-Stokes lines in delineating the range of the packet, and could not therefore provide an analytic argument for its location and structure.

Acknowledgements

We are grateful to the National Science Foundation for an award (EAR-8846267) which at various times supported each of us.

References

Acheson, D. J., "On the hydromagnetic stability of a rotating fluid annulus," *J. Fluid Mech.* **52,** 592–541 (1972).

Acheson, D. J., "On the instability of toroidal magnetic fields and differential rotation in stars," *Phil. Trans. R. Soc. Lond.* **A289,** 459–500 (1978).

Acheson, D. J., "Local analysis of thermal and magnetic instabilities in a rapidly rotating fluid," *Geophys. Astrophys. Fluid Dynam.* **27,** 123–136 (1983).

Braginsky, S. I., "Magnetohydrodynamics of the Earth's core," *Geomag. Aeron.* **4,** 898–916 (1964a) [Transl. 698–712].

Braginsky, S. I., "Self-excitation of a magnetic field during the motion of a highly conducting fluid," *Zh. Eksp. Teoret. Fiz.* (*SSSR*) **47,** 1084–1098 (1964b) [Transl. *Soviet Phys. JETP* **20,** 726–735 (1965)].

Braginsky, S. I., "Magnetic waves in the Earth's core," *Geomag. Aeron.* **7,** 1050–1060 (1967) [Transl. 851–859].

Braginsky, S. I., "Magnetic waves in the Earth's core, II," *Geophys. Astrophys. Fluid Dynam.* **14,** 189–208 (1980).

Braginsky, S. I. and Roberts, P. H., "Magnetic field generation by baroclinic waves," *Proc. R. Soc. Lond.* **A347,** 125–140 (1975).

Condi, F. J., "Sidewall boundary layers in rapidly rotating convection," Woods Hole Oceanographic Institution Technical Report WHOI-78-67, Vol. II, pp 18–24 (1978).

Fearn, D. R., "Boundary conditions for a rapidly rotating hydromagnetic system in a cylindrical container," *Geophys. Astrophys. Fluid Dynam.* **25,** 65–75 (1983a).

Fearn, D. R., "Hydromagnetic waves in a differentially rotating annulus I. A test of local stability analysis," *Geophys. Astrophys. Fluid Dynam.* **27,** 137–162 (1983b).

Fearn, D. R., "Hydromagnetic waves in a differentially rotating annulus II. Resistive instabilities," *Geophys. Astrophys. Fluid Dynam.* **30,** 227–239 (1984).

Fearn, D. R., "Hydromagnetic waves in a differentially rotating annulus IV. Insulating boundaries," *Geophys. Astrophys. Fluid Dynam.* **44,** 55–75 (1988).
Fearn, D. R. and Weiglhofer, W. S., "Resistive instability and the magnetostrophic approximation," *Geophys. Astrophys. Fluid Dynam.* **63,** 111–138 (1992).
Golub, G. H. and van Loan, C. F., *Matrix Computations*, 2nd Edition, Johns Hopkins University Press, Baltimore (1989).
Heading, J., *An Introduction to Phase-Integral Methods*, Methuen & Co., London (1962).
Huerre, P. and Monkowitz, P. A., "Local and global instabilities in spatially developing flows," *Ann. Rev. Fluid Mech.* **22,** 473–537 (1990).
Kreiss, H.-O., "Difference approximations for boundary and eigenvalue problems for ordinary differential equations," *Math. of Comp.* **26,** 605–624 (1972).
Kuang, W. and Roberts, P. H., "Resistive instabilities in rapidly rotating fluids: Linear theory of the tearing mode," *Geophys. Astrophys. Fluid Dynam.* **55,** 199–239 (1990).
Kuang, W. and Roberts, P. H., "Resistive instabilities in rapidly rotating fluids: Linear theory of the g-mode," *Geophys. Astrophys. Fluid Dynam.* **60,** 259–333 (1991).
Lifschitz, A. E., *Magnetohydrodynamics and Spectral Theory*, Kluver, Dordrecht (1989).
Olver, F. W. J., "Uniform asymptotic expansions for Weber parabolic cylinder functions of large orders," *J. Res. Nat. Bureau Stand.* **63B**, 131–169 (1959).
Roberts, P. H. and Loper, D. E., "On the diffusive instability of some simple steady magnetohydrodynamic flows," *J. Fluid Mech.* **90,** 641–668 (1979).
Soward, A. M. and Jones, C. A., "The linear stability of the flow in the narrow gap between two concentric rotating spheres," *Quart. J. Mech. Appl. Math.* **36,** 19–43 (1983).
Whittaker, E. T. and Watson, G. N., *Modern Analysis*, 4th Edition, Cambridge University Press, Cambridge U.K. (1962).

APPENDIX A: ASYMPTOTIC JUSTIFICATIONS

We claimed in Section 5 that the fourth order term $C_4D^4b_s$ in (25) is unimportant in determining the eigenvalue. We shall now justify this statement by asymptotic methods.

As stated in Section 5, if we drop the fourth order term in (23) and make the approximation (33) for $|s-s_0|\ll 1$, the eigenvalue problem involves the parabolic cylinder equation, the required solution of which ($k=0$) is of the form

$$b_s = e^{\alpha(s-s_0)^2}. \tag{39}$$

Here $\alpha = -\frac{1}{4}\sqrt{[-2Q''(s_0)]} = O(d^{-2})$. It follows from (39) that

$$D^4b_s = [12\alpha^2 + 48\alpha^3(s-s_0)^2 + 16\alpha^4(s-s_0)^4]b_s. \tag{40}$$

Easy calculations demonstrate that, when $|s-s_0|\ll 1$, $C_4D^4b_s \ll |C_0b_s|$ in (25), except in the immediate neighborhood of the turning points s_1 and s_2. More precisely, the fourth order term is negligible if $\varepsilon^2 d \ll |s-s_i| \ll 1$, where $\varepsilon \equiv \Lambda^{-1}d^{-3} \ll 1$. There may therefore exist inner boundary layers around the turning points.

Let $\xi = (s-s_i)/\delta$ near the turning points $s=s_i$ ($i=1$ or 2), where $\delta = \varepsilon d$ is the thickness of the boundary layer; note $\delta \ll d \ll 1$. The solution of the turning point equation is of the form

$$b_s = d_1 + d_2\xi + d_3 e^{\kappa\xi} + d_4 e^{-\kappa\xi}, \tag{41}$$

where $\kappa=(-\delta^2 C_2/C_4)^{1/2}$ may be taken to be an O(1) constant because s is very near the turning point s_i; d_1, d_2, d_3 and d_4 are undetermined constants.

The solutions (39) and (41) are required to match within the overlap domain in which they are both valid: $\varepsilon^2 d \ll |s-s_i| \ll 1$. It is found that, to leading order, $d_1=\exp[\alpha(s_i-s_0)^2]$ and that $d_2=d_3=d_4=0$.

These results show that the fourth order term is essentially irrelevant to the determination of the eigenvalue. That term does, however, affect the eigenfunctions when $|s-s_0| \gtrsim 1$, and therefore changes the form of the anti-Stokes lines (see Figure 10). Instead of (32), the anti-Stokes lines should be defined, when s is far from the turning points s_1 and s_2, by

$$\mathrm{Re}\int_{s_i}^{s}\left[\frac{-C_2+(C_2^2-4C_0C_4)^{1/2}}{2C_4}\right]^{1/2} ds=0, \qquad i=1,2. \tag{42}$$

When $|s-s_0| \ll 1$, (42) tends to (32) and the anti-Stokes lines coincide with those of the (scaled and rotated) parabolic cylinder equation, which are straight except near the turning points.

APPENDIX B: DIFFERENTIAL ROTATIONS

This Appendix generalizes Section 4 to situations in which a zonal differential rotation, $\mathbf{V}_0$, is present that is aligned with the prevailing zonal field, $\mathbf{B}_0$. With an appropriately defined pressure, P_0, this, as previously, defines an equilibrium state over the τ_s time scale, i.e. it satisfies (1) when $\eta=0$. Diffusive evolution of the state can, as before, be ignored in discussing waves and instabilities whose time scales are short compared with τ_η. Again, we study linear stability by perturbing as in (4), with the obvious difference that now $\mathbf{V}=\mathbf{V}_0+\mathbf{v}$, where $\mathbf{V}_0=V_0(s)\hat{\boldsymbol{\phi}}$. Then $(5)_2$ is replaced by

$$(\partial_t-\Lambda^{-1}\nabla^2)\mathbf{b}=\nabla\times(\mathbf{v}\times\mathbf{B}_0)+\nabla\times(\mathbf{V}_0\times\mathbf{b}). \tag{43}$$

The boundary conditions are unchanged.

The subsequent derivation of the basis equations goes through much as in Section 2. An added term, $(-imV_0/s)b_s$, appears on the left-hand side of (9) and (13), and one more term, $(-imV_0/s)b_z$, is present on the left-hand side of (10) and (14).

The form of (23) is unchanged but the leading order analytic expressions of C_2 and C_0 in (24) are changed to

$$\begin{aligned} C_2&=\left(-2p_K+i\frac{4mB_0^2}{s^2}-i\frac{2mV_0}{s}\right)\Lambda^{-1}-\frac{m^4B_0^4}{n^2s^4},\\ C_0&=p_K^2-i\frac{4mB_0^2}{s^2}p_K-\frac{2m^2B_0^3B_0'}{s^3}+\frac{(m^2-2)m^2B_0^4}{s^4}\\ &\qquad +i\frac{2mV_0}{s}p_K+\frac{m^2B_0^2}{s^2}V_0'+\frac{3m^2B_0^2V_0}{s^3}-\frac{m^2V_0^2}{s^2}. \end{aligned} \tag{44}$$

As a result, (27) is replaced by

$$p_K = \frac{mB_0^2}{s^2}\left\{2i \pm \sqrt{\Delta - m^2\left(1+\frac{l^2}{n^2}\right)}\right\} - i\frac{mV_0}{s} - \frac{l^2}{\Lambda}, \tag{45}$$

where the Δ defined in (21) is abandoned in favor of

$$\Delta = \frac{s^3}{B_0^2}\frac{d}{ds}\left(\frac{B_0^2}{s^2} - \frac{V_0}{s}\right). \tag{46}$$

It follows that the form (29) of L_r is unchanged although the Δ it contains is different.

It is easily seen that, when no critical points are present in $s_{ib} < s < 1$, criterion (21) holds as before with (22) replaced by (46); see Acheson (1983). It then follows that an eastward zonal motion that increases outward, or a westward differential rotation that decreases outward tends to stabilize field gradient instability. When critical points are present, we may use (45) in the same way as we used (27) in Section 4: a necessary condition for instability is that the maximum value of $L_r - K$ be positive.

The double-turning point analysis for a system with nonzero differential rotation is, *mutatis mutandis*, the same as in Section 5, as is the whole procedure for obtaining a quantitatively accurate estimate of the growth rate, p, when Λ and n are large.

15 April 1996

PHYSICS LETTERS A

ELSEVIER

Physics Letters A 213 (1996) 59–64

Structure and stability of a spherical implosion

P.H. Roberts [a], C.C. Wu [b]

[a] *Department of Mathematics, University of California, Los Angeles, CA 90095, USA*
[b] *Department of Physics, University of California, Los Angeles, CA 90095, USA*

Received 14 August 1995; revised manuscript received 29 December 1995; accepted for publication 22 January 1996
Communicated by M. Porkolab

Abstract

Similarity solutions and the stability for strong spherical implosions are studied for both ideal and van der Waals gases. When the van der Waals excluded volume is sufficiently large, a new type of solution is found and the shock may be linearly stable. Implications for inertial confinement fusion and sonoluminescence are discussed.

PACS: 47.40.−x; 43.25.+y; 47.20.−k; 47.40.Nm
Keywords: Spherical implosion; Shock stability; Inertial confinement fusion; Sonoluminescence

The structure and stability of shocks generated by spherical and cylindrical implosions in an ideal gas have received much attention in the past decade. The prime impetus has come from plasma fusion research where such studies are relevant to target design for ICF (inertial confinement fusion) [1]. More recently it has been conjectured [2] that such shocks are an essential part of the mechanism responsible for sonoluminescence (SL), i.e. the light which under certain conditions is emitted from a bubble of gas trapped in a liquid and compressed by incident spherically symmetric sound waves [3]. We found [2] that the density of air in an SL bubble in water is increased during compression until it approaches that of water and the assumption of an ideal gas became suspect. We therefore modeled the air by a simple van der Waals law and found that, as the moment $t = 0$ of implosion approached, the solution became closer and closer to a similarity form [4] in which the radius, $R_S(t)$, of the shock and other variables are proportional to $(-t)^{\alpha}$, where α is a function of the ratio of specific heats, γ, and of the van der Waals excluded volume, b. Similarity solutions are ubiquitous [5]. It has been conjectured recently [3] that the instability of the imploding shock provides an upper limit at which SL is quenched, and we therefore study the stability of the similarity solution below.

Fusion research has also inspired several experimental [6] and theoretical [7] studies of cylindrical and spherical implosions, and the stability of the resulting shock waves. The theoretical analyses of shock stability are linear and are based on the so-called "CCW approximation", so named after Chester, Chisnell and Whitham [8] (see also Ref. [9]). We find below that the CCW approximation performs badly unless the gas is close to ideal. Indeed, it misses a branch of solutions of an entirely different structure, this being the only branch that exists when b is sufficiently large. We show that shocks are made more stable if b or n increases, where n is the spherical harmonic number of the perturbation. The imperfect (van der Waals) shock is apparently more stable than the

PII S0375-9601(96)00082-5

corresponding shock in an ideal gas, but the most dangerous modes of instability are, in both cases, those of longest wavelength. Since sonoluminescence has recently been observed [10] in gases, such as ethane, for which γ is close to unity, we pay special attention to the limit $\gamma \to 1$, in which both the similarity solution and its stability can be solved analytically.

The basic equations are the Euler equations

$$\frac{\partial \rho}{\partial t} + \nabla\cdot(\rho \boldsymbol{v}) = 0, \tag{1}$$

$$\frac{\partial E}{\partial t} + \nabla\cdot[(E+p)\boldsymbol{v}] = 0, \tag{2}$$

$$\frac{\partial}{\partial t}(\rho \boldsymbol{v}) + \nabla\cdot(\rho \boldsymbol{v}\boldsymbol{v}) = -\nabla p, \tag{3}$$

where ρ is the gas density, $\boldsymbol{v}$ is the fluid velocity, p is the pressure and $E = \rho e + \frac{1}{2}\rho v^2$ is the total energy density, e being the internal energy density. We assume that the gas obeys a van der Waals equation of state of the form

$$p = \frac{\mathcal{R}T}{\mathcal{V} - b}, \quad e = c_v T = \frac{\mathcal{V} - b}{\gamma - 1}p,$$
$$S = c_v \ln[p(\mathcal{V} - b)^\gamma] + \text{const}, \tag{4}$$

where $\mathcal{V} = 1/\rho$ is the specific volume, T is the temperature, S is the specific entropy, $\mathcal{R}$ is the gas constant, and $c_v = \mathcal{R}/(\gamma - 1)$ is the specific heat at constant volume. It should be noted that this is a restricted form of the van der Waals equation for which the molecular attraction constant is zero. If the molecular attraction constant were nonzero, making the pressure nonmonotonic as a function of density, serious questions would arise as to whether the new mode propagates in a thermodynamically unstable state with negative compressibility. These questions are not addressed in this Letter.

Eqs. (1)–(4) break down at a shock front, where they are replaced by the Rankine–Hugoniot conditions. For the study of the implosion, we shall use $\boldsymbol{\xi} = \boldsymbol{x}/A_i(-t)^\alpha$ as an independent variable in place of $\boldsymbol{x}$ and write

$$R_S(t) = A_i(-t)^\alpha, \quad \rho = \rho_0 G(\boldsymbol{\xi}, t),$$
$$a^2 = \left(\frac{\alpha r}{t}\right)^2 Z(\boldsymbol{\xi}, t), \quad \boldsymbol{v} = \frac{\alpha r}{t}\boldsymbol{V}(\boldsymbol{\xi}, t), \tag{5}$$

where a is the sound speed and ρ_0 is the density ahead of the shock.

Eqs. (1)–(4) admit spherically symmetric similarity solutions (G_0, V_0, and Z_0) which depend only on $\xi = |\boldsymbol{\xi}|$; $\boldsymbol{V}$ is radial, the shock front is $\xi = 1$, and the shock velocity is $U = -\alpha R_S/t$. Eqs. (1)–(3) imply

$$(1 - V_0)\frac{\xi}{G_0}\frac{dG_0}{d\xi} = \xi\frac{dV_0}{d\xi} + 3V_0, \tag{6}$$

$$(1 - V_0)\xi\frac{dZ_0}{d\xi} + 2\left(\frac{1}{\alpha} - V_0\right)Z_0 = \frac{\gamma - 1 + 2\tilde{b}G_0}{1 - \tilde{b}G_0}Z_0\left(\xi\frac{dV_0}{d\xi} + 3V_0\right), \tag{7}$$

$$(1 - V_0)\xi\frac{dV_0}{d\xi} + \left(\frac{1}{\alpha} - V_0\right)V_0 = \frac{1 - \tilde{b}G_0}{\gamma - 1 + 2\tilde{b}G_0}\left(\xi\frac{dZ_0}{d\xi} + 2Z_0 + \kappa(1 - 2\tilde{b}G_0)\frac{Z_0}{1 - V_0}\right), \tag{8}$$

where $\kappa = 2(1-\alpha)/\alpha\gamma$ and $\tilde{b} = \rho_0 b$. When the shock has passed through the origin it becomes an outgoing shock, which again has the form (5) with the same value of α, but with t replacing $-t$ and a different constant A_0 replacing A_i; v is positive behind the shock but is still negative ahead of the shock.

Given values of V_0, Z_0 and G_0, Eqs. (6)–(8) determine $\xi\,dV_0/d\xi$, $\xi\,dZ_0/d\xi$, and $\xi\,dG_0/d\xi$ uniquely except where the determinant of the coefficients of these derivatives vanishes which, as Eqs. (7) and (8) show, occurs on the crucial surface $Z_0 = (1-V_0)^2$ in $V_0Z_0G_0$-space. For such values of V_0 and Z_0, the right-hand of these equations require that $[3V_0 - \kappa(1 - \tilde{b}G_0)]Z_0 = V_0(1 - V_0)(1/\alpha - V_0)$, a second crucial surface in $V_0Z_0G_0$-space. The intersection of these two crucial surfaces defines "critical curves". It is readily found by an application of de l'Hôpital's rule that, at any point on a critical curve, dZ_0/dV_0 can only take one of two possible values, and a solution curve in $V_0Z_0G_0$-space that crosses from one side of the crucial surfaces to the other must, if it is to be physically realistic (i.e., continuous and single-valued), cross the crucial surfaces at a critical curve with one of these two values of dZ_0/dV_0 [4]. This requirement closes the mathematical problem and determines α. For an ideal gas or when $\tilde{b}$ is small, solutions cross one of these curves; we call this the *Guderley branch*. For large $\tilde{b}$ they cross the other curve, so defining a *new branch* of solutions. For intermediate $\tilde{b}$ both types of solution coexist. Fig.

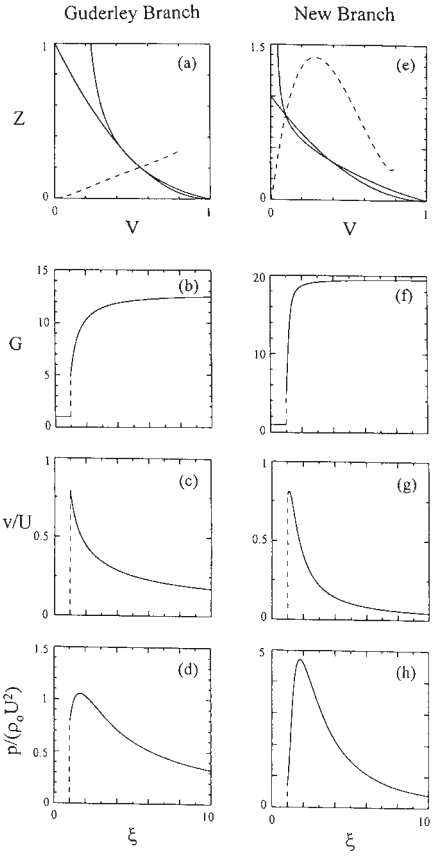


Fig. 1. Comparison between a similarity solution of Guderley type ((a)–(d)) for $\gamma = 1.4$ and $\tilde{b} = 0.05$, and the corresponding new similarity solution ((e)–(h)); see text.

Table 1
Summary of Integrations for the case $\gamma = 1.4$

$\tilde{b}$	α	ρ_c/ρ_0	A_0/A_i	ρ_{ou}/ρ_0	ρ_{od}/ρ_0	M_o
0.0 (G)	0.717	20.1	0.492	64.2	145.	1.73
0.05(G)	0.661	12.6	2.05	15.1	17.9	2.01
0.05(N)	0.494	19.5	6.44	19.6	19.9	2.88
0.25(N)	0.534	3.70	3.66	3.77	3.91	1.94
0.8 (N)	0.554	1.22	2.99	1.22	1.24	1.49

I shows such a case: $\gamma = 1.4$ and $\tilde{b} = 0.05$, which corresponds to an initially well compressed state. (It should perhaps be remarked here that in calculations for sonoluminescence [2] the surface of the imploding bubble compresses the trapped air to high densities before it creates the imploding shock which cuts off the interior of the bubble from the exterior.) Panels (a)–(d) are for the Guderley solution; panels (e)–(h) are for the new solution. In panels (a) and (d) the solid curves are the intersections of the two critical surfaces with the V_0Z_0-plane; the two intersections of these curves mark the intersection of the critical curves with the V_0Z_0-plane. The dashed curves are the solutions, which clearly cross different critical curves. Panel (a) resembles that of Refs. [4] but panel (e) does not. Panels (b)–(d) and (f)–(h) show ρ, v and p. The compression of the gas is greater for the new solution than for the Guderley solution; compare panels (b) and (f). In fact, the densities in the former approach closely the limit $\rho_\infty \equiv \rho_0/\tilde{b} = 20\rho_0$ set by the van der Waals excluded volume. The greater densities in the new solution are paralleled by greater pressure at $\xi = \infty$; compare panels (d) and (h).

Some properties of the $\gamma = 1.4$ solutions are summarized in Table 1; G stands for Guderley and N for new. The third column of this table gives the central compression, $\rho_c/\rho_0 = G$, at the moment of focussing. The last three columns refer to the outgoing shock, and give the upstream (ρ_{ou}) and downstream (ρ_{od}) densities at the shock front; M_o is the Mach number based on the upstream fluid velocity and sound speed in the shock frame. The larger the α, the faster (for the same R and t) the shock moves. The constant A_o/A_i determines whether the outgoing shock moves slower ($A_o/A_i < 1$) or faster ($A_o/A_i > 1$) than the ingoing shock at the same radius. In the case of the ideal gas, it moves more slowly; in the other cases shown, it moves more rapidly.

The new branch of solutions is important because it covers a much larger range of $\tilde{b}$ than the Guderley branch and is the only one that exists for $\tilde{b} \neq 0$ when $\gamma \to 1$. When $\gamma = 1.4$, a Guderley type solution exists for $0 \leqslant \tilde{b} < 0.08$ and the new solution in $0.05 \leqslant \tilde{b} \leqslant 1$. For $\gamma = 1.2$, the Guderley branch only exists over a tiny range, $0 \leqslant \tilde{b} \leqslant 0.0001$; the new branch was found for $0.05 \leqslant \tilde{b} \leqslant 1$. We did not obtain a similarity solution in the range $0.0001 \leqslant \tilde{b} < 0.05$. Compu-

tational difficulties arise when γ and $\tilde{b}$ are small. For example, a satisfactory solution for $\gamma = 1.2$ and $\tilde{b} = 0.0001$ requires α to be determined to 7 decimal place accuracy. (For an ideal gas, α has been determined to more than a dozen places, see e.g. Ref. [9]; however, we have not tried to be that accurate for $\tilde{b} \neq 0$ in this paper.) Asymptotic analysis shows that, in the limit $\gamma \to 1$ (for $\tilde{b} \neq 0$),

$$\alpha = 2/(5 - \tilde{b}). \tag{9}$$

There is a strong motive for studying the similarity shock in the limit $\gamma \to 1$. Sonoluminescence has recently been observed [10] from a bubble of ethane, a gas in which the ratio of specific heats, γ, is close to unity. If ethane were an ideal gas, the shock heating would vanish for $\gamma \to 1$, for then $p/\rho \propto T = \text{const.}$ But for a van der Waals gas, ρ_2 is limited by $1/b$; then $T_2 \propto p_2/\rho_2 \sim M^2 b \rho_1 \propto M^2$ for $M \to \infty$; the subscripts 1 and 2 denote, respectively, the upstream and downstream of a shock. It is therefore reasonable to believe that shock heating, due to the imperfections of ethane, can explain its ionization and the emission of light.

We perturbed the self-similar solution and solved the resulting linearized perturbation equations numerically. We set

$$G = G_0(1 + G_1 f),$$
$$V = V_0 + \left(V_1 f, W_1 \frac{\partial f}{\partial \theta}, \frac{W_1}{\sin\theta}\frac{\partial f}{\partial \phi}\right),$$
$$Z = Z_0 + Z_1 f, \tag{10}$$

where G_1, V_1, W_1 and Z_1 are perturbations that are functions of ξ, $f = S_n(\theta, \phi)(-t)^{\alpha\beta}$ and S_n is a spherical harmonic of order n; β is "the complex decay rate", which is an eigenvalue.

There are three solutions of no physical interest: (A) The difference between two of the original similarity solutions corresponding to infinitesimally different A_i provides a $\beta = 0$ mode for $n = 0$; (B) The difference between two of the original similarity solutions corresponding to infinitesimally different t provides a $\beta = -1/\alpha$ mode for $n = 0$; (C) The difference between two of the original similarity solutions separated by infinitesimally different origins provides a $\beta = -1$ mode for $n = 1$.

The stability of the similarity solutions in the limit $\gamma \to 1$ can be obtained by asymptotic analysis. After a lengthy analysis [12], we found that solutions (A) and (B) are the only discrete modes for $n = 0$ but that for all $n \geqslant 1$ and all $\tilde{b}$ there is an infinite discrete spectrum of eigenmodes which are the roots of the transcendental equation

$$n - 2 - \beta + \tfrac{1}{3}(n+1)(1+\beta) \times \tilde{b}^{-\mu}(1-\tilde{b})^{1+\mu-n/3} B_{1-\tilde{b}}(\tfrac{1}{3}n - \mu, \mu) = 0, \tag{11}$$

where $\mu = \frac{1}{6}(2\beta + \tilde{b} - 1)$ and $B_x(a, b)$ is the incomplete beta function, which is defined for $\mathrm{Re}(a) > 0$ by

$$B_x(a,b) = \int_0^x t^{a-1}(1-t)^{b-1}\,\mathrm{d}t, \quad 0 \leqslant x < 1, \tag{12}$$

and which can be analytically continued to all complex a and b, apart from nonpositive integral a. It is conjectured (but not proved) that the eigenvalues determined from (11) are complete, i.e., that the corresponding eigenfunctions form a complete set, in terms of which the solution to the initial value problem for any n ($\leqslant 1$) and $\tilde{b}$ can be expressed. Values of β satisfying (11) are shown in Fig. 2. They correspond to the most unstable modes (MUM), i.e., the mode or modes for which $\mathrm{Re}\,\beta$ is smallest. The β for MUM is complex when $\tilde{b}$ or n is small, apart from the $n = 1$ mode which is direct ($\mathrm{Im}\,\beta = 0$). As $\tilde{b}$ approaches 1, the β for MUM is real for all n, and all are positive for $\tilde{b} > 0.4654$. The shock is then linearly stable to all non-spherically-symmetric perturbations. The oscillations in $\mathrm{Re}\,\beta$ and $\mathrm{Im}\,\beta$ as $\tilde{b} \to 0$ for every $n \geqslant 2$ are real; they arise from the complex exponent, μ, in $\tilde{b}^{-\mu}$ in (11).

Similar trends were discovered for other γ by numerical integration. The real and imaginary parts of β are shown in Fig. 3. Dashed curves refer to solutions of Guderley type and the solid curves to the new solutions. The real part is positive for $\tilde{b}$ greater than about 0.7. The new $\tilde{b} = 0.05$ solution does not appear to conform with the others, apparently because of an oscillation analogous to those noted above for $\gamma \to 1$. Application of the WKBJ method shows that, for the ideal gas, β is purely imaginary as $n \to \infty$,

$$\beta \sim \pm i n \sqrt{\frac{\gamma - 1}{\gamma + 1}}, \quad n \to \infty. \tag{13}$$

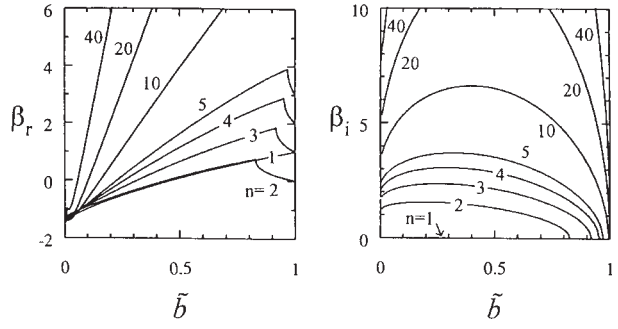


Fig. 2. The complex decay rate β, of the most unstable mode of the shock front as a function of $\tilde{b}$ for different harmonic numbers n in the limit $\gamma \to 1$.

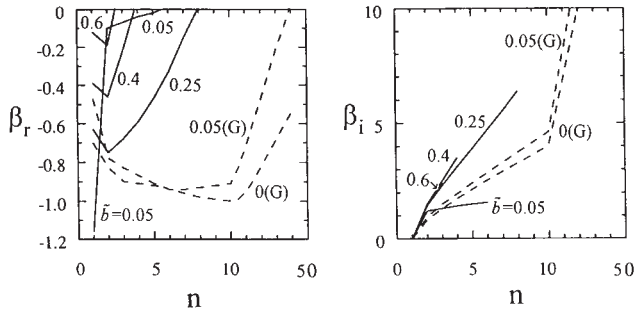


Fig. 3. The complex decay rate β, of the most unstable mode of the shock front as a function of n for different $\tilde{b}$; in all cases, $\gamma = 1.4$. Dashed curves: solutions of Guderley type; solid curves: new solutions.

The Chester–Chisnell–Whitham (CCW) approximation [8] has been used with success to solve for both the structure and the linear stability of the Guderley solution in an ideal gas [7]. In the CCW approximation, the influence of the characteristics which are overtaking the shock from behind is ignored. We found [12] that the CCW approximation performs increasingly poorly as $\tilde{b}$ is increased. For example for $\gamma = 1.4$ it gives $\alpha = 0.717$ for $\tilde{b} = 0$, $\alpha = 0.589$ for $\tilde{b} = 0.25$, and $\alpha = 0.486$ for $\tilde{b} = 0.8$. The last two values differ substantially from those shown in Table 1. In the limit $\gamma \to 1$, the CCW approximation gives $\alpha = \frac{1}{3}$ for $\tilde{b} \neq 0$, which is very different from (9). For $\tilde{b} = 0$, the CCW approximation gives $\alpha = 1$ in the limit $\gamma \to 1$. Thus a nonuniformity arises in this double limit $\gamma \to 1$, $\tilde{b} \to 0$. The CCW approximation provides only two modes of the stability problem for each n; it gives incorrect β for the van der Waals gas. In the limit $\gamma \to 1$, it gives $\beta = -1.5 \pm 0.5\,[9 - 4n(n+1)]^{1/2}$ for $\tilde{b} \neq 0$. Accordingly $\mathrm{Re}\,\beta = -1.5$ for $n \geqslant 2$ independent of both n and $\tilde{b}$ and negative (instability). This is inconsistent with (11) and the results shown in Fig. 2.

One disappointing feature of the linear stability analysis is that it gives no information about $n = 0$ modes. In particular, in parameter domains where both the Guderley and the new solutions exist, it does not determine which, if either, is unstable. We therefore numerically integrated (1)–(4) for $\gamma = 1.4$

both for initial states close to the Guderley solution and for initial states close to the new similarity solution. We followed their evolution for as long as was numerically feasible. Remarkably, we discovered no tendency for one solution to transform into the other. It appears that each is relatively stable with respect to the other.

A desire to understand better one aspect of the physics of sonoluminescence motivated our study. How far can our findings help in the search for configurations that provide greater luminosities or, in cases where deuterium and/or tritium is present, that result in greater fusion rates [11]? It has been conjectured [3] that a limit to these processes is set by the stability of the imploding shock. And it is natural to seek ways of stabilizing the shocks. The replacement of the air in the bubble by gaseous D_2 or DT is potentially less promising than the use of compounds, such as ethane, but in which D and/or T atoms replace H atoms [2]. The greater atomic weight of such compounds lowers the sound speed and leads to greater compressions during the implosion. This idea gains additional support from the present study, which shows that shocks in a material, such as ethane, are more stable than they would be in gaseous D_2 or DT, which have a larger γ. The idea may also be found useful in the design of fuel capsules for ICF. The following point is also of interest. During the acoustic compression of the bubble, the trapped gas is first compressed adiabatically. The shock that later forms cuts off the central part of the bubble from its exterior, and central compression is halted. It now appears, as a result of the present study, that it is desirable to prolong the initial adiabatic compression for as long as possible, in order that the density, ρ_0, ahead of the shock, and hence $\tilde{b}$, are as large as possible, so promoting the stability of the shock. Such a prolongation might be achieved by shaping the wave form of the acoustic driver. It may also be recalled that, in many situations of practical interest, the compression of the bubble is accompanied by two shocks in succession [2]. The second of these, which travels across the gas pre-compressed by the first shock, is usually responsible for the greater light emission and, in the thermonuclear application, for the larger fusion rate. The greater value of $\tilde{b}$ for the second shock implies that it will also be the more stable. Our discussions on the similarity solutions and their stability for strong spherical implosions in this paper are based on a simple equation of state. Thus in the ICF and SL applications, the results are valid at low temperatures when the equation of state is applicable. Further studies of shock implosions are required at very high temperatures when changes in the equation of state occur.

We thank Seth Putterman for the suggestion that similarity solutions of Guderley type would exist for the van de Waals gas. CCW was supported by NSF grant ATM-92-01662 and NASA grant NAGW-2848. Computing facilities were provided by the San Diego Supercomputer Center, which is sponsored by the NSF.

References

[1] M.D. Cable, S.P. Hatchett, J.A. Caird, J.D. Kilkenny, H.N. Kornblum, S.M. Lane, C. Laumann, R.A. Lerche, T.J. Murphy, J. Murray, M.B. Nelson, D.W. Phillion, H. Powell and D.B. Ress, Phys. Rev. Lett. 73 (1994) 2316.
[2] C.C. Wu and P.H. Roberts, Phys. Rev. Lett. 70 (1993) 3424; Proc. R. Soc. A 445 (1994) 323.
[3] S.J. Putterman, Sci. Am. 272 (1995) 46.
[4] G. Guderley, Luftfahrtforschung 19 (1942) 302;
L.D. Landau and E.M. Lifshitz, Fluid mechanics, 2nd Ed. (Pergamon, Oxford, 1987) §89.
[5] S. Abarbanel and M. Goldberg, J. Comput. Phys. 10 (1972) 1;
H. Matsuo, Y. Ohya, K. Fujiwara and H. Kudoh, J. Comput. Phys. 75 (1988) 384;
Y. Nakamura and H. Matsuo, Mem. Fac. Eng. Kumamoto Univ. 34 (1989) 161;
H.P. Greenspan and A. Nadim, Phys. Fluids A 5 (1993) 1065.
[6] D.P. Singh, M.A. Harith, V. Palleschi, G. Tropiano, M. Vaselli, N. Salingaros and E. Panarella, Laser Part. Beams 8 (1990) 253;
M. Watanabe and K. Takayama, JSME Int. (Ser. II) 35 (1992) 218.
[7] J.H. Gardner, D.L. Book and I.B. Bernstein, J. Fluid Mech. 114 (1982) 41;
C.R. DeVore and E.S. Oran, Phys. Fluids 4 (1992) 835.
[8] W. Chester, Philos. Mag. 45 (1954) 1293;
R.F. Chisnell, J. Fluid Mech. 2 (1957) 286;
G.B. Whitham, J. Fluid Mech. 2 (1957) 145.
[9] M. Yousaf, J. Fluid Mech. 66 (1974) 577.
[10] S.J. Putterman, private communication.
[11] B.P. Barber, C.C. Wu, R. Löfstedt, P.H. Roberts and S.J. Putterman, Phys. Rev. Lett. 72 (1994) 1380.
[12] C.C. Wu and P.H. Roberts, Structure and stability of a spherical shock wave in a van der Waals gas, submitted to Q. J. Mech. Appl. Math.

Thermal inertial waves in a rotating fluid layer: Exact and asymptotic solutions

Keke Zhang[a] and Paul H. Roberts
Institute of Geophysics and Planetary Physics, University of California, Los Angeles, California 90095-1567

(Received 25 September 1996; accepted 19 March 1997)

The problem of the onset of convective instability in a rapidly rotating fluid layer heated from below with various velocity boundary conditions is investigated by constructing exact solutions and by asymptotic analysis. It is shown that convective motions at sufficiently small Prandtl numbers are described in leading order by a thermal inertial wave. It is at the next order that buoyancy forces drive the wave against the weak effect of viscous dissipation. On the basis of the perturbation of the thermal inertial wave, asymptotic convection solutions for rigid boundaries can be expressed in simple analytic form. A new asymptotic power law between the critical Rayleigh number and the Ekman number is derived. In addition, solutions for both stationary and time-dependent convection are calculated numerically. Comparison between the numerical and asymptotic solutions is then made to show that a satisfactory quantitative agreement has been achieved. © *1997 American Institute of Physics.* [S1070-6631(97)01907-7]

I. INTRODUCTION

Thermal convection in a rotating fluid layer heated from below has been widely used to illustrate in a simple way the effect of rotation in various astrophysical and geophysical contexts. The linear stability problem was first studied by Chandrasekhar;[1] in what follows we shall however refer to his well-known monograph;[2] see also Fearn, Roberts, and Soward[3] and Busse.[4] We also find it convenient to introduce Rayleigh, Ekman, and Prandtl numbers:

$$R=\frac{g\alpha\beta d^2}{\Omega\kappa},\quad E=\frac{\nu}{\Omega d^2},\quad Pr=\frac{\nu}{\kappa}. \tag{1}$$

Here ν is the kinematic viscosity of the fluid, κ its thermal diffusivity, α is its coefficient of thermal expansion, all assumed constant so that the temperature gradient across the layer, β, is uniform; the depth of the layer is d, the acceleration due to gravity is g (also uniform) and the layer rotates about the upward vertical with angular velocity Ω. The horizontal wave number, a, of the convective planform is measured in units of d^{-1} and the frequency, ω, of oscillatory convection is measured in units of ν/d^2. The Rayleigh number defined in (1) is smaller, by a factor of E, than that employed by Chandrasekhar, who also used a Taylor number, $T=4E^{-2}$, in favor of E.

Chandrasekhar showed that, when both boundaries of the layer are stress-free, the eigenfunctions of the linear stability problem are very simple, and lead to an equally simple exact analytic solution. For steady convection between free boundaries, he established that, in the limit of high rotation, the critical Rayleigh number R_c (the smallest value of R for which convection can occur), and the corresponding wave number a_c, are to leading order,

$$R_c\sim 3(2\pi^2)^{2/3}E^{-1/3},\quad a_c\sim(2\pi^2)^{1/6}E^{-1/3},\quad \text{as } E\to 0. \tag{2}$$

The origin of the power law (2) was discussed in detail by Chandrasekhar.[2] The role of viscosity is inverted: instead of being purely dissipative, its presence is required to offset the rotational constraint and so allow convection to take place.

When one or both of the bounding surfaces are rigid surfaces, on which the flow satisfies the "no-slip" boundary conditions, the eigenfunctions are no longer simple. Chandrasekhar derived numerical results with the help of a variational principle. From these he argued that (2) also holds when one or both boundaries are rigid, but he believed that the constants of proportionality are different. That this is not the case was clearly brought out by Roberts,[5] who explained the apparent discrepancy through the omitted second terms in the expansions (2) which, relative to the leading terms, are $O(E^{1/6})$. For the values of E that Chandrasekhar attained in his variational calculations, this "correction term" was large enough to mislead him. What is more significant in the context of the present paper is that, to dominant order, the nature of the boundaries has no effect on the critical mode (2).

Chandrasekhar[2] showed that the simple exact solution for stress-free boundaries also describes oscillatory convection, and leads to asymptotic laws of the same form as (2) but with constants of proportionality that depend on Pr:

$$R_c\sim\frac{6(2\pi^2)^{2/3}Pr^{4/3}}{(1+Pr)^{1/3}}E^{-1/3},$$

$$a_c\sim\frac{(2\pi^2)^{1/6}Pr^{1/3}}{(1+Pr)^{1/3}}E^{-1/3}, \tag{3}$$

$$\omega_c\sim\pm\frac{(2\pi^2)^{1/3}(2-3Pr^2)^{1/2}}{Pr^{1/3}(1+Pr)^{2/3}}E^{-2/3},\quad \text{as } E\to 0.$$

It is therefore clear that an oscillatory mode can exist only if Pr is sufficiently small, but that its critical Rayleigh number may then be considerably less that that of the steady mode

[a]Permanent and corresponding address: Department of Mathematics, University of Exeter, EX4 4QJ, United Kingdom. Telephone: 44-1392-263992; Fax: 44-1392-263997. Electronic mail: KZhang@maths.exeter.ac.uk

 1070-6631/97/9(7)/1980/8/$10.00

(2). Nevertheless, it cannot be zero for $Pr=0$, and in fact (2) is in flat contradiction with the result of Chandrasekhar (1961, Sec. 32) for that case:

$$R_c=\frac{27\pi^4}{2}E, \quad a_c=\frac{\pi}{\sqrt{2}},$$

$$\omega_c=\pm 2\left(\frac{2}{3}\right)^{1/2}E^{-1}, \quad \text{for } Pr=0. \tag{4}$$

Noting this, Chandrasekhar cautioned that (3) is valid not for $E\to 0$ but for $\tau\to\infty$, where

$$\tau=\frac{Pr}{E}=\frac{\kappa}{\Omega d^2}. \tag{5}$$

Comparison of (3) and (4) demonstrates that the limit $E\to 0$ is singular, and it highlights the need for further asymptotic analysis, valid for $E\to 0$ for fixed τ. We anticipate (and show) that the resulting asymptotic expressions for R_c, a_c and ω_c tend continuously to (3) for $\tau\to\infty$ and continuously to (4) for $\tau\to 0$. The new asymptotic theory uncovers some interesting fluid behaviors. The motion is to leading order an inviscid inertial wave, which carries the temperature and associated density differences passively. We call it a *thermal inertial wave*. At next order the buoyancy force maintains the wave against the weak viscous dissipation that would otherwise obliterate it. In contrast to the modes on which Chandrasekhar concentrated, which we shall simply call *convective modes*, the inertial mode releases the rotational constraint through its fast oscillation and not by viscosity, which consequently plays its usual dissipative role.

The primary objective of the paper is to examine the existence and preference of oscillatory convection in the form of a thermal inertial wave in a rapidly rotating fluid layer. We have performed three analyses to achieve this objective. First, we suppose that both boundaries are stress-free. By perturbing the thermal inertial wave, we obtain complete agreement with the exact results of Chandrasekhar:[2] for fixed τ,

$$R_c=O(E), \quad a_c=O(1), \quad \omega_c=O(E^{-1}), \quad \text{as } E\to 0. \tag{6}$$

This provides some confidence in the validity of the asymptotic method. Second, we consider the case when the lower boundary is rigid and the upper boundary is stress-free. We again obtain the exact solution, but it is no longer simple. We present results for a number of cases of both stationary and oscillatory convection. As a byproduct we find that Chandrasekhar's variational results are generally satisfactory. Third, still in the context of one rigid and one free boundary, we generalize our asymptotic method and obtain simple asymptotic solutions, which show satisfactory quantitative agreement with the solutions obtained numerically. In sharp contrast to the convection modes, where the transition to rigid boundaries does not affect the leading term (3), the asymptotic laws governing the thermal inertial modes take new forms in dominant order,

$$R_c=O(E^{1/2}), \quad a_c=O(1), \quad \omega_c=O(E^{-1}), \quad \text{as } E\to 0. \tag{7}$$

This is because the viscous dissipation in the main body of the layer, which is all important in the case of free boundaries, is negligible compared with dissipation in the Ekman layer at the rigid wall.

We do not consider the case when both boundaries are rigid, since the solution in that case can be obtained by a symmetry argument from the solution obtained when one wall is rigid and the other free.

In what follows we shall begin by presenting the mathematical formulation of the problem (Sec. II). Asymptotic solutions for two stress-free boundaries are discussed in Sec. III. This is followed (Sec. IV) by a discussion of the exact solution for one stress-free and one rigid boundary. The corresponding asymptotic solution is presented in Sec. V. We close (Sec. VI) the paper with a few remarks.

II. MATHEMATICAL FORMULATION OF THE PROBLEM

The equations governing convection in a rotating Boussinesq fluid may be written in the form (Chandrasekhar, 1961)

$$\partial_t \boldsymbol{u}+\boldsymbol{u}\cdot\nabla\boldsymbol{u}+2\boldsymbol{\Omega}\times\boldsymbol{u}=-\rho^{-1}\nabla\hat{p}+g\alpha\hat{\theta}\mathbf{1}_z+\nu\nabla^2\boldsymbol{u}, \tag{8a}$$

$$\nabla\cdot\boldsymbol{u}=0, \tag{8b}$$

$$(\partial_t-\kappa\nabla^2)\hat{\theta}=\beta\mathbf{1}_z\cdot\boldsymbol{u}-\boldsymbol{u}\cdot\nabla\hat{\theta}, \tag{8c}$$

where ρ is density, $\hat{p}$ and $\hat{\theta}$ are the departures in pressure and temperature from those of the static conduction solution, $\boldsymbol{u}$ is fluid velocity, ∂_t is the Eulerian time derivative and $\mathbf{1}_q$ is the unit vector in the direction of increasing coordinate q, here the direction of the upward vertical z. Equations (8a)–(8c) are nondimensionalized as follows:

$$\boldsymbol{x}\to\boldsymbol{x}d, \quad t\to td^2/\nu, \quad \hat{\theta}\to\hat{\theta}\beta d\nu/\kappa,$$

$$\hat{p}\to\hat{p}\rho\nu\Omega, \quad \boldsymbol{u}\to\boldsymbol{u}\nu/d. \tag{9}$$

On neglecting the nonlinear terms in (8), we find that the nondimensional equations governing the onset of convection are

$$E\partial_t\boldsymbol{u}+2\mathbf{1}_z\times\boldsymbol{u}=-\nabla\hat{p}+R\hat{\theta}\mathbf{1}_z+E\nabla^2\boldsymbol{u}, \tag{10a}$$

$$\nabla\cdot\boldsymbol{u}=0, \tag{10b}$$

$$(Pr\partial_t-\nabla^2)\hat{\theta}=\mathbf{1}_z\cdot\boldsymbol{u}. \tag{10c}$$

According to (10b), the velocity $\boldsymbol{u}$ may be expressed as the sum of poloidal and toroidal parts:

$$\boldsymbol{u}=\nabla\times\nabla\times[\hat{\phi}(x,y,z,t)\mathbf{1}_z]+\nabla\times[\hat{\psi}(x,y,z,t)\mathbf{1}_z]. \tag{11}$$

Equations (10a) and (10c) permit solutions of the form

$$[\hat{\phi},\hat{\psi},\hat{\theta},\hat{p}]=[\phi(z),\psi(z),\theta(z),p(z)]e^{i(a_x x+a_y y+\omega t)}, \tag{12}$$

where a_x and a_y are horizontal wave numbers and ω is frequency. After applying the operators $\mathbf{1}_z\cdot\nabla\times$ and $\mathbf{1}_z\cdot\nabla\times\nabla\times$ to (10a) and using expressions (11) and (12), we obtain

$$E(D^2-a^2)\psi-iE\omega\psi+2D\phi=0, \tag{13a}$$

$$E(D^2-a^2)^2\phi-iE\omega(D^2-a^2)\phi-2D\psi-R\theta=0, \tag{13b}$$

$$(D^2-a^2)\theta - i\omega Pr\,\theta + a^2\phi = 0. \tag{13c}$$

The differential operator D and the total horizontal wave number a are defined as

$$D = \frac{d}{dz}, \quad a^2 = a_x^2 + a_y^2. \tag{14}$$

To complete the convection problem, we must specify boundary conditions on the walls $z=0$ and $z=1$. At a stress-free, impenetrable boundary ϕ and ψ must satisfy

$$\phi = D^2\phi = 0, \quad \psi = 0. \tag{15a}$$

We refer to such a surface as "free." At a non-slip, impenetrable boundary we require

$$\phi = D\phi = 0, \quad \psi = 0. \tag{15b}$$

We refer to such a surface as "rigid." We further assume that the temperatures of the upper and lower boundaries are held constant, so that

$$\theta = 0 \quad \text{at } z = 0, 1. \tag{15c}$$

In this paper, three cases are considered: (a) $z=0$ and $z=1$ are both free; (b) $z=0$ is rigid and $z=1$ is free; and (c) $z=0$ and $z=1$ are both rigid.

III. THERMAL INERTIAL WAVES: STRESS-FREE BOUNDARIES

When both the bounding surfaces are free, exact solutions of simple analytic form are readily obtained for any Pr and E. The simplicity of the exact solutions provides a unique opportunity that allows us to demonstrate clearly and convincingly that convection at a sufficiently small Prandtl number is described at leading order by a thermal inertial wave.

The basic physical mechanism for the existence and preference of a thermal inertial wave is that convective motions may oscillate so rapidly that the rotational constraint is relaxed, and convection can take place. Viscosity is not required to release the constraint, and it plays only a dissipative role. This suggests that we may assume the following asymptotic expansion for $E \to 0$:

$$\psi = \psi_0 + \epsilon\psi_1, \tag{16a}$$

$$\phi = \phi_0 + \epsilon\phi_1, \tag{16b}$$

$$\theta = \theta_0 + \epsilon\theta_1, \tag{16c}$$

$$R = \epsilon(R_1 + \epsilon R_2), \tag{16d}$$

$$\omega = E^{-1}(\omega_0 + \epsilon\omega_1), \tag{16e}$$

where $\epsilon = E$ for the case of two free boundaries and

$$\psi_0 = O(1), \phi_0 = O(1), \theta_0 = O(1), \omega_0 = O(1). \tag{17}$$

Inserting (16a)–(16e) into (13a)–(13c), we find that to leading order,

$$-i\omega_0\psi_0 + 2D\phi_0 = 0, \tag{18a}$$

$$-i\omega_0(D^2-a^2)\phi_0 - 2D\psi_0 = 0, \tag{18b}$$

$$(D^2-a^2)\theta_0 - i\tau\omega_0\theta_0 + a^2\phi_0 = 0. \tag{18c}$$

We shall later require that $\tau \ll E^{-1}$ in order that a thermal inertial wave is the preferred mode of convection. The system of equations (18) is of lower order than the system (12), and its solution is constrained only by (15c) and the first of (15a) or (15b):

$$\phi_0 = 0, \quad \text{at } z = 0, 1. \tag{19}$$

Eliminating ψ_0 from (18a)–(18b), we derive the leading order eigenvalue problem: to determine ϕ_0 satisfying (19) and

$$D^2\phi_0 + \frac{\omega_0^2 a^2}{4-\omega_0^2}\phi_0 = 0. \tag{20}$$

It follows that

$$\omega_0^2 = \frac{4\pi^2}{a^2+\pi^2} \quad \text{and} \quad 0 < |\omega_0| < 2. \tag{21}$$

Without loss of generality, we may assume henceforward that $\omega_0 > 0$. The eigenfunctions corresponding to (21) are proportional to

$$[\phi_0, \psi_0, \theta_0, p_0] = \left[\sin \pi z, -\frac{2\pi i \cos \pi z}{\omega_0}, \frac{a^2 \sin \pi z}{a^2+\pi^2+i\tau\omega_0}, \frac{i\pi(4-\omega_0^2)\cos \pi z}{\omega_0}\right]. \tag{22}$$

In (21) and (22), we have chosen the eigenmode that has the simplest vertical structure since this is the one most readily excited to convection. Equation (22) determines how the isotherms move in response to the inertial wave. It is clear that, since the associated buoyancy force does not feature in (18b), the thermal field is completely passive at dominant order.

So far, we have determined only the eigenfrequency. To determine R_c and a_c we must carry the analysis to the next order. It is important to notice that the leading order solution (22) satisfies all the boundary conditions imposed at a stress-free boundary, even though only the condition of vanishing normal flow has been imposed. Generally speaking, an inviscid solution in other geometries (such as a sphere or a cylinder; see Greenspan[6]) cannot satisfy all the required boundary conditions at a stress-free boundary. A significant implication in the present case of free boundaries is that there are no Ekman layers at the walls. The next order problem is therefore

$$-i\omega_0\psi_1 + 2D\phi_1 = -(D^2-a^2)\psi_0 + i\omega_1\psi_0, \tag{23a}$$

$$-i\omega_0(D^2-a^2)\phi_1 - 2D\psi_1 = R_1\theta_0 - (D^2-a^2)^2\phi_0 + i\omega_1(D^2-a^2)\phi_0. \tag{23b}$$

In general these inhomogeneous equations have no solution. The solvability condition (the condition that solutions exist) is obtained by multiplying (23a) by ψ_0^*, the complex conjugate of ψ, multiplying (23b) by ϕ_0^*, summing the results, integrating over the layer, integrating by parts making use of (18a) and (18b), and applying the boundary condition (15) that $\phi_1 = 0$ on the walls. We obtain

$$R_1\int_0^1 \phi_0^*\theta_0 dz=\int_0^1[-\psi_0^*(D^2-a^2)\psi_0 + \phi_0^*(D^2-a^2)^2\phi_0]dz+i\omega_1 \times\int_0^1[|\psi_0|^2+a^2|\phi_0|^2+|D\phi_0|^2|]dz. \quad (24)$$

Real and imaginary parts of (24) determine the Rayleigh number R_1 and the frequency correction ω_1, which give the marginal Rayleigh number R and the corresponding frequency ω,

$$R=\frac{8E}{4-\omega_0^2}\left(\frac{16\pi^4}{\omega_0^4}+\tau^2\omega_0^2\right), \quad (25a)$$

$$\omega=E^{-1}\omega_0(1-Pr). \quad (25b)$$

For given E and Pr, this Rayleigh number determines the marginal state for a given ω_0 or equivalently [see (21)] given a. It is necessary to find the value ω_c of ω_0 that minimizes R. The resulting cubic has only one real root, and this gives

$$R_c=\frac{3\pi^4}{2}(1+s_1^2+s_2^2)^2E, \quad (26a)$$

$$a_c=\frac{\pi}{\sqrt{2}}(-1+s_1^2+s_2^2)^{1/2}, \quad (26b)$$

$$\omega_c=\frac{2}{E\sqrt{Q}}(s_1-s_2)^{1/2}, \quad (26c)$$

where

$$s_1=[\sqrt{(Q^2+1)}+Q]^{1/3}, \quad s_2=[\sqrt{(Q^2+1)}-Q]^{1/3}=s_1^{-1},$$

$$Q=2\tau/\pi^2. \quad (26d)$$

Equations (26) agree with (3) in the limit $\tau\to\infty$ and with (4) for $\tau\to 0$. For the parameter range $0<\tau\lesssim O(1)$, convection in the form of the thermal inertial wave is the most unstable mode and we have $Q=O(1)$ and $s_1=s_2=O(1)$, giving rise to

$$R_c=O(E), a_c=O(1).$$

Figure 1 displays both the exact and asymptotic solutions for different Prandtl numbers and for $E=10^{-4}$. Our asymptotic solutions are consistent with the exact solutions of Chandrasekhar[2] in the limit of small Prandtl number. For small $P_r(\lesssim 0.1)$, there are no noticeable differences between the asymptotic and exact solutions, but differences become large as Pr or τ is increased. When Pr is of order unity, convection becomes steady and is then independent of Pr, as discussed in detail by Chandrasekhar. This is also shown in Fig. 1.

The principal significance of our asymptotic analysis is that it convincingly demonstrates that the preferred mode of convection in the limit $E\to 0$ and $\tau\ll E^{-1}$ is indeed a thermal inertial wave. It holds the clear promise that a similar asymptotic analysis will succeed in more realistic cases in which one or both of the bounding surfaces are rigid, and when in consequence viscous dissipation in the Ekman boundary layer(s) must be included. Because of the mathematical complexity of finding exact solutions in such cases, an asymptotic solution of simple analytical form is not only practically useful but also shows explicitly how the solution depends on the governing physical parameters, so providing greater insight into the nature of the convective motions.

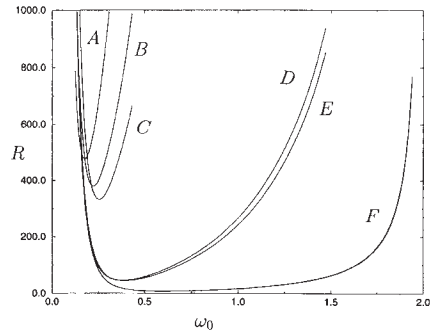


FIG. 1. The Rayleigh number at the onset of convection plotted as a function of frequency ω_0 at $E=10^{-4}$ for different Prandtl numbers. Both boundaries are stress-free. Line A is for steady convection; line B for the exact (oscillatory) solution at $Pr=0.5$ and line C for the corresponding asymptotic solution; line D for the exact solution at $Pr=0.1$ and line E for the corresponding asymptotic solution. Line F is for both the exact and asymptotic solutions at $Pr=0.025$ but their differences cannot be identified in the figure.

IV. EXACT SOLUTIONS: ONE RIGID AND ONE FREE BOUNDARY

In this case, variational solutions were obtained and discussed by Chandrasekhar.[2] Here we present the results of a direct analytical solution of the marginal convection problem, for both stationary and time-dependent modes. Our primary aim is to check the accuracy of our asymptotic solutions and to identify whether the most unstable mode is associated with the thermal inertial wave (22). The complexity of the exact solution will highlight the usefulness of our simple asymptotic solution. The secondary aim of this section is to use our solutions to assess the accuracy of Chandrasekhar's variational solution.

Let us first consider stationary convection ($\omega=0$). It can be readily shown that the general solution satisfying the stress-free boundary conditions (15a) and (15c) at $z=0$ has the form

$$\phi=A_1\sin zr+[A_2\sinh(z\sqrt{\rho}e^{i\theta/2})+\text{c.c.}], \quad (27a)$$

$$\theta=A_1\frac{a^2\sin zr}{a^2+r^2}+\left[A_2\frac{a^2\sinh(z\sqrt{\rho}e^{i\theta/2})}{a^2-\rho e^{i\theta}}+\text{c.c.}\right] + A_3\sinh az, \quad (27b)$$

$$\psi = A_1 \frac{2r\cos zr}{E(a^2+r^2)} + \left[A_2 \frac{2\sqrt{\rho} e^{i\theta/2} \cosh(z\sqrt{\rho} e^{i\theta/2})}{E(a^2-\rho e^{i\theta})} + \text{c.c.} \right] - A_3 \frac{R}{2a} \cosh az, \tag{27c}$$

where c.c. denotes the complex conjugate of the term preceding it, $[A_1, A_2, A_3]$ are complex constants and $-r^2, \rho e^{i\theta}$ and $\rho e^{-i\theta}$, where r, θ and ρ are real, are the three roots of the cubic equation,

$$E^2 Z^3 - 3E^2 a^2 Z^2 + (4 + 3E^2 a^4) Z + Ea^2 (R - Ea^4) = 0. \tag{28}$$

Substitution of the expressions for $[\phi, \psi, \theta]$ given by (27a)–(27c) into the four conditions (15b) and (15c) for the rigid surface at $z=1$ yields the "discriminant," a transcendental equation that determines the marginal R for given a and E:

$$\mathrm{Rl}\{\det M_{4\times 4}(E, a, R)\} = 0, \tag{29}$$

where $M_{4\times 4}$ is a 4th order complex square matrix. After satisfying (29), it is possible to solve the four linear equations for $[A_1, A_2, A_3]$ that led to (29). We may then, if we wish, derive the exact eigenfunctions from the expressions (27a)–(27c). The critical Rayleigh number, R_c, for given E is the smallest value of R for which (29) can be satisfied; the corresponding value, a_c, of a is the critical horizontal wave number.

For example, the most unstable mode for $E = 10^{-3}$ is given by $a_c = 15.2$ for which $R_c = 202.75$. At the smaller Ekman number $E = 5\times 10^{-4}$, we obtain $a_c = 19.4$ and $R_c = 252.98$, which is consistent with the asymptotic law (2). This suggests that stationary convection is then not the preferred form of convection in the small Prandtl number fluid that is the focus of this paper.

We compared our exact solutions with Chandrasekhar's variational results, and obtained good agreement. For example, for $E = 2\times 10^{-4} (T = 10^8)$ and $a = 26.55$, Chandrasekhar (Table 9, p. 105, Chap. 3) obtained $R_c = 343.46$, while our exact solution is $R_c = 342.56$.

Let us now consider the more complicated case of time-dependent (oscillatory) convection. The general solution satisfying the boundary conditions (15a) and (15c) at $z=0$ may be written as

$$\phi = \sum_{j=1}^{j=4} A_j \sinh \xi_j z, \tag{30a}$$

$$\theta = -\sum_{j=1}^{j=4} \frac{A_j a^2 \sinh \xi_j z}{\xi_j^2 - a^2 - i\omega Pr}, \tag{30b}$$

$$\psi = -\sum_{j=1}^{j=4} \frac{2A_j \xi_j \cosh \xi_j z}{E(\xi_j^2 - a^2 - i\omega)}, \tag{30c}$$

where the four complex constants A_j, $j = 1-4$, are to be determined and the complex numbers $\xi_j^2, j = 1,2,...,4$ are given by

$$\xi_j^2 = a^2 + \frac{1}{4} i[\omega(2+Pr) - S_1\Delta] + iS_2 \left\{ \frac{1}{16}[\omega(2+Pr) - S_1\Delta]^2 + Y_0 - \frac{1}{S_1\Delta} \left[\omega(2+Pr) Y_0 + \omega Pr \times \left(\omega^2 - \frac{4}{E^2} \right) - \frac{ia^2}{E^2}(4+ER) \right] \right\}^{1/2}, \tag{31}$$

where S_1 and S_2 take the values ± 1 independently, and

$$\Delta = \left[\left(\frac{16}{E^2} - 4\omega^2 Pr + \omega^2 Pr^2 \right) - 8Y_0 \right]^{1/2}.$$

In the preceding equations, Y_0 is any root of the cubic equation

$$Y^3 + a_2 Y^2 + a_1 Y + a_0 = 0.$$

The coefficients a_i are given by

$$a_2 = \frac{1}{2}\left(\omega^2 - \frac{4}{E^2} \right) + \omega^2 Pr,$$

$$a_1 = \frac{1}{4}\omega^2 Pr(2+Pr)\left(\omega^2 - \frac{4}{E^2} \right) - \frac{ia^2\omega}{4E^2} \times (8 - 12Pr - 2ER + PrER),$$

$$a_0 = \frac{1}{8}\left\{ \omega^2 Pr^2 \left(\omega^2 - \frac{4}{E^2} \right)^2 - \frac{a^4}{E^4}(4+ER)^2 - \frac{ia^2\omega}{E^2}\left[(4Pr+ER)\left(\frac{16}{E^2} - 4\omega^2 Pr + \omega^2 Pr^2 \right) + 2Pr(4+ER)\left(\omega^2 - \frac{4}{E^2} \right) \right] \right\}.$$

Substitution of the expressions (30a)–(30c) for $[\phi, \psi, \theta]$ into the four boundary conditions (15b) and (15c) at $z=1$ yields the discriminant

$$\det\{M_{4\times 4}(E, a, Pr, R)\} = 0. \tag{32}$$

The real and imaginary parts of this transcendental equation give the marginal $R(a, E, Pr)$ and $\omega(a, E, Pr)$. When (32) is satisfied, the four linear equations for $[A_1, A_2, A_3, A_4]$ that led to (32) can be solved. We may then, if we wish, determine the eigenfunctions from (30a)–(30c).

The values we obtain agree well with those of Chandrasekhar (Table 8, p. 102, Chap. 3). For example, for $E = 6.325\times 10^{-4}$ $(T = 10^6)$ and $a = 5.85$, he obtained $R = 10.81$ and $E\omega = 0.9126$; we found $R = 10.85$ and $E\omega = 0.9128$. For $E = 3.651\times 10^{-5}$ $(T = 3\times 10^9)$ and $a = 15.58$, he obtained $R = 17.56$ and $E\omega = 0.3801$; we found $R = 17.57$ and the same ω. Several examples for marginal R as a function of the oscillation frequency are shown in Fig. 2, together with the corresponding asymptotic values from the analysis of the next Section.

If $R(a, E, Pr)$ is minimized over a we obtain the critical mode, R_c, a_c and ω_c, for the given E and Pr. We have performed that minimization for two cases which are shown in Fig. 2.

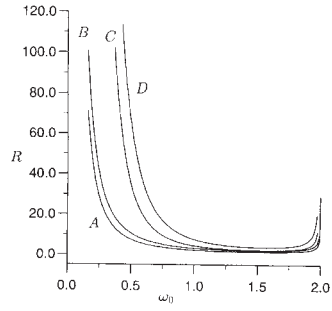


FIG. 2. The Rayleigh number at the onset of convection plotted as a function of frequency ω_0 for different Ekman numbers at $\tau=1$ for the case when one bounding surface is stress-free and the other is rigid. Line D is for the exact solution at $E=10^{-3}$ and line B for the corresponding asymptotic solution; line C is for the exact solution at $E=5\times10^{-4}$ and line A for the corresponding asymptotic solution.

V. THERMAL INERTIAL WAVES: RIGID BOUNDARIES

When either boundary is rigid, the no-slip conditions require that an Ekman layer is present on it. The viscous damping of the thermal inertial wave is now dominated by this layer; damping in the main body of the fluid is insignificant to a first approximation. As a result, the expansion parameter ϵ introduced in Sec. III is $E^{1/2}$ and not E.

Ekman layers usually arise in the context of flows that are slowly varying on the timescale of adjustment of the layers, Ω^{-1}. In the present case this is not true, and we therefore start by developing the necessary theory for a rapidly varying Ekman layer.

A. The Ekman boundary layer solution

The equation governing the flow $\boldsymbol{u}_b$ in an Ekman layer on $z=0$ may be written as

$$i\omega_0\boldsymbol{u}_b+2\mathbf{1}_z\times\boldsymbol{u}_b+(\mathbf{1}_z\cdot\nabla p_b)\mathbf{1}_z=E\partial_z^2\boldsymbol{u}_b, \tag{33}$$

where $\boldsymbol{u}_b$ and the associated pressure perturbation, p_b, are nonzero only in the layer whose thickness is $O(E^{1/2})$; $\partial_z\equiv\partial/\partial z$. The boundary flow $\boldsymbol{u}_b$ must satisfy the boundary condition

$$\boldsymbol{u}_b=\left[-\left(\frac{2a_y}{\omega_0}+ia_x\right)\mathbf{1}_x+\left(\frac{2a_x}{\omega_0}-ia_y\right)\mathbf{1}_y\right]\times\pi e^{i(a_xx+a_yy+\omega_0t)},\quad\text{on } z=0, \tag{34}$$

so that the no-slip conditions are obeyed by the sum of $\boldsymbol{u}_b$ and the mainstream, given by (2.4) and (3.7). Applying the operators $\mathbf{1}_z\times$ and $\mathbf{1}_z\times\mathbf{1}_z\times$ to (5.1) gives

$$(E\partial_z^2-i\omega_0)(\mathbf{1}_z\times\boldsymbol{u}_t)+2\boldsymbol{u}_t=0, \tag{35a}$$

$$(E\partial_z^2-i\omega_0)\boldsymbol{u}_t-2(\mathbf{1}_z\times\boldsymbol{u}_t)=0, \tag{35b}$$

where $\boldsymbol{u}_t=\boldsymbol{u}_b-(\mathbf{1}_z\cdot\boldsymbol{u}_b)\mathbf{1}_z$ is the tangential component of the boundary flow. Combining these two equations yields the fourth-order differential equation

$$(\partial_\eta^2-i\omega_0)^2\boldsymbol{u}_t+4\boldsymbol{u}_t=0, \tag{36}$$

where we have introduced a stretched boundary layer variable:

$$\eta=E^{-1/2}z, \tag{37a}$$

$$\partial_\eta\equiv\frac{\partial}{\partial\eta}=E^{1/2}\partial_z. \tag{37b}$$

It can be shown that the solution to (36) satisfying (34) is

$$\boldsymbol{u}_t=\left[\boldsymbol{C}_1\exp\left\{-\frac{(1+i)}{\sqrt{2}}\sqrt{(2+\omega_0)}\,\eta\right\}+\boldsymbol{C}_2\exp\left\{-\frac{(1-i)}{\sqrt{2}}\sqrt{(2-\omega_0)}\,\eta\right\}\right]e^{i(a_xx+a_yy+\omega_0t)}, \tag{38}$$

where the complex vectors $\boldsymbol{C}_1$ and $\boldsymbol{C}_2$ are

$$\boldsymbol{C}_1=\frac{\pi}{2\omega_0}(2-\omega_0)(a_x+ia_y)(i\mathbf{1}_x+\mathbf{1}_y), \tag{39a}$$

$$\boldsymbol{C}_2=-\frac{\pi}{2\omega_0}(2+\omega_0)(ia_x+a_y)(\mathbf{1}_x+i\mathbf{1}_y). \tag{39b}$$

In order to link the Ekman boundary layer flow to the thermal inertial wave, the normal component of the flow, W_b, at the edge of the lower boundary layer is needed. This can be written as (see Greenspan[6] for details)

$$W_b=-E^{1/2}\int_0^\infty\mathbf{1}_z\cdot\nabla\times(\mathbf{1}_z\times\boldsymbol{u}_t)\,d\eta. \tag{40}$$

Substituting from (38) and (39), we obtain

$$W_b=-\frac{\pi(2E)^{1/2}a^2}{4\omega_0}\left[(1+i)\frac{(2+\omega_0)}{\sqrt{(2-\omega_0)}}-(1-i)\frac{(2-\omega_0)}{\sqrt{(2+\omega_0)}}\right]e^{i(a_xx+a_yy-\omega_0t)}. \tag{41}$$

B. Asymptotic solutions

The solution again has the form (16), but with $\epsilon=E^{1/2}$. We will find that the thermal inertial mode is preferred over the convective mode provided that $\tau\ll E^{-1/2}$. While the zeroth order problem is still described by Eqs. (18a)–(18c), the first order equations are, in vector form,

$$i\omega_0\boldsymbol{u}_1+2\mathbf{1}_z\times\boldsymbol{u}_1+\nabla p_1=R_1\theta_0\mathbf{1}_z-i\omega_1\boldsymbol{u}_0, \tag{42a}$$

$$\nabla\cdot\boldsymbol{u}_1=0, \tag{42b}$$

which are subject to the velocity boundary condition

$$\mathbf{1}_z\cdot\boldsymbol{u}_1=E^{-1/2}W_b, \tag{43}$$

at the edge of the Ekman boundary layer. The solvability condition for the inhomogeneous Eq. (42a) can be obtained by the same procedure as that used in Sec. II. Instead, here

we carry out the equivalent steps of taking the scalar product of (42) with $\boldsymbol{u}_0^*$, integrating over the layer, using the divergence theorem and (42b), and applying the boundary condition (43) at $z=0$ together with $\mathbf{1}_z\cdot\boldsymbol{u}_1=0$ at the free surface $z=1$. With the help of the expressions (22), we obtain

$$-W_b p_0^*(z=0)=-R_1 a^2\int_0^1\theta_0\phi_0^* dz+i\omega_1\int_0^1|\boldsymbol{u}_0|^2 dz. \quad (44)$$

The real and imaginary parts of (44) determine R_1 and the correction, ω_1, to the frequency, which give the marginal Rayleigh number R and the corresponding frequency ω,

$$R=\frac{\sqrt{2E}\omega_0^2}{8\pi^2}\left[\frac{16\pi^4}{\omega_0^4}+(\tau\omega_0)^2\right]\times\left[\frac{(2+\omega_0)}{\sqrt{(2-\omega_0)}}+\frac{(2-\omega_0)}{\sqrt{(2+\omega_0)}}\right], \quad (45a)$$

$$\omega=\frac{\omega_0}{E}-\frac{\sqrt{2}(4-\omega_0^2)}{16\sqrt{E}}\left\{\frac{\tau\omega_0^3}{4\pi^2}\left[\frac{(2+\omega_0)}{\sqrt{(2-\omega_0)}}+\frac{(2-\omega_0)}{\sqrt{(2+\omega_0)}}\right]+\left[\frac{(2+\omega_0)}{\sqrt{(2-\omega_0)}}-\frac{(2-\omega_0)}{\sqrt{(2+\omega_0)}}\right]\right\}. \quad (45b)$$

The final step is to obtain R_c and a_c by minimizing the expression (45a) for R_1 over ω_c. The minimization of R over ω_0 in Eq. (45a) gives rise to an equation for ω_0 that is too complicated to be useful. But it is at once clear that, for fixed τ,

$$R_c=O(E^{1/2}),\quad a_c=O(1), \quad (46a)$$

$$\omega_c=O(E^{-1}),\quad \text{for } E\to 0. \quad (46b)$$

This is a new asymptotic law for the onset of convection in a rapidly rotating fluid layer heated from below. The complete asymptotic solution at leading order may be written as

$$\boldsymbol{u}=\boldsymbol{u}_t(x,y,z,t)+\left[\mathbf{1}_x\left(ia_x+\frac{2a_y}{\omega_0}\right)\pi\cos\pi z+\mathbf{1}_y\left(ia_y-\frac{2a_x}{\omega_0}\right)\times\pi\cos\pi z+\mathbf{1}_z a^2\sin\pi z\right]e^{i(a_x x+a_y y+\omega_0 t)}, \quad (47a)$$

$$\theta=\frac{a^2\sin\pi z}{a^2+\pi^2+i\tau\omega_0}e^{i(a_x x+a_y y+\omega_0 t)}, \quad (47b)$$

where $\boldsymbol{u}_t$ is given by Eq. (38).

Several examples of the asymptotic solutions are shown in Fig. 2. It should be noted that the asymptotic solutions obtained for a stress-free boundary at $z=1$ have the same critical Rayleigh numbers as those (Sec. IV) obtained with a stress-free boundary at $z=0$. The agreement between the exact and asymptotic solutions are satisfactory. The larger discrepancies at smaller frequencies (larger wave numbers) seen in Fig. 2 are not unexpected because our asymptotic analysis is valid for the most unstable mode only when $a=O(1)$. Note that the frequency ω is given to good accuracy by the leading order asymptotic value, the $O(E^{1/2})$ correction being small.

Asymptotic solutions for the case with two rigid boundaries require no extra analysis. As a consequency of symmetry of the thermal inertial wave solution (18a)–(18c) and the Ekman boundary layer solution (38) with respect to the mid plane $z=\frac{1}{2}$, the critical Rayleigh number R_c and the correction to the frequency for the case with two rigid boundaries are obtained by simply changing 8 to 4 in Eq. (45a), 16 to 8 in Eq. (45b), and replacing $\boldsymbol{u}_t(x,y,z)$ in Eq. (47a) by $\boldsymbol{u}_t(x,y,z)+\boldsymbol{u}_t(x,y,1-z)$.

In summary, when $E\ll 1$ and $\tau\ll E^{-1/2}$ the most unstable mode of convection is to leading order the thermal inertial wave (18a)–(18c). When one or both boundaries are rigid, the role of viscosity in this type of convection is only dissipative, and mainly takes place in the Ekman boundary layer(s). The critical Rayleigh number is $R_c=O(E^{1/2})$ instead of $R_c=O(E^{-1/3})$, as is the case for the convective modes. The simple nature of convection is paralleled by an equally simple asymptotic solution given by Eqs. (45a), (45b) and (47a) and (47b).

VI. CONCLUDING REMARKS

We have derived a relatively simple asymptotic solution for the onset of convection as overstability in a rapidly rotating fluid layer. This mode is preferred when the Prandtl number is sufficiently small. New asymptotic relations between R and E have also been derived, not only for the case when both boundaries are stress-free but also for the cases in which the flow at one or both boundaries satisfies the no-slip condition. Comparison between the asymptotic and exact solutions has shown a satisfactory quantitative agreement. Several issues in this classical problem that were not discussed by Chandrasekhar[2] have been clarified.

The existing experimental results appear to be consistent with our theory, although existing experiments were not particularly designed to investigate the thermal inertial waves discussed in the paper. A experimental study for a small Prandtl number (mercury, $P_r=0.025$) was carefully carried out by Rossby.[7] He found that, for Taylor numbers T_a between 1.8×10^4 and 10^5, his experimental results and Chandrasekhar's linear theory do not quite agree, but for the larger Taylor numbers, there is a good agreement. He also found that the frequencies of oscillation jump almost discontinuously at about $T_a=4\times10^4$, based on which he suggested that this may be an indication of a different type of instability. The different type of instability is likely to be the thermal inertial waves. Let us take $T_a=4\times10^4$ and note that the relationship between his Taylor number and our Ekman number is

$$E=2/\sqrt{T_a}$$

and

$$\tau=Pr\sqrt{T_a}/2.$$

If $Pr=0.025$ and $T_a=4\times10^4$, we have $\tau=2.5$. According to our theory, convection in the form of a thermal inertial wave that is characterized by a fast oscillation is the most unstable mode for this parameter range.

Finally, it should be observed that the geometry of the convecting region is significant. In recent studies of rotating spherical convection, Zhang,[8] showed that, when the Prandtl number is sufficiently small, the preferred mode of convection is a thermal inertial wave, obtained in leading order by solving the Poincaré equation. But there are fundamental differences between the spherical and plane-layer problems. Convective motions in rotating spherical systems are always in the form of travelling waves regardless of the value of the Prandtl number.[9] Because of the spherical geometry, the relevant solutions of Poincaré's equation are in the form of equatorially trapped waves.[10] The direction of wave propagation, i.e., whether it is eastward or westward, plays a key role in determining the marginal value of the Rayleigh number in the spherical problem, whereas in planar geometry the sign of ω has no significance. Furthermore, since the spherical Poincaré modes do not satisfy the conditions of zero tangential stress, Ekman layers are required even when the surface of the sphere is stress-free. These Ekman layers are, however, weaker, by a factor of $E^{1/2}$ than the Ekman layers that arise at no-slip walls, such as those considered in Sec. V. The Ekman pumping, W_b, is therefore $O(E)$, and not $O(E^{1/2})$. This pumping plays a key role in the solubility condition that determines R_1 and ω_1. The integrated viscous dissipation in a main body of fluid turns out to be zero; only that in the Ekman layer is nonzero, and requires a buoyancy source (proportional to R_1) to offset it.

ACKNOWLEDGMENTS

K.Z. would like to thank Professor G. Schubert for support and hospitality during his half-year sabbatical leave at the Institute of Geophysics and Planetary Physics, University of California, Los Angeles. P.R. was partially supported by an Australian Research Grant No. A69330372 and K.Z. is partially supported by the Leverhulme Trust.

[1] S. Chandrasekhar, "The instability of a layer of fluid heated below and subject to Coriolis Force," Proc. R. Soc. London Ser. A **217**, 306 (1953).

[2] S. Chandrasekhar, *Hydrodynamic and Hydromagnetic Stability* (Clarendon Press, Oxford, 1961).

[3] D. R. Fearn, P. H. Roberts, and A. M. Soward, "Convection, stability and the dynamo, energy, stability and convection," edited by B. Straughan and P. Galdi (Longman, New York, 1988), pp. 60–324.

[4] F. H. Busse, "Convection driven zonal flows in vortices in the major planets," Chaos **4**, 123 (1994).

[5] P. H. Roberts, "On the thermal instability of a self-gravitating fluid sphere containing heat sources," Philos. Trans. R. Soc. London Ser. A **263**, 93 (1965).

[6] H. P. Greenspan, *The Theory of Rotating Fluids* (Cambridge University Press, Cambridge, 1968).

[7] H. T. Rossby, "A study of Bénard convection with and without rotation," J. Fluid Mech. **36**, 309 (1969).

[8] K. Zhang, "On coupling between the Poincaré equation and the heat equation," J. Fluid Mech. **268**, 211 (1994).

[9] P. H. Roberts, "On the thermal instability of a highly rotating fluid sphere," Astrophys. J. **141**, 240 (1968).

[10] K. Zhang, "On equatorially trapped boundary inertial waves," J. Fluid Mech. **248**, 203 (1993).

BRIEF COMMUNICATIONS

The purpose of this Brief Communications section is to present important research results of more limited scope than regular articles appearing in Physics of Fluids. Submission of material of a peripheral or cursory nature is strongly discouraged. Brief Communications cannot exceed three printed pages in length, including space allowed for title, figures, tables, references, and an abstract limited to about 100 words.

The decay of bubble oscillations

P. H. Roberts
Department of Mathematics, University of California, Los Angeles, California 90095

C. C. Wu
Department of Physics and Astronomy, University of California, Los Angeles, California 90095

(Received 18 August 1998; accepted 27 August 1998)

We study the initial value problem posed by the small amplitude free oscillations of a bubble in a viscous fluid. The solution consists of a linear combination of discrete normal modes and an integral over a continuous spectrum. The continuous spectrum dominates the solution for large times. As a result, the surface deformation ultimately decays algebraically and not as a modulated damped wave, as has sometimes been suggested. © *1998 American Institute of Physics.*
[S1070-6631(98)01712-7]

The decay through viscosity of small deviations in the shape of an otherwise spherical gas bubble (radius R) in an incompressible fluid poses a classical problem with a long and distinguished history.[1–3] The decay is governed by linear partial differential equations, and this linearity allows the evolution of each surface harmonic in the shape of the bubble to be considered in isolation from the rest, i.e., we may focus on bubble shapes of the form

$$r=R+a_n(t)Y_n(\theta,\phi), \tag{1}$$

where (r,θ,ϕ) are spherical coordinates and Y_n is a spherical harmonic of degree n. In what follows we shall exclude the case $n=1$, which concerns the bodily translation of the bubble. Most interest attaches to the oblate/prolate disturbance, $n=2$.

The distortion of the bubble surface creates a potential flow in the surrounding fluid and, when the fluid is viscous, it also generates vorticity, $\boldsymbol{\omega}$. We follow Prosperetti[2] by writing

$$\boldsymbol{\omega}=\nabla\times[T(r,t)Y_n(\theta,\phi)\mathbf{l}_r] + \nabla\times\nabla\times[S(r,t)Y_n(\theta,\phi)\mathbf{l}_r], \tag{2}$$

where $\mathbf{l}_r$ is the unit radial vector. The poloidal scalar, S, is of no great interest since it is uncoupled to a_n; we therefore set $S=0$. The toroidal scalar, T, is governed by a diffusion equation that describes the radial diffusion of vorticity from the bubble surface:

$$\frac{\partial T}{\partial t}=\nu\left[\frac{\partial^2 T}{\partial r^2}-n(n+1)\frac{T}{r^2}\right], \tag{3}$$

where $\nu(\neq 0)$ is the kinematic viscosity of the fluid. Prosperetti[2] shows that a_n and T are governed by the coupled equations

$$\ddot{a}_n-\frac{2\nu}{R^2}(n-1)(n+1)(n+2)\dot{a}_n+\omega_n^2 a_n + \frac{\nu}{R^2}n(n+1)(n+2)T(R,t)=0, \tag{4}$$

$$T(R,t)+\frac{2}{R}\int_R^\infty\left(\frac{R}{r}\right)^n T(r,t)\,dr=\frac{2(n+2)}{n+1}\dot{a}_n, \tag{5}$$

in which the overdot denotes time differentiation and

$$\omega_n^2=(n-1)(n+1)(n+2)\frac{\sigma}{\rho R^3} \tag{6}$$

defines the natural frequency, ω_n, of bubble oscillation in the absence of viscosity due to interfacial tension, σ, between liquid (with density ρ) and gas. Equations (4) and (5) express, respectively, continuity of normal and tangential stress across the bubble surface.

It is clear that (3)–(5) possess discrete normal mode solutions of the form

$$a_n=a_{n0}\exp(i\omega t), \tag{7}$$

and it is intuitively clear that, since there is no source of energy to maintain the bubble pulsation against viscous losses, $\mathrm{Im}(\omega)>0$. It is far less clear, and in fact untrue, that the discrete modes (7) provide a complete spectrum of modes, complete in the sense that an arbitrary initial state can be expressed as a linear combination of the discrete

1070-6631/98/10(12)/3227/3/$15.00

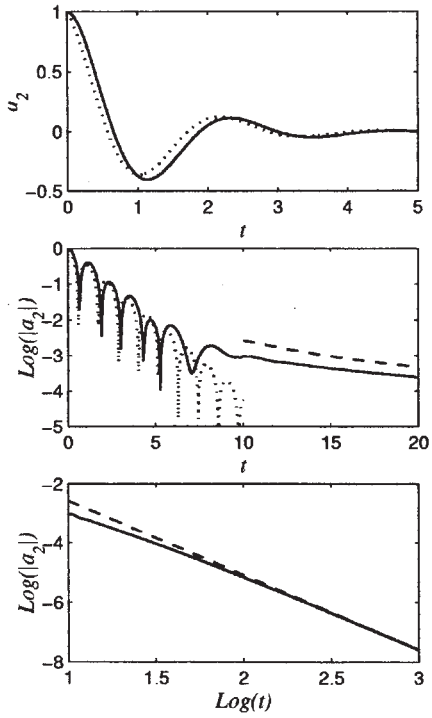


FIG. 1. The evolution of $a_n(t)$ initiated from $a_n(0)=1$, $\dot{a}_n(0)=0$, $T(r,0)=0$, for the case $n=2$, $R=4.5$ μm, $\nu=1.5\times10^{-6}$ m^2/s, $\sigma=0.073$ kg/s^2 and $\rho=1000$ kg/m^3. The top panel shows $a_2(t)$ for $0\leqslant t\leqslant5$ μs, the middle panel shows $\mathrm{Log}(|a_2|)$ for $0\leqslant t\leqslant20$ μs, and the bottom figure plots $\mathrm{Log}(|a_2|)$ vs $\mathrm{Log}(t)$ for $10\leqslant t\leqslant1000$ μs. The initial behavior of a_2 is well represented by a single mode of the form (7): The dotted curves in the top two panels show the single mode solution of $a_2=\exp(-0.942t)\times\cos(2.75t)$, with t in units of μs, for $0\leqslant t\leqslant10$ μs. At large time, a_2 decays algebraically: The asymptote (15) is plotted as dashed lines in the lower two figures for $10\leqslant t\leqslant1000$ μs.

modes; the solution of the initial value problem in general necessarily involves an integral over a continuous spectrum of modes. This is clearly seen in Figure 1, which shows the evolution of $a_n(t)$ when initiated from

$$a_n(0)=1, \quad \dot{a}_n(0)=0, \quad T(r,0)=0, \tag{8}$$

for the case $n=2$, $R=4.5$ μm, $\nu=1.5\times10^{-6}$ m^2/s, $\sigma=0.073$ kg/s^2, and $\rho=1000$ kg/m^3. The initial behavior of a_n is well represented by a single mode of the form (7) and the solution decays as a damped wave. The later evolution is dramatically different: a_n decays algebraically.

As Prosperetti[2] recognized, the initial value problem is most conveniently attacked by Laplace transform methods in which we write

$$\bar{a}_n(s)=\int_0^\infty e^{-st}a_n(t)dt, \quad \bar{T}(r,s)=\int_0^\infty e^{-st}T(r,t)dt. \tag{9}$$

Then, as demonstrated by Prosperetti,[2]

$$s^2\bar{a}_n-\dot{a}_n(0)-sa_n(0)+\omega_n^2\bar{a}_n+\frac{2\nu}{R^2}(n+2)\times\left[2n+1-2n(n+2)\frac{\delta}{R}\right][s\bar{a}_n-a_n(0)]=0, \tag{10}$$

where

$$\frac{\delta}{R}=\frac{K_{n-1/2}(q)}{qK_{n+1/2}(q)+2K_{n-1/2}(q)}, \tag{11}$$

$$q=R\left(\frac{s}{\nu}\right)^{1/2}, \quad \mathrm{Re}(q)>0. \tag{12}$$

The toroidal scalar is given by

$$\bar{T}=\bar{T}_0(s)\left(\frac{r}{R}\right)^{1/2}\frac{K_{n+1/2}(qr/R)}{K_{n+1/2}(q)}, \tag{13}$$

where

$$\bar{T}_0(s)=\frac{2(n+2)}{n+1}[s\bar{a}_n-a_n(0)]\times\frac{qK_{n+1/2}(q)}{qK_{n+1/2}(q)+2K_{n-1/2}(q)}. \tag{14}$$

In (11)–(14), $K_p(z)$ is the associated Bessel function of the second kind of order p.

The presence of the $s^{1/2}$ term in (12) reveals the existence of a branch point at $s=0$ in the complex s-plane. When carrying out the inversion of (9) by a Bromwich integral it is necessary to cut the s-plane along the negative s-axis. The distorted Bromwich contours give rise to contributions from integrals along either side of the cut, and these are what give rise to the continuous spectrum. When t is sufficiently large, the contributions from the vicinity of $s=0$ dominate those from the poles associated with the discrete modes (7). The technique of obtaining the large t expansions is very well known.[4] In brief, it consists in expanding $a_n(s)$ and $T(r,s)$ for small s and retaining the parts that contribute to the Bromwich integrals. We omit the details, which are straightforward, and retain only the dominant terms for $t\gg R^2/\nu$.

We find that

$$a_n\sim-\frac{n(n+2)^2(2n-1)}{2^{(2n-3)}(2n+1)^2\Gamma(n-\frac{1}{2})}\left(\frac{\nu}{\omega_nR^2}\right)^2\left(\frac{R^2}{\nu t}\right)^{n+1/2}. \tag{15}$$

The agreement between (15) and the results of numerical integration is shown in the bottom panel of Figure 1. If $R\leqslant r\ll(\nu t)^{1/2}$, we obtain

$$T\sim\frac{(n+2)(2n-1)^2\nu}{2^{(2n-1)}(n+1)(2n+1)\Gamma(n+\frac{3}{2})R^2}\left(\frac{R}{r}\right)^n\left(\frac{R^2}{\nu t}\right)^{n+1/2}, \tag{16}$$

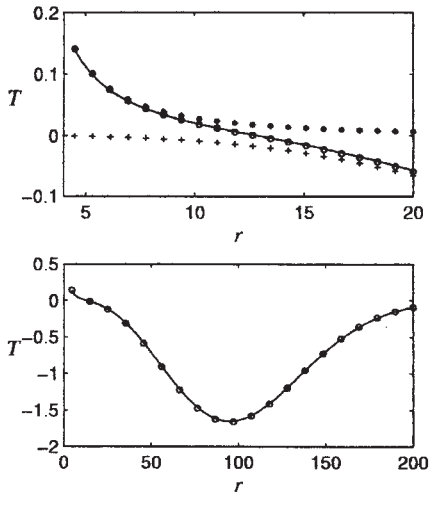


FIG. 2. The toroidal scalar function, $T(r)$, at $t=1$ ms for the calculations specified in Figure 1. The top panel shows T for $4.5 \leqslant r \leqslant 20$ μm. The solid curve represents the numerical solution. Points marked with * and + are, respectively, for the expression (16), for the small r solution, and (17), for the large r solution, and points marked with o represent their composite values. The bottom panel shows T for $4.5 \leqslant r \leqslant 200$ μm, with points marked o showing the composite solution.

which is compared with numerical results in the top panel of Figure 2. For $r \gg (\nu t)^{1/2}$ we have

$$T \sim -\frac{(n+2)(2n-1)}{2^{(2n+1)}(n+1)\Gamma(n+\frac{3}{2})(\nu t)^{1/2}} \left(\frac{R}{\nu}\right)^n \left(\frac{r}{t}\right)^{n+1} e^{-r^2/4\nu t}, \tag{17}$$

which is compared with numerical results in the bottom panel of Figure 2. Figure 2 also shows numerical results for $r = O(\nu t)^{1/2}$, and the results obtained from a composite expansion for T derived by adding the right-hand sides of (16) and (17).

The composite solution is in excellent agreement with the numerical results. For instance, at $t=1$ ms, the numerical solution gives $a_2 = -2.56 \times 10^{-8}$, $T(R) = 0.139$, and $\min(T) = -1.65$ and the composite solution gives $a_2 = -2.62 \times 10^{-8}$, $T(R) = 0.141$, and $\min(T) = -1.66$. Indeed, one can show that the composite solution satisfies (3), (4) and (5) at large times. We have also computed cases for $n=3$ and $n=4$, which compare equally well with the asymptotic solutions.

When $t \ll R^2/\nu$, vorticity has not spread far from the bubble surface; it is significant only in a boundary layer of thickness $O(\delta)$, i.e., of order $(\nu t)^{1/2}$. In fact the boundary layer approximation $\delta \ll R$ in (10) is useful for small t. But it seems clear that the large time algebraic behavior described here will not emerge from such an approximation.

One purpose of this paper is to point out that a suggestion made by Prosperetti[3] is not borne out by the present analysis. Prosperetti suggested that, for the oscillation of both droplets and bubbles,

$$a_n(t) \propto \exp\{[-b(t) \pm i\omega(t)]t\}; \tag{18}$$

see Eq. (23) of his paper. Prosperetti went on to say that "The asymptotic values of ω and b for $t \to \infty$ are those given by the normal-mode analysis." The present analysis shows that this is not correct for bubbles.

A second purpose of this paper is to report that the reduced decay rate of a_2 shown in the center panel of Figure 1 (and associated with algebraic decay) also arises for acoustically-driven bubbles.[5] This may be significant in evaluating the suggestion by Brenner *et al.*[6] that shape instabilities explain the quenching of light for sonoluminescencing bubbles as the acoustic drive is increased beyond a certain threshold.

ACKNOWLEDGMENTS

We are grateful to Professor Seth Putterman for the interest he has taken in this problem. The computations were performed at the San Diego Supercomputing Center, which is supported by the National Science Foundation.

[1] Lord Kelvin, *Mathematical Papers* (Clay & Sons, London, England, 1890), Vol. 3, pp. 384–386; Lord Rayleigh, *The Theory of Sound*, 2nd ed. (Macmillan, London, 1894; reprinted by Dover, New York, 1945); H. Lamb, *Hydrodynamics*, 6th ed. (Cambridge University Press, Cambridge, England, 1940), p. 640; S. Chandrasekhar, "The oscillations of a viscous liquid globe," Proc. London Math. Soc. **9**, 141 (1959); W. H. Reid, "The oscillations of a viscous liquid drop," Quart. Appl. Math. **18**, 86 (1960); S. Chandrasekhar, *Hydrodynamic and Hydromagnetic Stability* (Clarendon, Oxford, 1961), p. 466.

[2] A. Prosperetti, "Viscous effects on perturbed spherical flows," Q. Appl. Math. **34**, 339 (1977).

[3] A. Prosperetti, "Free oscillations of drops and bubbles: the initial-value problem," J. Fluid Mech. **100**, 333 (1980).

[4] See, for example, Sec. 126 of H. S. Carslaw and J. C. Jaeger, *Operational Methods in Applied Mathematics*, 2nd ed. (Oxford University Press, Oxford, 1948; reprinted by Dover, New York, 1963).

[5] C. C. Wu and P. H. Roberts, "Bubble shape instability and sonoluminescence," submitted to Phys. Lett.

[6] M. P. Brenner, D. Lohse, and T. F. Dupont, "Bubble shape oscillations and the onset of sonoluminescence," Phys. Rev. Lett. **75**, 954 (1995).

Part 3

SUPERFLUIDITY

ROTON-PHONON RELAXATION TIMES IN HELIUM II ‡

R. J. DONNELLY and P. H. ROBERTS *
Department of Physics, University of Oregon, Eugene, Oregon 97403, USA

Received 5 November 1969

The conversion of rotons to phonons in helium II is considered as a diffusion process, rather than a phonon-assisted collision process.

A sound wave in liquid helium II will cause a shift in the roton and phonon populations which will relax in times which may be comparable with the frequency of the sound wave. Absorption will then occur. A number of processes contribute to relaxation effects and these are summarized by Wilks [1]. Our discussion has to do only with the roton-phonon conversion process, and we shall have reference to the work of Dransfeld et al. [2].

In his pioneering work on the absorption of sound, Khalatnikov [3] considered the roton-phonon process

$$R_1 + P_1 \rightleftharpoons R_2 + R_3 \qquad (1)$$

which will occur with appreciable probability only if the assisting phonon has an energy of the order of Δ (the minimum roton energy $\Delta/k \approx \approx 8{\cdot}8^{\circ}\mathrm{K}$). He finds the corresponding relaxation time θ_{rp} is given by

$$\frac{1}{\theta_{rp}} = 4 \times 10^{49} \exp(-2\Delta/kT)\left(\frac{\partial \mu_r}{\partial N_r} + \frac{\partial \mu_p}{\partial N_p}\right)_{\rho,s} \mathrm{sec}^{-1}, \qquad (2)$$

where the μ's are chemical potentials, the N's number densities and the suffixes r and p refer to rotons and phonons, respectively.

Instead of the collision process (1), we consider that rotons can diffuse over the barrier which separates them from phonons in the Landau dispersion spectrum. This barrier has a maximum energy $E_c/k \approx 13{\cdot}9^{\circ}\mathrm{K}$. In doing so we take the view, first discussed by Feynman [4], that the rotons belong to the same family of excitations as rings. Note that the energy of a vortex ring with $p/\hbar = 4\text{Å}^{-1}$ is approximately $E/k = 23^{\circ}\mathrm{K}$. This assumption allows us to estimate the appropriate diffusion constants from the scattering of vortex rings by rotons and phonons and leads to relaxation times comparable with (2).

Deviations from equilibrium decay aperiodically by the roton-phonon process at an e-folding rate.

$$\lambda = P_{rp} + P_{pr} = (1 + N_r/N_p)\, P_{rp}, \qquad (3)$$

where

$$P_{rp} = (\Lambda_c \omega_A \omega_c/2\pi)(p_c/p_0)^2 \exp(-\Delta E/kT), \qquad (4)$$

is a probability per second for conversion of a roton to a phonon calculated by methods analogous to those outlined previously by us [5], and P_{pr} is the probability per second for conversion to a roton. Here $\omega_A^2 (= 1/\mu)$ is the curvature of the roton minimum, $\omega_c^2 (= 1/0.5\, M_{He})$ is the curvature of the maximum of the dispersion curve in units of the mass of the helium atom, $p_0 (\approx 1{\cdot}91\hbar\text{Å}^{-1})$ is the minimum roton momentum, $p_c (\approx 1{\cdot}11\hbar\text{Å}^{-1})$ is that of the maximum in the dispersion curve, $\Delta E = E_c - \Delta$ is the height of the barrier to be transversed starting from the roton well. We define a diffusion constant Λ_c,

$$\Lambda_c = 4\alpha(\pi/\rho_s k^3)^{\frac{1}{2}} p_c^{\frac{1}{2}},$$

which enters the Langevin equation for vortex rings. The drag on a vortex is discussed by a coefficient α defined by Rayfield and Reif [6]. Results for λ are shown in table 1 compared with the results for $1/\theta_{rp}$ given by Dransfeld et al. [2], (and using their numerical constants).

The present calculations are easily extended to take into account effects of pressure or added helium 3 by modifying the constants α_r, α_p and

‡ Research supported by the Air Force Office of Scientific Research under a grant AF-AFOSR-785-65, and by the National Science Foundation under a grant NSF-GP-6473.

* Visiting Professor. Permanent address: School of Mathematics, University of Newcastle upon Tyne.

Table 1

T°K	0.6	0.8	1.0	1.2	1.4	1.6
$1/\theta_{rp}$	9.29 ± 0.04	1.02 ± 0.07	2.78 ± 0.08	2.64 ± 0.09	1.24 ± 0.10	3.75 ± 0.10
λ	1.02 ± 0.03	3.40 ± 0.05	1.46 ± 0.07	2.29 ± 0.08	1.88 ± 0.09	1.00 ± 0.10

Table 2

%^{3}He	0	0.5	1.0	2.0	3.0	5.0
λ	2.29 ± 0.08	3.63 ± 0.08	4.97 ± 0.08	7.63 ± 0.08	1.03 ± 0.09	1.56 ± 0.09

α_3 which are summed to give α. Here α_3 refers to the ^{3}He vortex collisions. Results for λ for a series of concentrations of ^{3}He at $T = 1.2$°K are shown in table 2.

Added ^{3}He increases λ substantially, the effect being in some ways equivalent to raising the temperature.

Our model assumes that rotons can diffuse over a barrier of about $5kT$ by a series of fortunate collisions with other quasi particles. The relative importance of each type of scattering is automatically included. This approach provides an alternative to Khalatnikov's process [1], which must accomplish the same result by a single collision with a very energetic (and hence statistically rare) phonon. This model has the further advantage that the relevant scattering cross sections may be estimated from experiments on vortex rings [5], rather than from fitting experiments on sound attenuation.

References

1. J. Wilks, Liquid and solid helium (Oxford, Clarendon Press, 1967) Ch. 8.
2. K. Dransfeld, J. A. Newell and J. Wilks, Proc. Roy. A243 (1958) 500.
3. I. M. Khalatnikov, Theory of superfluidity (W. A. Benjamin, Inc., New York, 1965).
4. R. P. Feynman, Progress in low temperature physics, Vol. I, ed. C. J. Gorter (North Holland, Amsterdam, 1965) Ch. II.
5. R. J. Donnelly and P. H. Roberts, Proc. Roy. Soc. A312 (1969) 519.
6. G. W. Rayfield and F. Reif, Phys. Rev. 136 (1964) A1194.

* * * * *

NUCLEATION OF QUANTIZED VORTEX RINGS BY IONS IN HELIUM II *

Russell J. Donnelly and Paul H. Roberts †
Department of Physics, University of Oregon, Eugene, Oregon 97403
(Received 13 October 1969)

We suggest that a quantized vortex ring is formed when a roton, localized near a moving ion, is expanded by stochastic processes to become a small ring, attached to the ion. Calculations based on this notion are compared with experiment.

The purpose of this Letter is to suggest that quantized vortex rings are nucleated from moving ions by stochastic processes. It is easy to show that conservation of energy and momentum lead to estimates for the critical velocity far too high to agree with experiment.[1] We adopt the view, first advanced by Iordanskii, that there exists a Maxwell distribution of vortex rings in He II.[2] We suppose further that there exists a smallest ring (proto-ring) and that higher states are filled by thermal collisions.

Suppose that proto-rings are somehow located near the equator of an ion moving with velocity v_I. The effect of collisions of other quasiparticles with the ion–proto-ring complex will be, on occasion, to make one of these rings grow to finite size. In this case we know that the natural attraction felt by an ion near the core of a vortex (which has been extensively studied by the authors[3]) will cause the ring to attach to the ion. If the ring grows to be of such radius that its forward velocity is equal to that of the ion it will be a "critical fluctuation" in the sense that further growth will be at the expense of the electric field, and will continue until the drag on the ring is equal and opposite to the electric force. A possible sequence of events is pictured in Fig. 1(a); the advantage of this view is that estimates may be made of the shape, energy, momentum, and velocity of the critical fluctuation which allow calculation of the free energy difference ΔF between the proto-ring and the critical fluctuation. While we cannot specify how the critical fluctuation is formed, ergodic arguments tell us that it will occur with a probability P proportional to $\exp(-\Delta F/kT)$.

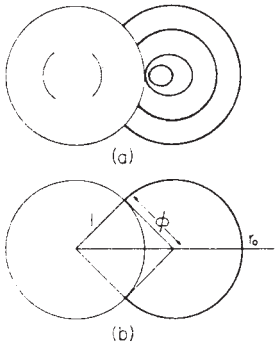


FIG. 1. (a) Possible sequence of growth of a localized roton (inner ring) to a critical fluctuation (outer line segment) attached normally to the ion. Motion of ion is into the paper. (b) Geometry for calculating the properties of the critical fluctuation. $\varphi = R/R_I$.

We compute the proportionality factors by considering vortex rings as Brownian particles. We shall develop in a separate communication Langevin and Fokker-Planck equations for vortex rings, which introduce a diffusion constant (dimensions gm/sec) defined by

$$\Lambda_c = 4\alpha(\pi/\rho_s\kappa^3)^{1/2}p^{1/2}, \tag{1}$$

where ρ_s is the superfluid density, κ the quantum of circulation, p the impulse of the ring, and α (dimensions erg/cm) the drag coefficient for rings as defined by Rayfield and Reif.[4] This drag coefficient is composed of contributions from phonons, rotons, and solvated He^3 atoms, allowing us to compute the influence of each factor in vortex nucleation:

$$\alpha = \alpha_p + \alpha_r + \alpha_3. \tag{2}$$

These quantities are computed, measured and plotted in Rayfield and Reif's paper: α is highly temperature dependent, being dominated by rotons above 0.6°K, by phonons between 0.6 and 0.3°K, and by the natural well concentration of He^3 below 0.3°K.

The evolution of a proto-ring to a critical fluctuation may be thought of as diffusion from a well in momentum space over a saddle point lying ΔF higher in free energy than the bottom of the well. The probability per second of this diffusion taking place is computed by methods discussed by Chandrasekhar[5] and the authors[3] to be

$$P = (2\pi)^{-1}\omega_A\omega_c\Lambda_c e^{-\Delta F/kT}, \tag{3}$$

where ω_A and ω_c are curvatures of the well and

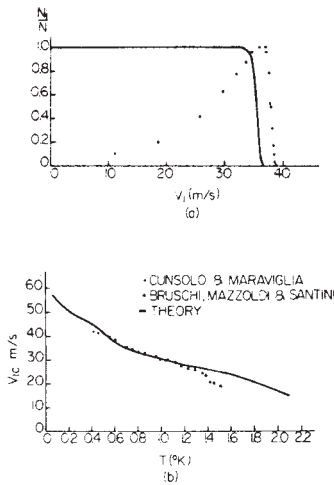


FIG. 2. (a) Theoretical function N_I/N (solid line) for $T = 0.67$°K, positive ion. The experimental points represent ion pulse height, which increases with the increasing electric field required to raise v_I. Lifetime effects cause a sudden drop. The value of $v_{Ic}{}^+$ is 35.8 m/sec theoretically and 38.0 m/sec experimentally. (b) $v_{Ic}{}^+$ calculated from the present theory, compared with two sets of experimental data.

barrier in momentum space (and are each of order 10^{10} gm$^{-1/2}$). Equation (3) contains the probability factor $\exp(-\Delta F/kT)$ referred to earlier.

Suppose now we do an experiment where N ions enter a drift space containing an electric field. The fraction of ions N_I/N which will survive for τ seconds to arrive without nucleating a ring is

$$N_I/N = \exp(-\nu\tau), \tag{4}$$

where $\nu = n_r P$ is the product of the number of proto-rings localized near the ion and the probability per second of one of them becoming a ring. It turns out that this quantity decreases from unity only at quite large velocities, as shown by the solid line in Fig. 2(a). (The experimental points are pulse heights.)

Let us calculate the quantities appearing in the expressions above. The local superfluid velocity with respect to the ion is given by ideal theory as

$$V^2 = v_I{}^2[1 + (R_I{}^3/2r^3)]^2\sin^2\theta + v_I{}^2[1-(R_I{}^3/r^3)]^2\cos^2\theta, \tag{5}$$

where R_I is the radius of the ion and θ is the polar angle measured from the forward direction. The critical fluctuation must be computed in the

presence of the superflow (5) about the ion. A detailed investigation of the shape of this segment of line, sketched in Fig. 1(b), has been made using the Arms-Hama concept of self-induction,[6] concluding that, to good approximation, it may be represented by a circular segment of radius of curvature R, and length of arc $R_I\varphi(\pi + 2\tan^{-1}\varphi)$, where $\varphi = R/R_I$. This gives a vortex energy

$$E = (1/4\pi)\rho_s\kappa^2 R_I[\ln(8R_I\varphi/a) - \tfrac{3}{2}] \times \varphi(\pi + 2\tan^{-1}\varphi), \quad (6)$$

where a is the core radius of the vortex line. The momentum p is taken to be the hydrodynamical impulse corresponding to the area lying between the vortex and the ion, viz.

$$p = \rho_s\kappa R_I^2[\tfrac{1}{2}\varphi^2(\pi + 2\tan^{-1}\varphi) + \varphi - \tan^{-1}\varphi], \quad (7)$$

and the self-induced velocity

$$\frac{\partial E}{\partial p} = \frac{\kappa}{4\pi R_I\varphi}\left[\ln\left(\frac{8R_I\varphi}{a}\right) - \frac{1}{2}\right] + \frac{2\varphi}{(1+\varphi^2)(\pi + 2\tan^{-1}\varphi) + 2\varphi} \quad (8)$$

is made equal to the superfluid velocity at r_0,

$$v_I\left[1 + \frac{1}{2[(1+\varphi^2)^{1/2} + \varphi]^3}\right]. \quad (9)$$

The activation free energy of the critical fluctuation is

$$\Delta F = E - (\partial E/\partial p)p. \quad (10)$$

The determination of the pre-exponential factors in ν require more detail about the proto-ring. Feynman[7] has argued that the roton represents the smallest possible ring. We can demonstrate that rotons may be localized near the ion by examining, as Glaberson, Strayer, and Donnelly did for a vortex core,[8] the effect of the superflow (5) around the ion. This flow lowers the roton energy gap as much as $6kT$ at the equator. The reduction in roton gap, which is already small for $r = 3R_I$, forms an effective well centered about the equator of the ion which inhibits the escape of a roton within it. Furthermore these localized rotons are predominantly polarized against the oncoming superflow and hence are in the position envisaged in Fig. 1(a). A numerical estimate of the number of localized rotons uses the transformation relating energy and $(\vec{v}_n - \vec{v}_s)$. Statements about $\vec{v}_n$, however, are complicated by the fact that mean paths are often large compared with the size of the ion. A conservative estimate of the total number of trapped rotons is given by the expression

$$n_r = (\pi/15)N_r R_I^3(2K^2 + 1 - K\coth K), \quad (11)$$

where $K \equiv p_0 v_I/kT$ is typically of order 5, N_r is the equilibrium roton density, and p_0 is the roton momentum. Note that while n_r may be less than 1, $n_r/\frac{4}{3}\pi R_I^3 N_r > 1$ (e.g., 2.31 at $K = 5$) where the denominator gives a measure of those rotons which are, on the average, near the ion but randomly oriented and hence not candidates for nucleation. The assumption that the smallest rings are rotons gives $\omega_A = \mu^{-1/2}$, where μ is the effective mass of a roton, and ω_c is of the same order. This completes the estimates of all quantities entering the theory.

The results of calculating N_i/N in Eq. (4) at $T = 0.67$°K are shown in Fig. 2(a). The experimental points, obtained by D. M. Strayer of this laboratory, show the pulse height measured over a 5-cm drift space as a function of the velocity v_I of the ion. The rising pulse height at low velocities reflects the current characteristics of the ion source. Our theory suggests that this rising function will be cut off by the factor N_i/N to produce a fall owing to lifetime effects. This "lifetime edge" is observed and has the predicted magnitude and slope. The calculated critical-ion velocity v_{Ic}, defined when $\nu\tau = 1$ or $N_I/N = e^{-1}$, is $v_{Ic}^{+} = 35.8$ m/sec assuming $R_I^{+} = 7.9$ Å,[9] and found to be 38.0 m/sec experimentally. For the negative ion at the same temperature we predict $v_{Ic}^{-} = 32.6$ m/sec assuming $R_I^{-} = 16$ Å,[9] and observe $v_{Ic}^{-} = 32.0$ m/sec.

One can see also that when $N_I/N \cong 1$ only a bare-ion pulse should be observed. When $N_I/N \cong 0$, only ion-ring complexes should be observed. Both should appear in the lifetime-edge region, but the ring pulse, being composed of nucleation events occurring randomly through the drift space, should be spread out in time. All these predictions are realized experimentally.

Calculations have been carried out for the positive ion as a function of temperature. In these calculations ΔF is the most important factor; the core size $a = 1.28$ Å is clearly influential in the logarithmic correction to ΔF and has been taken from Rayfield and Reif.[4] The pre-exponential factors are, fortunately, less important: An order-of-magnitude change influences v_{Ic} by only a few percent. The results are shown in Fig. 2(b) assuming a is independent of temperature. The changes in slope at $T \sim 0.6$°K and $T \sim 0.3$°K correspond to temperatures where rotons,

then phonons, fail to dominate the fluctuations as characterized by the coefficient α discussed above. Experiments by Cunsolo and Maraviglia[10] from 0.4 to 1.0°K are shown, together with experiments above 1°K by Bruschi, Mazzoldi, and Santini.[11] The precise agreement in the magnitude of v_{ic}^{+} with these experiments, which became available after the calculations were made, is fortuitous considering our limited knowledge of the vortex-core and pre-exponential parameters. The divergence in slope above 1.2°K reflects the temperature dependence of a, omitted from the calculation for simplicity. The agreement is, on the whole, gratifying.

The effect of applying pressure on the positive ion is to change Δ/k, μ, and p_0. This is not a large effect, and positive-ion nucleation should be relatively unaffected by pressure. The negative ion, however, is reduced in size on applying pressure (cf. Springett and Donnelly[12]). For example at 20 atm the radius is reduced to about 10 Å. At $T=0.3$°K this is calculated to raise v_{ic}^{-} from 42.1 to 46.9 m/sec. Rayfield,[13] working between 0.3 and 0.6°K, finds a larger rise in v_{ic}^{-} than can be accounted for here.

The effect of adding He^3 is described for the positive ion by a change in α_3 [Eq. (2)] which is proportional to the factor n_3/n_4 of solvated atoms. The decrease in critical velocity reflects the increase in fluctuations due to the He^3.

The effect of adding He^3 on the negative ion may be further enhanced because of localized states on the surface of the ion bubble.[13,14] These trapped He^3 atoms increase the fluctuation rate, and effectively α_3, by an amount proportional to $N_s/R_i n_3$, where N_s is the density of surface states defined by Andreev.[15] Using experimental parameters of Zinovyeva and Bolarev[16] we have included these localized states for the negative-ion calculations in Table I. Note that the effect of doping with He^3 is slightly greater for the negative ion than for the positive ion owing to this effect. Under pressure, the localization of He^3 will tend to counteract the rise in critical velocity mentioned above. Our calculations have the right trend, but do not quantitatively account for Rayfield's experimental observations.[13] Perhaps the bulk parameters used in the calculations[16] must be modified for a microscopic bubble.

Note that on this theory only the velocity v_f is important for nucleation; the electric field plays no essential part. At low enough temperatures, where α is small, the formation of fast ions by α particles gives enough velocity to the ions at the source to produce rings directly. A field given by

$$E_{\min} \cong (\alpha/e)[\ln(8R_i/a)-\tfrac{1}{2}], \qquad (12)$$

which at $T=0.1$°K is only 5 mV/cm, can maintain such rings, so that it should be nearly impossible to observe bare ions, as noted by Neeper and Meyer.[17]

Table I. Variation of critical velocity with He^3 concentration, $T=0.3$°K.

n_3/n_4	v_{ic}^{+} (m/sec)	v_{ic}^{-} (m/sec)
10^{-10}	48.7	42.7
10^{-8}	48.6	40.8
10^{-6}	46.7	38.6
10^{-4}	44.2	36.8
10^{-2}	42.1	35.3

In summary, the fluctuation theory gives an excellent account of nucleation experiments on positive and negative ions at the vapor pressure. While it shows the correct qualitative behavior for the negative ion under pressure and with added He^3, quantitative agreement is lacking. Systematic experiments are urgently needed.

The authors are indebted to Mr. D. M. Strayer for use of his data before publication, and to Professor G. W. Rayfield for a number of discussions.

---

*Research supported by the U. S. Air Force Office of Scientific Research under Grant No. AF-AFOSR-785-65, and by the National Science Foundation under Grant No. NSF-GP-6473.

†Visiting professor. Permanent address: School of Mathematics, University of Newcastle upon Tyne, Newcastle upon Tyne, England.

[1]D. Pines in Quantum Fluids, edited by D. F. Brewer (North-Holland Publishing Company, Amsterdam, The Netherlands, 1966), p. 328.

[2]S. V. Iordanskii, Zh. Eksperim. i Teor. Fiz. 48, 708 (1965) [translation: Soviet Phys.–JETP 21, 467 (1965)].

[3]R. J. Donnelly and P. H. Roberts, Proc. Roy. Soc. (London), Ser. A 312, 519 (1969).

[4]G. W. Rayfield and F. Reif, Phys. Rev. 136, A1194 (1964).

[5]S. Chandrasekhar, Rev. Mod. Phys. 15, 1 (1943).

[6]R. J. Arms and F. R. Hama, Phys. Fluids 8, 553 (1965).

[7]R. P. Feynman, in Progress in Low-Temperature Physics, edited by C. J. Gorter (North-Holland Publishing Company, Amsterdam, The Netherlands, 1965), Vol. I, Chap. II, p. 51.

[8]W. I. Glaberson, D. M. Strayer, and R. J. Donnelly, Phys. Rev. Letters 21, 1740 (1968); W. I. Glaberson, J. Low Temp. Phys. 1, 289 (1969).
[9]P. E. Parks and R. J. Donnelly, Phys. Rev. Letters 16, 45 (1966).
[10]S. Cunsolo and B. Maraviglia, Phys. Rev. (to be published).
[11]L. Bruschi, P. Mazzoldi, and M. Santini, Phys. Rev. Letters 21, 1738 (1968).
[12]B. E. Springett and R. J. Donnelly, Phys. Rev. Letters 17, 364 (1966).
[13]G. W. Rayfield, Phys. Rev. Letters 20, 1467 (1968).
[14]A. J. Dahm, Phys. Rev. 180, 259 (1969).
[15]A. F. Andreev, Zh. Eksperim. i Teor. Fiz. 50, 1415 (1966) [translation: Soviet Phys.– JETP 23, 939 (1966)].
[16]K. N. Zinovyeva and S. T. Boldarev, in Proceedings of the Eleventh International Conference on Low Temperature Physics, St. Andrews, Scotland, 1968, edited by J. F. Allen, D. M. Finlayson, and D. M. McCall (St. Andrews University, St. Andrews, Scotland, 1969), Vol. I, p. 367.
[17]D. A. Neeper and L. Meyer, Phys. Rev. 182, 223 (1969).

THEORY OF THE ONSET OF SUPERFLOW*

Paul H. Roberts† and Russell J. Donnelly
Department of Physics, University of Oregon, Eugene, Oregon 97403
(Received 19 January 1970)

Spontaneous production of quantized vortices in He II by thermal fluctuations is considered. Implications of this theory for superflow in finite channels are discussed.

In a recent communication[1] we examined the problem of the nucleation of quantized vortex rings by ions. There we presented arguments to suggest that the smallest vortex rings are rotons and that larger rings form higher-energy states which are filled by collisions. In order to fix our ideas and notation about superflow, consider first spontaneous vortex production in an unbounded fluid, a concept introduced by Iordanskiĭ,[2] and applied to the interpretation of the decay of persistent currents by Langer and Fisher,[3] Fisher,[4] and Langer and Reppy.[5] Rotons in a counterflow will be polarized against the oncoming superfluid and will become rings if they can fluctuate over a saddle point C in momentum space constituting a free-energy barrier ΔF. We shall show elsewhere that the probability per unit time of such a diffusion taking place may be estimated by Brownian-motion theory of vortex rings to be

$$P = \frac{\Lambda_c kT}{4\pi\mu^{1/2}p_0^2}\left[\frac{v_s p_0/kT}{\sinh(v_s p_0/kT)}\right]\left(\frac{\omega_c}{s_c^2}\right)e^{-\Delta F/kT}, \tag{1}$$

where Λ_c is the diffusivity constant for vortex rings (and includes contributions from rotons, phonons, and solvated He^3 atoms as explained in Ref. 1), v_s is the relative velocity of the counterflow, μ and p_0 are the effective mass and momenta of rotons, ω_c and S_c are the principal curvatures of momentum space at the saddle point. The derivation of (1) follows from considerations similar to those leading to Eq. (133) of Donnelly and Roberts.[6] In a counterflow the density of rotons is given by

$$N_r = 2(2\pi kT)^{3/2}\mu^{1/2}p_0h^{-3}v_s^{-1}\sinh(p_0v_s/kT)\exp(-\Delta/kT). \tag{2}$$

The rate of production of vortex rings per unit volume is therefore given by

$$\nu \equiv PN_r = (2\pi)^{1/2}(kT)^{3/2}\Lambda_c h^{-3}\omega_c S_c^{-2}\exp(-\Delta F/kT), \tag{3}$$

where ΔF is now corrected for Δ. Equation (3) coincides with Iordanskiĭ's expression below (3.14)[2] (except for a factor of $\sqrt{2}\pi^2$ whose origin is not clear to us).

Experiments on the decay of persistent currents use small v_s and large kT, hence we may regard the quantity in square brackets in (1) as unity and write

$$p = f \exp(-\Delta F/kT), \tag{4}$$

where to a reasonable approximation

$$f = \frac{\beta \rho_s \kappa^3 kT}{64\pi^2 p_0^2 [2\mu v_s^5]^{1/2}} \left[\ln\left(\frac{c\kappa}{4\pi v_s a}\right) \right]^2, \tag{5}$$

and

$$\frac{\Delta F}{kT} = \frac{\rho_s \kappa^3}{16\pi kT v_s} \left[\ln\left(\frac{c\kappa}{4\pi v_s a}\right) \right]^2 - \frac{\Delta}{kT}. \tag{6}$$

In these equations $\Lambda_c = \beta p_c^{1/2}$ where $p_c = \rho_s \kappa \pi R_c^2$, R_c is the radius of the ring at C, $c = 8/\sqrt{e}$, and a(=1.28 Å) is the radius of the vortex core. Equation (6) contains the velocity

$$v_{th} = \rho_s \kappa^3 / 16\pi kT \tag{7}$$

which might be called the "thermal characteristic velocity." Typical values for v_{th} are 188 m/sec at 1.1°K, 46.5 m/sec at 2.0°K, and 2.2 m/sec at 2.17°K. These high velocities contain the essence of the difficulty of vortex nucleation. For, unless $v_s \approx v_{th}$, appreciable rates of vortex production will not occur. For example at 2.1°K we find [on a somewhat more precise extimate than is given by (6)] $\nu = 2.90 \times 10^{-73}$ for $v_s = 200$ cm/sec, $\nu = 1.03$ for 388 cm/sec, and $\nu = 1.07 \times 10^{20}$ for $v_s = 600$ cm/sec. If we adopt $\nu = 1$ as a rate characteristic of values of v_s observed in the laboratory, we find the critical velocity is almost an order of magnitude too large to agree with the experiments of Kukich, Henkel, and Reppy[7] and Notarys.[8] It may be noted, however, that the values of fN_r given by (5) (e.g., $fN_r = 6.04 \times 10^{35}$ sec^{-1} cm^{-3} at $v_s = 388$ cm/sec) are of the order of those estimated by Langer and Fisher[3] and stated by them to be a characteristic atomic frequency. We see that since Λ and N_r vary rapidly with temperature, fN_r should be extremely temperature dependent and (5) provides a means of estimating this dependence as well as those changes created by pressure variations, He^3 concentrations, etc., as outlined in Ref. 1.

Consider now the effect of boundaries on the spontaneous nucleation process discussed above. For simplicity we suppose the fluid is confined to a tube of radius R_0 and is at rest. We imagine a ring to nucleate from a roton at the center of the tube polarized along its axis, to a circular ring whose plane is perpendicular to the axis of the tube on which it is centered. This model is common to studies of the Feynman critical-velocity mechanism by Fineman and Chase[9] and Gopal.[10] We will follow the analysis of the latter author assuming, as he did, that the rings are classical "solid core" rings when their radius R is large compared with their cross section a. We imagine that for $R \lesssim a$ the spectrum has a roton well. Also, by analogy, for $R \lesssim R_0 - a$ the quantum pressure effects associated with healing at the wall and the image of the ring produce an analogous minimum which we call the "image well." Following Gopal, we assume this occurs at $R_I \cong R_0 - a$, and evaluate the energy E_I and momentum p_I of that state accordingly. We assume for simplicity that the curvature $\partial^2 E/\partial p^2$ of the image well is $1/\mu^{1/2}$. These considerations lead to the curve sketched in Fig. 1. The theory here applies strictly only to those rotons on the axis of the tube but we apply it (perhaps less accurately) to all rotons in the tube. The energy of a ring in an unbounded fluid is a monotonically increasing function of R; the situation in a tube is crucially different. The confinement of the vortex flow to the tube depresses E at large R, and creates a single maximum C between the roton and image wells, and nucleation can occur in either direction over this barrier. By "nucleation outwards" we will mean the nu-

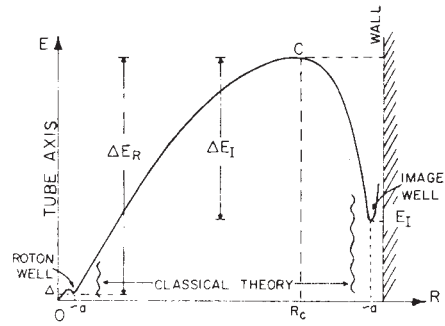


FIG. 1. Schematic diagram of the dependence of the energy of a vortex ring, centered on the axis of a tube, as a function of the radius R of the ring. The curve is also, parametrically, a dispersion curve since $p \propto R^2$. The region in which classical theory is applied is indicated, and the roton and image wells greatly exaggerated for clarity.

cleation of vortex rings in the image well by rotons in the interior. By "nucleation inwards" we mean the reverse process of the collapse of rings from the well into rotons in the interior. In either process, the probability of nucleation is independent of the sense of circulation. Once a nucleation has taken place, however, a flow v_s is created in the tube, and this enhances the probability that a nucleation of the opposite sense will take place next (see below). The mechanism, then, is self-correcting and on average maintains the original state of superfluid at rest. The time scale of this process is, for nucleation outwards, of order $1/(\nu V)$ where V is the volume of helium in the apparatus and ν is the rate parameter introduced above. For nucleation inwards, it is of order $1/(\nu_s S)$ where S is the surface area of the tube, $\nu_s = P_V S_V$ where P_V is the probability for inward nucleation and S_V is the density of vortex line on the walls. We show below that in the presence of a superflow the rate of production of vortices by either of these processes adjusts itself in such a way as to reduce the superflow. We may infer that if τ is "sufficiently small," say 1 second, the phenomenon of superfluidity will be greatly inhibited. This is the well known "suppression of the λ point," or more accurately, the onset temperature T_0 for superflow.

The free energy F is simply the kinetic energy of superflow and can, in the classical range, be obtained from Gopal's analysis.[10] For E_I, we have taken Gopal's estimate

$$E_I = \tfrac{1}{2}\rho_s \kappa^2 R_I(\ln 2 + 7/4). \tag{8}$$

It is clear that since $E_I \gg \Delta$, $\Delta E_I \ll \Delta E_R$ (cf. Fig. 1) and the probability of nucleation inwards greatly exceeds that of nucleation outwards. Balancing the inward rate $(2\pi)^{-1} S_V \Lambda \omega_A \omega_c \exp(-\Delta E_I/kT)$ with the outward rate $(2\pi)^{-1} N_I \pi R_0^2 \Lambda \omega_A \omega_c \times \exp(-\Delta E_R/kT)$ we find the equilibrium density of vortex rings on the wall to be (in units cm^{-1})

$$S_V = (\mu kT/2\pi)^{1/2} \hbar^{-3} (p_0 R_0)^2 \exp(-E_I/kT). \tag{9}$$

Whether such a population as (9) exists may well depend on how the sample is prepared.

For P we find a one-dimensional approximation suffices:

$$P = (2\pi)^{-1} \Lambda_c \omega_A \omega_c \exp(-\Delta E_R/kT), \tag{10}$$

and we use (10) to evaluate T_0 as that temperature at which, on the average, one vortex per unit volume per unit time is fluctuating over the barrier at C in Fig. 1. We have ignored inward nucleations in this definition, but it is easily shown that were they added, the change in the estimate of T_0 would be negligible.

Nothing in the discussion above implies that the superfluid is destroyed above T_0. Rather any attempted superflow in this range would rapidly decay owing to fluctuations. An alternating flow such as is induced by third or fourth sound of frequency f will become strongly damped at a temperature such that $f \approx \nu$. But ν increases so rapidly with T that third and fourth sound should extinguish at nearly the same temperature as direct flow. Evidence for this is seen in Fig. 6 of the article by Guyon and Rudnick.[11]

T_0 is plotted in Fig. 2 as a function of temperature and tube diameter, $d = 2R_0$. With the reservation that the core radius a must be an increasing function of T (we adopt $a = 1.28$ Å throughout), the onset curve is in gratifying agreement with the experiments of Guyon and Rudnick[11] and Fokkens, Taconis, and de Bruyn Ouboter.[14] The latent heat of the vortex cores will contribute a specific heat anomaly with a maximum in the interval $T_0 < T < T_\lambda$.

Now consider the situation if the normal fluid is at rest, but the superfluid is moving everywhere with the uniform velocity v_s (>0) down the tube (plug flow). The difference in free energy between this state and one including a vortex of radius R coaxial with the tube is related to $E(R)$, the expression for $v_s = 0$, by

$$F(R) = E(R) + \vec{v}_s \cdot \vec{p}(R), \tag{11}$$

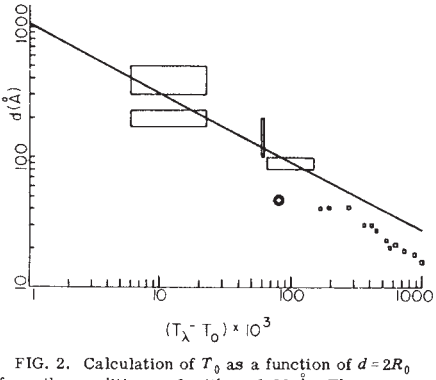


FIG. 2. Calculation of T_0 as a function of $d = 2R_0$ from the condition $\nu = 1$ with $a = 1.28$ Å. The experimental onset temperatures indicated by rectangles and one circle were obtained by Guyon and Rudnick, Ref. 11, and the squares indicate unsaturated film measurements by Fokkens, Taconis, and de Bruyn Ouboter, Ref. 14.

where $\tilde{p}(R) = \rho_s \tilde{\kappa} \pi R^2$. For $\kappa < 0$, the last term of (11) is negative so that $F(R) < E(R)$ and the probability of outward nucleation is increased relative to its value for $v_s = 0$. Since, however, $|p(R)|/E(R)$ increases monotonically with R, the probability of inward nucleation is decreased. Conversely, for $\kappa > 0$ these conclusions are reversed. Outward nucleation becomes less probable. There are, then, four fluctuating processes, inward and outward nucleation of rings of positive and negative circulation. The dominant processes, the outward nucleation of rings of negative κ and the inward nucleation of rings of positive κ, act in a sense to destroy v_s, which will ultimately cease unless an agency acts to maintain it (i.e., a source of thermodynamic potential). Superfluidity is observed in the laboratory when the time scale for this decay greatly exceeds the duration of the experiment.

If we compare the present nucleation process with that for an unbounded fluid we see that the energy barrier at C is always lower for outward nucleation of rings of negative κ than it is at the same v_s in the Iordanskiĭ process. For large flows (compared with the Feynman velocity $\kappa/4\pi R_0$) C moves away from the wall (Fig. 1), image effects are less important, and the results are asymptotically the same as Iordanskiĭ's theory would predict. For small flow rates $R_c(v_s)$ lies close to $R_c(0)$, the maximum of $E(R)$ shown in Fig. 1. This prevents the critical ring from ever exceeding the size of the tube, a problem raised by Notarys.[8]

We may further, in extension of our ideas of the onset of superflow, define $T_0(R_0, v_s)$ as that temperature at which $\nu = 1$. Figure 3 illustrates this generalization. For $v_s = 0$ the curve coincides with that of Fig. 2. The curves for $v_s = 350$ cm/sec and 700 cm/sec agree asymptotically with Iordanskiĭ's theory, and in this region of relatively large channels it is correct to speak of an "intrinsic" critical velocity, i.e., one independent of channel size. This is no longer appropriate for $d \lesssim 800$ Å.

Let us now consider the process of "cooling through the onset temperature T_0." While $T > T_0$ the processes of creation and destruction of vortex rings occur so rapidly that the populations in the fluid and on the walls can adjust continuously to the changing temperature. Below T_0, however, this process will be so slow that further cooling is adiabatic (no matter how slowly it is carried out on the laboratory time scale) and populations on the walls are now "frozen" to their values at T_0 as roughly estimated by (9). Normally, as we have seen, the numbers of vortices of positive and negative circulation should be equal: $S_V^+ = S_V^-$. If on cooling a superfluid gyroscope through T_0 fluctuations should leave one sign of vortex predominant, a test would reveal a spontaneous persistent current of magnitude $\kappa(S_V^+ - S_V^-)$ to have been generated, as indeed Mehl and Zimmerman have reported.[13]

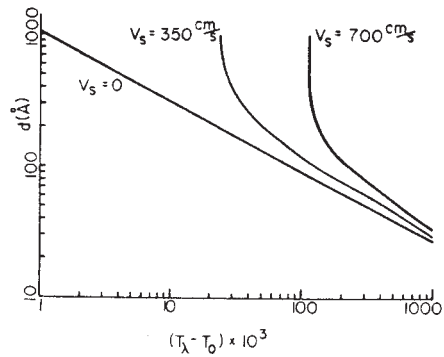


FIG. 3. Generalized onset temperature for $v_s = 0$, 350, and 700 cm/sec as a function of channel size. The whole family of curves $T_0(R_0, v_s)$ may be thought of as a map of critical velocities.

Now suppose, for example, we were to rotate a gyroscope with 200 Å channels at a speed corresponding to $v_s = 700$ cm/sec, cool down to a temperature below $(T_\lambda - T_0) = 200$ mdeg K, and stop. A persistent current of magnitude 700 cm/sec will be observed corresponding to a population difference $(S_V^+ - S_V^-)$ induced by the rotation. If, now, the temperature is raised to $(T_\lambda - T) = 60$ mdeg K, fluctuations will occur, and the persistent current will decay, at first rapidly, then more slowly until the critical value of v_s (350 cm/sec) is reached and passed. This decay has been observed experimentally by Kukich, Henkel, and Reppy under the general conditions described.[7,5] As one would expect, the rate of decay reported by them is greater, the higher the temperature.

While the qualitative effects described here are in striking accord with experiments, the calculated velocities are still higher than those observed experimentally. We believe core-size corrections and density-of-states arguments will improve the agreement but we must await better experimental knowledge of these quantities and their variation with temperature.

*Research supported by the Air Force Office of Scientific Research under grant No. AF-AFOSR 67-1239, and by the National Science Foundation under grant No. NSF-GP-6473.

†Visiting Professor. Permanent address: School of Mathematics, University of Newcastle upon Tyne, Newcastle upon Tyne, England.

[1]R. J. Donnelly and P. H. Roberts, Phys. Rev. Letters 23, 1491, 1969.

[2]S. V. Iordanskiĭ, Zh. Eksperim. i Teor. Fiz. 48, 708 (1965) [Soviet Phys. JETP 21, 467 (1965)].

[3]J. S. Langer and M. E. Fisher, Phys. Rev. Letters 19, 560 (1967).

[4]M. E. Fisher, in Proceedings of the Conference on Fluctuations in Superconductors, Stanford Research Institute, Menlo Park, Calif., March 1968, edited by W. S. Goree and Frank Chilton (unpublished).

[5]J. S. Langer and J. D. Reppy, "Progress in Low-Temperature Physics" (to be published). We are grateful to Professor Langer and Professor Reppy for a preprint of their interesting paper.

[6]R. J. Donnelly and P. H. Roberts, Proc. Roy. Soc. Ser. A 312, 519 (1969).

[7]G. Kukich, R. P. Henkel, and J. D. Reppy, Phys. Rev. Letters 21, 197 (1968).

[8]H. A. Notarys, Phys. Rev. Letters 22, 1240 (1969).

[9]J. C. Fineman and C. E. Chase, Phys. Rev. 129, 1 (1963).

[10]E. S. R. Gopal, Ann. Phys. 25, 196 (1963).

[11]E. Guyon and J. Rudnick, J. Phys. (Paris) 29, 1081 (1969).

[12]It may be argued that for a ring near the wall $p = \rho_s \kappa \pi (R_0^2 - R^2)$. Since we are interested only in free-energy differences, the alternate definition is immaterial.

[13]J. B. Mehl and W. Zimmerman, Jr., Phys. Rev. 167, 214 (1968).

[14]K. Fokkens, K. W. Taconis, and R. de Bruyn Ouboter, Physica 32, 2129 (1966).

Motion of Ions at Finite Velocities in Helium II†

Donald M. Strayer and Russell J. Donnelly
Department of Physics, University of Oregon, Eugene, Oregon 97403

and

Paul H. Roberts*
National Center for Atmospheric Research,‡ Boulder, Colorado 80301
(Received 4 December 1970)

The drag force on an ion moving through liquid helium II at finite velocities is discussed in terms of localized roton states near the ion. These ideas are compared with experiments on positive and negative ions below 1°K.

The relationship between the drift velocity v_i and the electric field E for positive and negative ions in helium II, up to the nucleation velocity, resembles the curves shown in Figs. 1 and 2. The ion mobility μ, which should strictly be defined from $1/\mu = (\partial E/\partial v_i)_{v_i=0}$, was first discussed and related to experiment by Reif and Meyer on the basis of kinetic-theory arguments.[1] More sophisticated theory[2] and experiments[3] have produced a quite satisfactory account of ion mobilities. The motion at finite velocities, however, has remained less well understood and our purpose is to advance a hypothesis to account for the

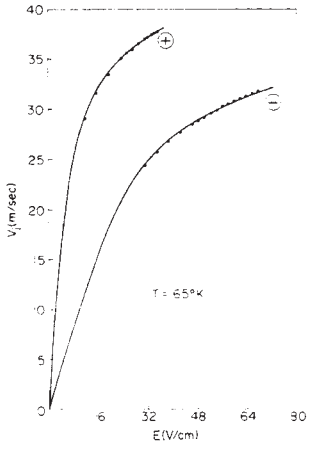


FIG. 1. The average drift velocity of positive and negative ions at $T = 0.65$°K. The points are experimental data. The curves result from Eq. (8) using the fitted constants $f_+ = 0.065$, $\mu = 400$ cm^2/V sec, $f_- = 0.010$, and $\mu = 65$ cm^2/V sec.

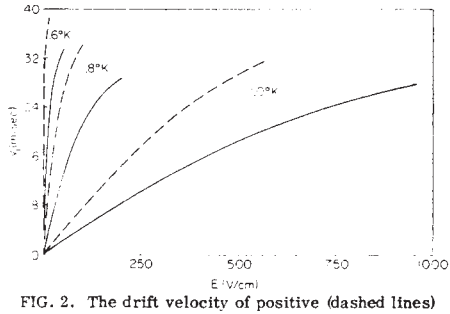


FIG. 2. The drift velocity of positive (dashed lines) and negative (solid lines) ions at $T = 0.6$, 0.8, and 1.0°K. The curves result from Eq. (8) using mobility data of Ref. 1 and the constants $f_+ = 0.060$, $f_- = 0.012$.

shape of the drag curve.

A continuum theory of the drag on an ion would begin by solving the Landau equations (see, for example, Donnelly,[1] Chap. 4). At low speeds the velocity field for the normal fluid with respect to the ion would be given, at a radial distance r from the ion small compared with ν/v_i (where ν is the kinematic viscosity of the normal fluid), by the Stokes flow:

$$v_{nr} = -v_i \cos\theta\left[1 - \frac{3R_i}{2r} + \frac{R_i^3}{2r^3}\right], \tag{1a}$$

$$v_{n\theta} = v_i \sin\theta\left[1 - \frac{3R_i}{4r} - \frac{R_i^3}{4r^3}\right], \tag{1b}$$

and for the superfluid by the potential flow:

$$v_{sr} = -v_i \cos\theta\left[1 - \frac{R_i^3}{r^3}\right], \tag{2a}$$

$$v_{s\theta} = v_i \sin\theta\left[1 + \frac{R_i^3}{2r^3}\right], \tag{2b}$$

where $\theta = 0$ is the direction of motion and R_i is the radius of the sphere. At finite velocities the variables, particularly ρ_n, are functions of $\vec{w} = (\vec{v}_n - \vec{v}_s)$ (e.g., Donnelly,[4] pp. 112-113) so that from (1) and (2) one could calculate $\vec{w}$, then $\rho_n(w)$, then solve anew for $\vec{v}_n(w)$ and $\vec{v}_s(w)$. One would hope that this iterative procedure would converge. We approach the problem in this Letter by means of a single iteration, and the use of the localized roton scheme first advanced by Glaberson, Strayer, and Donnelly[5] for calculating the properties of the vortex core. We cannot expect a quantitative solution to the problem from such a drastic procedure, nor from the adoption of a continuum picture when mean free paths of excitations are much greater than the size of the ion. We can hope, nevertheless, to identify the governing physical processes.

The roton density in the presence of relative velocity of normal and superfluid is (in standard notation)

$$N(T, w) = h^{-3}\int\{\exp[(\Delta + (p-p_0)^2/2\mu_0 - pw\cos\varphi)/kT] - 1\}^{-1} d^3p, \tag{3}$$

where φ is the angle between $\vec{w}$ and $\vec{p}$. The integral in Eq. (3) can be evaluated by noting that the effect of w is to depress the roton mininum to $\Delta_m = \Delta - \frac{1}{2}\mu_0 w^2 - p_0 w$ at $\varphi = 0$ and $p_m = p_0 + \mu_0 w$. The principal contribution to the integral arises from the neighborhood of the minimum (Δ_m, p_m). Since in the present application we find that Δ_m/kT may still be considered significantly greater than unity, we drop the 1 in Eq. (3) and, observing that $p_0 \gg \mu_0 w$, we find

$$\frac{N(T, w)}{N(T, 0)} = \frac{\sinh(p_0 w/kT)}{(p_0 w/kT)}, \tag{4}$$

where $N(T, 0) = N_r$, the ordinary roton density. The number of trapped rotons n_r should come by integrating $N(T, w) - N_r$, using Eqs. (1) and (2), over all space outside the ion. One finds that this quantity diverges, and that the origin of the divergence is the Stokes approximation which leads to Eq. (1). (See, for example, Landau and Lifshitz.[6]) This divergence may be removed by a proper discussion of the Oseen flow which occurs at distances from the ion large compared with ν/v_i, and by considering the incipient wake which occurs at these distances. It would appear, however, that although the integrated enhancement in roton density from these regions is large, it does not correspond to rotons trapped in the sense described below. For this reason, a finite value of n_r was obtained by cutting off the integration. The choice of cutoff may be appreciated by reference to Fig. 3 which shows contours of constant $|\vec{w}|$ plotted in the unit v_i. These contours can be seen from Eq. (3) to be contours of constant well depth for the trapped rotons, and hence turning points for the motion of rotons whose energy is $p_0 w$ below Δ. Those surfaces which intersect the ion surround the deepest energy well (at the equator), and we have chosen to integrate only to the last such surface, characterized by $|\vec{w}| = v_i/\sqrt{3}$. This produces a finite estimate for n_r. It also introduces a new length into the problem, the distance l over which trapping occurs, and Fig. 3 shows that this is broadly of order of the radius of the ion.

The calculation above assumes that the roton is a point particle. Actually the roton wavelength is nearly 4 A, compared with the presently accepted ion radii of $R_i^- = 16$ A, $R_i^+ = 5$ Å.[3] We have attempted, as in Ref. 5, to make a first-order correction for this by an uncertainty-principle argument. If V is the volume available to a trapped roton as estimated from Fig. 3 and the

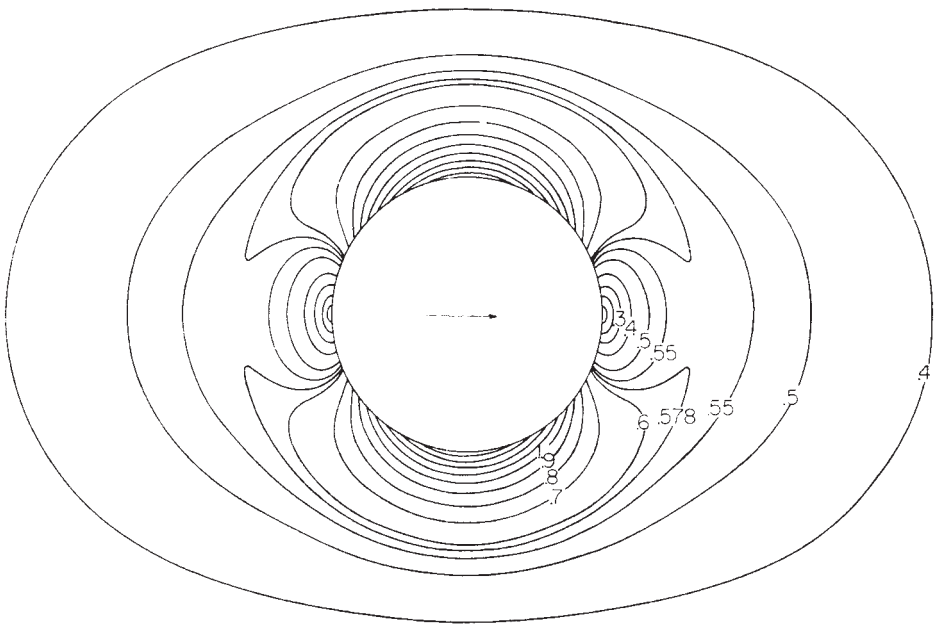


FIG. 3. Lines of $|\vec{w}| =$ constant near a moving ion. The curves are labeled with the corresponding constant value of $|\vec{w}|$ measured in units of v_i. The direction of motion is indicated by the arrow.

argument above, the spread in momenta is of order $\Delta p \sim \hbar/V^{1/3}$ and so $\Delta E \approx (\Delta p)^2/2\mu_0$. This energy is added to the argument of the exponential in Eq. (3). Final values of n_r come from numerical integration by Simpson's one-third rule.

The results require extensive tabulation as they are temperature dependent and are not represented over the entire range of the variable $K = p_0 v_i/kT$ by any simple function. We find, however, that in the temperature range of interest ($0.6^\circ \leq T \leq 1.0$ K) and the velocity range $3 < K < 8$ the following formula approximates the results:

$$n_r = N_r R_i^3 \mathcal{E}_r, \tag{5}$$

where

$$\mathcal{E}_r = c\left[\frac{\sinh mK}{mK} - 1\right], \tag{6}$$

and c and m are chosen to best fit the exact calculations: $m_+ = 0.9$, $m_- = 1.0$; at $T = 0.6$, 0.8, and 1.0°K, we find $c_- = 4.94$, 6.08, and 6.94; $c_- = 7.97$, 8.33, and 8.56, respectively. Note that (6) is not a good approximation outside the velocity range indicated. We consider these results to be more reliable than Eq. (11) of a previous publication,[7] which was computed on the assumption that $\vec{v}_n$ was zero near the ion.

Viewing the ion in the laboratory system is instructive. We note first that $n_r \ll 1$, i.e., the ion is often bare: A roton which approaches the ion is localized for a time which may be estimated from Eq. (3) of Ref. 7. For example, at $T = 0.6^\circ$K and $v_i = 30$ m/sec, the mean lifetime of a roton localized near a negative ion is $\sim 3 \times 10^{-5}$ sec, whereas the interval between collisions with bath rotons is of order $2 \cdot 10^{-7}$ sec. The trapped roton has a thermal velocity of order $v_r = (kT/\mu_0)^{1/2}$ or 91 m/sec and hence fluctuates back and forth at a frequency $\approx (kT/\mu_0)^{1/2}/R_i = 6 \times 10^{10}$ sec^{-1}. During the time the trapped roton stays near the ion, the ion will experience about 150 collisions with rotons of the bath and the trapped roton will fluctuate in its well perhaps 1000

times between each of these collisions. We have in addition, of course, phonon collisions.

We are now in a position to estimate the drag on the ion-roton complex. Let us take a view consistent with our continuum calculation of $\tilde{w}$. We suppose that each time the localized roton fluctuates away from the ion, it comes into contact with the surrounding normal fluid which is essentially at rest. Some momentum is lost each time this occurs, the exact amount depending upon what constitutes the "normal fluid" outside, i.e., rotons, phonons, and He^3 atoms. The effective mass associated with a roton in the well is $p_0^2/3kT$ [cf. Donnelly,[4] Eq. (3.73)]. The amount of momentum transmitted by our fluctuating roton is written as $f(p_0^2/3kT)v_i$, where f (the only free parameter in this theory) is empirical, and may be a function of temperature, pressure, He^3 concentration, and ion species. The resulting drag on the ion, F_r, is the product of the number of trapped rotons, the fluctuation rate, and the momentum exchange per fluctuation:

$$F_r = f(p_0^2/3kT)v_i(kT/\mu_0)^{1/2}R_i^{-1}n_r$$
$$= f\rho_{nr}\mathcal{E}_r v_i(kT/\mu_0)^{1/2}R_i^2. \quad (7)$$

Note that $p_0^2N_r/3kT = \rho_{nr}$, the roton contribution to the normal fluid density evaluated far from the ion (cf. Donnelly,[4] p. 86). Observe also that the fluctuating roton plays the role of momentum transfer agent between the moving ion and the stationary fluid outside; in this sense we call the process "quasiviscous."

Equation (7) may be obtained a different way. Let us calculate the viscous stress $\mathfrak{F} = \eta(\partial v_n/\partial r)$ on the surface of the ion. The shear is $\partial v_n/\partial r \approx v_i R_i$ and the coefficient of viscosity is $\eta = nm\langle v\rangle l$ [cf. Donnelly,[4] Eq. (3.97)], where n is the number density $n \approx n_r/R_i^3$, $m \approx p_0^2/3kT$, $\langle v\rangle \approx (kT/\mu_0)^{1/2}$, and $l \approx R_i$. The drag on the ion is then the product of the stress and the area of the surface, which recovers Eq. (7). Thus the quasiviscous drag operates over the trapping length about the ion, and this length acts in place of the classical boundary-layer thickness. Because l acts in place of the free path, the coefficient of viscosity here contains a density, which is absent for an ideal gas.

Consider now the experimental situation. In the steady state the drag force on the ion is balanced by the electric field: $F = eE$. Thus the field E required is the dependent variable, being the force per unit charge necessary to maintain the ion drift velocity v_i. At low velocities the required field is given by the mobility relation $E = v_i/\mu$, where μ is made up of contributions from phonons, rotons, and He^3 atoms: $\mu^{-1} = \mu_r^{-1} + \mu_p^{-1} + \mu_3^{-1}$. We now add the quasiviscous drag (7):

$$E = \frac{v_i}{\mu} + v_i\left(\frac{f}{e}\right)\rho_{nr}\left(\frac{kT}{\mu_0}\right)^{1/2}R_i^2\mathcal{E}_r. \quad (8)$$

Equations (8) and (6) show that at low velocities the ordinary mobility law holds,[3] while at higher velocities the field required increases exponentially—indeed for $K \gg 10$ it increases as $\exp(3K/2)$. Comparison with data taken recently at the University of Oregon with a fast electrometer is shown in Fig. 1, where the constants $f_+ = 0.065$, $\mu_+ = 400$, $f_- = 0.010$, and $\mu_- = 65$ were adjusted to fit the data. The shape of the high-field drag is given fairly accurately. A family of curves for $T = 0.6$, 0.8, and 1.0°K for positive and negative ions is shown in Fig. 2. The curves are calculated from Eq. (8) using the constants m and c quoted above and $f_+ = 0.060$, $f_- = 0.012$. The mobilities chosen have been taken from Ref. 1: $\mu_+ = 1700$, 66, and 7.4 and $\mu_- = 155$, 26, and 4.2 at $T = 0.6$, 0.8, and 1.0°K, respectively. Considering experimental uncertainties in the absolute magnitude of μ, we have chosen f, the one empirical parameter, to be independent of temperature for Fig. 2. This choice is supported by the results of fitting to our data in the temperature range $0.6° \leq T \leq 0.75$°K: The fitted f's are nearly independent of temperature and have approximately the values chosen for Fig. 2.

Inspection of the curves in Fig. 2 show several features of interest. The fields required to bring the ions up to the nucleation velocity are highly temperature dependent. This reflects the dominance of μ in Eq. (8). The velocities at which the quasiviscous drag becomes apparent are <u>not</u> dramatically temperature dependent (they range from 20 to 30 m/sec in Fig. 2). This reflects the dominance of the dimensionless velocity K in $\mathcal{E}_r$.

The connection between the shape of the curves and nucleation of vortex rings should be noted. At low velocities the ion is bare and obeys a mobility law. As the velocity increases, rotons appear near the equator, predominantly polarized in the direction of motion. There is much agitation of these rotons: They are struck by the various excitations belonging to the normal fluid, and they are absorbed and emitted from the well about the ion at a high rate. Finally, one of them reaches critical size and nucleates

a ring, and we have an ion-ring complex where the ring grows to macroscopic size. The ion-ring complex grows (and slows down) until the total drag force again balances eE. This is what Careri has termed the "giant fall." The binding of the ion to the ring has been described earlier by Donnelly and Roberts[9]; clearly if the field is further increased the ions will begin to escape from the rings and the ion velocity will, on average, be seen to increase again. This has been observed by Bruschi, Mazzoldi, and Santini.[10] The entire shape of the (v_i, E) curve, then, reflects the making of complexes by the moving ion and the distinctive role played by fluctuations.

We are grateful to Professor G. D. Mahan and Professor G. H. Wannier for discussions of this problem.

---

†Research supported by the U. S. Air Force Office of Scientific Research under Grant No. AF-AFOSR-67-1239, and by the National Science Foundation under Grant No. NSF-GP-26361.

*Permanent address: School of Mathematics, University of Newcastle upon Tyne, Newcastle upon Tyne, England.

‡Supported by the National Science Foundation.

[1]F. Reif and L. Meyer, Phys. Rev. 119, 1164 (1960).

[2]G. Baym, R. G. Barrera, and C. J. Pethick, Phys. Rev. Lett. 22, 20 (1969).

[3]K. W. Schwarz and R. W. Stark, Phys. Rev. Lett. 21, 967 (1968), and 22, 1278 (1969).

[4]R. J. Donnelly, *Experimental Superfluidity* (Univ. of Chicago, Chicago, Ill., 1967).

[5]W. I. Glaberson, D. M. Strayer, and R. J. Donnelly, Phys. Rev. Lett. 21, 1740 (1968); also W. I. Glaberson, J. Low Temp. Phys. 1, 289 (1969).

[6]L. D. Landau and E. M. Lifshitz, *Fluid Mechanics* (Addison-Wesley, Reading, Mass., 1959), Sect. 20.

[7]R. J. Donnelly and P. H. Roberts, Phys. Rev. Lett. 23, 1491 (1969).

[8]The most accurate expressions for mobility come from the work of Baym, Barrera, and Pethick (Ref. 2). It is interesting to notice that the roton drag on an ion can be written down by direct analogy with Eq. (7) as $P_r^{-1} = (e/\mu_r) = f(p_0^2/3kT)N_r(kT/\mu_0)^{1/2}\sigma = (3\pi^2)^{-1}\hbar k_0^4\sigma \times \exp(-\Delta/kT)$, the expression obtained recently by Schwarz and Stark [Eq. (5) of the second paper of Ref. 3] if $f = (2/\pi)^{1/2}$.

[9]R. J. Donnelly and P. H. Roberts, Proc. Roy. Soc., Ser. A 312, 519 (1969).

[10]L. Bruschi, P. Mazzoldi, and M. Santini, Phys. Rev. Lett. 21, 1738 (1968).

DYNAMICS OF ROTONS *

P.H. ROBERTS and R.J. DONNELLY **
Niels Bohr Institute, Copenhagen Ø, Denmark

Received 15 December 1972

It is shown that point dipoles in a superfluid obeying the Landau dispersion relation for rotons can form stable bound states. Structure is added to the roton model by adding the octupole moment appropriate to a vortex ring. The binding energy is then computed to be 0.337°K. Scattering appears to be satisfactorily accounted for by the dipole approximation.

The interaction of rotons has become an active field of inquiry, especially since the discovery of an apparently bound state [1]. The theory of roton interactions has been discussed by Zawadowski, Ruvalds and co-workers [2], by Yau and Stephen [3] and others. We report a direct determination of roton orbits.

We proceed from Onsager's and Feynman's idea that rotons resemble microscopic vortex rings of momentum $\boldsymbol{p}$, having a dipole moment of strength $\boldsymbol{p}/4\pi\rho$, where $p = |\boldsymbol{p}|$ and ρ is the density [4]. Vortex rings interact at large distances by dipole-dipole forces; at closer distances octupole and higher moments become important. For rotons, close interactions will also involve quantum corrections. We show first that the dipole interaction plus the Landau dispersion relation for rotons, $\epsilon = \Delta + (p - p_0)^2/2\mu$, is sufficient to guarantee negative energy states, some of which are unstable, but others of which are stable. We then show that the addition of the next higher (octupole) moment is adequate recognition of the roton's internal structure to yield a binding energy in excellent agreement with experiment. Finally, positive energy states are studied, and the roton-roton scattering cross section is calculated.

Consider two rotons in a stationary superfluid with momenta close to p_0, the momentum of the minimum. Measuring energy from Δ, the Hamiltonian for a single roton is $H = (p - p_0)^2/2\mu$. Rotons for which $p > p_0$ move in the direction of $\boldsymbol{p}$, while those for which $p < p_0$ move in the opposite sense. We call the former 'R-rotons' to recall that their motion qualitatively resembles classical vortex rings, and the latter 'Q-rotons' to recall that quantum effects are the origin of this unusual behaviour [4].

Consider the interaction of two rotons with momenta $\boldsymbol{p}_1$ and $\boldsymbol{p}_2$ situated at $\boldsymbol{q}_1$ and $\boldsymbol{q}_2$. The total Hamiltonian is then

$$H = (p_1 - p_0)^2/2\mu + (p_2 - p_0)^2/2\mu + H_{12}(\boldsymbol{p}, \boldsymbol{P}, \boldsymbol{r}) , \tag{1}$$

where

$$\boldsymbol{p} = \boldsymbol{p}_1 - \boldsymbol{p}_2, \boldsymbol{P} = \boldsymbol{p}_1 + \boldsymbol{p}_2, \boldsymbol{r} = \boldsymbol{q}_1 - \boldsymbol{q}_2, \boldsymbol{R} = \tfrac{1}{2}(\boldsymbol{q}_1 + \boldsymbol{q}_1). \tag{2}$$

Since H_{12} is independent of time, the energy H is a constant of the motion; since it is independent of $\boldsymbol{R}$, the total momentum $\boldsymbol{P}$ is a constant of the motion. The interaction H_{12} may be written in operational form as

$$H_{12} = h(\theta_1, \theta_2, p_1, p_2)\,(1/r) , \tag{3}$$

where h is a rational function of $\theta_1 = \boldsymbol{p}_1 \cdot \boldsymbol{\nabla}$, $\theta_2 = \boldsymbol{p}_2 \cdot \boldsymbol{\nabla}$, p_1 and p_2, and $\boldsymbol{\nabla}_i = \partial/\partial r_i$. The simplest case is $h = \theta_1\theta_2/4\pi\rho$, which yields the dipole interaction

$$H_{12} = \left(\frac{1}{4\pi\rho}\right)(\boldsymbol{p}_1 \cdot \boldsymbol{\nabla})(\boldsymbol{p}_2 \cdot \boldsymbol{\nabla})\left(\frac{1}{r}\right)$$

$$= \frac{3(\boldsymbol{p}_1 \cdot \boldsymbol{r})(\boldsymbol{p}_2 \cdot \boldsymbol{r})}{4\pi\rho r^5} - \frac{\boldsymbol{p}_1 \cdot \boldsymbol{p}_2}{4\pi\rho r^3} . \tag{4}$$

When (3) holds, it follows from the canonical equations that the total angular momentum $\boldsymbol{M}$, is preserved

$$\tfrac{1}{2}\boldsymbol{r} \times \boldsymbol{p} + \boldsymbol{R} \times \boldsymbol{P} = \boldsymbol{M} = \text{constant} . \tag{5}$$

* Research supported by the Air Force office of Scientific Research, Grant AF-AFOSR-71-1999 and by the National Science Foundation, Grant NSF-GH-35898.

** Permanent addresses: P.H.R. University of Newcastle, NE1 7RU. R.J.D. University of Oregon, Eugene 97403, USA.

We introduce a dimensionalization by measuring length in units of $\lambda = (\mu/4\pi\rho)^{1/3}$, time in units of $\lambda\mu/p_0$, and energy in units of p_0^2/μ. At the vapor pressure, $\mu = 1.06 \times 10^{-24}$ gm, $p_0 = 2.01 \times 10^{-19}$ gm cm sec^{-1}, $\rho = 0.145$ gm cm^{-3} and $\lambda = 0.835$ Å, $\lambda\mu/p_0 = 4.40 \times 10^{-14}$ sec, $p_0^2/\mu k = 276°$K.

The theory is straightforward in the case $P = 0$. Orbits remain in the plane defined by the initial directions of $\boldsymbol{p}$ and $\boldsymbol{r}$. The dimensionless equivalents of (1), (4) and (5) give

$$H = (p - 2)^2/4 - p^2/2r^3 + 3M^2/r^5 . \qquad (6)$$

On examining the curves of constant H in the (r, p) plane, using (6), we find a variety of unbound orbits, some corresponding to scattering, and some corresponding to motion asymptotic to $r = 2^{1/3}$ or $r = 0$. The latter collisions represent the absorption of one roton by another. This is an unlikely process for real rotons since $r = 2^{1/3} \to 1.05$ Å, and quantum effects would be likely to exclude such a close encounter.

Negative energy states occur and, for all $M > M_1 = 16/3\sqrt{10} \approx 1.6865$, two circular orbits exist. These are found by minimizing H with respect to r and p; we find

$$p(1 - 2/r^3) = 2 , \quad pr = M/10 , \qquad (7)$$

the latter of which, with (5), shows that the angle between $\boldsymbol{p}$ and $\boldsymbol{r}$ is always arc $\cos\sqrt{\frac{3}{8}} = 53.1°$. Eq. (7) has two real roots for $M \geqslant M_1$. One corresponds to a saddle point in the energy surface, and hence to unstable circular orbits; the other gives stable orbits whose energy is a monotonically increasing function of M being $-1/15$ ($-18.41°$K) for $M = M_1$, and zero for $M = \infty$, (measured from 2Δ). Explicit integration of the canonical equations shows circular orbits, and neighbouring bound orbits of the same angular momentum oscillating stably about this potential minimum.

Let us include roton structure by adding an octupole term of strength γp^2 in dimensionless units, or $\gamma p^2(\lambda^2/4\pi\rho p_0)$ in dimensional units. We replace (4) in dimensionless units by

$$H_{12} = (1 + \gamma\theta_1^2/p_1)(1 + \gamma\theta_2^2/p_2)\theta_1\theta_2(1/r) . \qquad (8)$$

Two families of circular orbits are found for $\gamma \neq 0$, both of negative energy, and both involving R-rotons: one is unstable, the other stable. For $\gamma = p_0/8\pi\rho\kappa\lambda^2 \approx \approx 0.794$, the value appropriate to a vortex ring, the sequence of stable bound states now starts from a minimum energy state of radius of 8.125λ (6.7Å) and an energy of $-1.27 \times 10^{-3}\, p_0^2/\mu$ ($-0.337°$K). This is in good agreement with the experimental value $0.37 \pm 0.1°$K [1]. At $P = 25$ atm, the separation decreases to 6Å and the binding increases to $0.466°$K. These separations provide some confidence that our neglect of quantum corrections is justified.

We have examined open (scattering) orbits of rotons on the dipole approximation, for $P = 0$. We find two cross sections. One, a small cross section insensitive to the relative velocity of the rotons, corresponds to roton absorption and may be discarded as a defect of the dipole approximation for short distances. The second divides quite sharply into forward scattering at large impact parameters and backscattering at small impact parameters. For $H \to 0$ it is easy to show that the critical impact parameter dividing these cases is $\rho_0(p) = \lambda p_0^{2/3}/(p - p_0)^{2/3}$ and, after averaging over the Maxwell-Boltzmann distribution $f(\epsilon)$ at p_0, a representative cross section

$$\sigma = \frac{\int_0^\infty p^4 \mathrm{d}p \left[-\frac{\partial f(\epsilon)}{\partial \epsilon} \right] \left| \frac{\partial \epsilon}{\partial p} \right| \pi\rho_0^2(p)}{\int_0^\infty p^4 \mathrm{d}p \left[-\frac{\partial f(\epsilon)}{\partial \epsilon} \right] \left| \frac{\partial \epsilon}{\partial p} \right|} \approx$$

$$\approx \pi\Gamma(\tfrac{1}{3})\lambda^2 (p_0^2/2\mu kT)^{2/3} \qquad (9)$$

is obtained. Whilst a more precise formulation of the transport theory may change the constant in (9), it already is sufficiently accurate to account for the temperature and pressure variation of the roton viscosity and neutron line widths. Note that, on this model, roton-roton scattering is satisfactorily described by the long range dipole-dipole interaction coupled with the Landau dispersion law for rotons. There are no adjustable parameters, or repulsive core assumptions, in the calculations reported here.

[1] T.J. Greytak, R. Woerner, J. Yan and R. Benjamin, Phys.Lett. 25 (1970) 1547.
[2] A. Zawadowski, J. Ruvalds and J. Solana, Phys.Rev. A5 (1972) 399.
[3] J. Yau and M.J. Stephen, Phys.Rev.Lett. 27 (1971) 482.
[4] R.P. Feynman, Progress in low temperature physics, ed. C.J. Gorter (North Holland Publishing Company, Amsterdam, 1957) Vol.1, chapter II, section 11.

A SIMPLE THEORY OF TEMPERATURE-DEPENDENT ENERGY LEVELS: APPLICATION TO He II*

P.H. ROBERTS and R.J. DONNELLY
Department of Physics, University of Oregon, Eugene, Oregon 97403, USA

Received 23 December 1975

The statistical mechanics of interacting Bose quasi particles is considered for the case where energy levels are temperature-dependent and known from experiment. The results are applied to the thermodynamic properties of He II using neutron scattering data.

We examine here the problem of deducing thermodynamic quantities from statistical mechanics when the energy levels are temperature (as well as volume) dependent. Application of these ideas is to He II when the experimental energies for quasi particles are known, at least at relatively low temperatures, from inelastic neutron scattering.

We suppose that the internal energy $E(\underline{n}, V)$ of a system of elementary Bose excitations can be expressed in terms of the system volume V, and the occupation numbers n_α, in a set $\underline{n}$ defined by discrete modes α. We now define the energy level ϵ_α, for mode α, by

$$\epsilon_\alpha(\underline{n}, V) = \partial E(\underline{n}, V)/\partial n_\alpha \,. \tag{1}$$

We make the hypothesis that all statistical weighting of significance arises from $\underline{n}$ so close to the neighborhood of some $\underline{\omega}$, that the truncated expansion

$$E(\underline{n}, V) = E(\underline{\omega}, V) + \sum_\alpha (n_\alpha - \omega_\alpha)\,\epsilon_\alpha(\underline{\omega}, V) \,, \tag{2}$$

suffices. The choice of $\underline{\omega}$ is open, and a criterion for selecting it will be given presently. The mean occupation numbers $\bar{n}_\alpha$ can be obtained as $\bar{n}_\alpha = -\beta^{-1}(\partial \ln Z/\partial \epsilon_\alpha)$ where Z is $\Sigma_{\underline{n}} \exp -\beta E$ whose essential part for fixed $\underline{\omega}$ is

$$Z = \prod_\alpha \sum_{n_\alpha} \exp\{-\beta n_\alpha \epsilon_\alpha(\underline{\omega}, V)\} = \prod_\alpha (1 - \exp\{-\beta\epsilon_\alpha(\underline{\omega}, V)\})^{-1} \,, \tag{3}$$

* Research supported by grant DMR72-03221-A01 from the National Science Foundation and AFOSR76-2880 from the Air Force Office of Scientific Research.

where $\beta = 1/kT$. The result is

$$\bar{n}_\alpha = [\exp\{\beta\epsilon_\alpha(\underline{\omega}, V)\} - 1]^{-1} \,. \tag{4}$$

The equilibrium entropy

$$\bar{S} = \sum_\alpha [k \ln(1 + \bar{n}_\alpha) + \bar{n}_\alpha \epsilon_\alpha(\underline{\omega}, V)/T] \,, \tag{5}$$

comes from differentiating the Helmholtz free energy $F = -kT \ln Z$ with respect to T, holding V and the set $\underline{\omega}$ constant.

We adopt the criterion that the optimum choice of $\underline{\omega}$ comes by varying the trial set $\underline{\omega}$ to find the maximum $\bar{S}$, holding E, V and T constant. We may note that by (1) and (2)

$$\left(\frac{\partial E(\bar{n}, V)}{\partial \omega_\gamma}\right)_V = \sum_\alpha \epsilon_\alpha(\underline{\omega}, V) \left(\frac{\partial \bar{n}_\alpha}{\partial \omega_\gamma}\right)_{V,T} + \sum_\alpha (\bar{n}_\alpha - \omega_\alpha) \left(\frac{\partial \epsilon_\alpha(\underline{\omega}, V)}{\partial \omega_\gamma}\right)_V \,, \tag{6}$$

while (4) and (5) give

$$\frac{\partial \bar{S}(\bar{n}, V)}{\partial \omega_\gamma} = \frac{1}{T} \sum_\alpha \left(\frac{\partial \bar{n}_\alpha}{\partial \omega_\gamma}\right)_{V,T} \sum \epsilon_\alpha(\underline{\omega}, V) \,. \tag{7}$$

It follows that $\bar{S}$ is stationary if, for all γ

$$\sum_\alpha (\bar{n}_\alpha - \omega_\alpha)\,[\partial\epsilon_\alpha(\underline{\omega}, V)/\partial\omega_\gamma]_V = 0 \,, \tag{8}$$

which is obeyed if $\omega_\alpha = \bar{n}_\alpha$. This choice allows us to replace $\epsilon(\bar{n}, V)$ by the experimentally measured $\epsilon(T, V)$. The evaluation of $\bar{n}$ is from (4) $n = [e^{\beta\epsilon} - 1]^{-1}$, and the equilibrium entropy (5) becomes

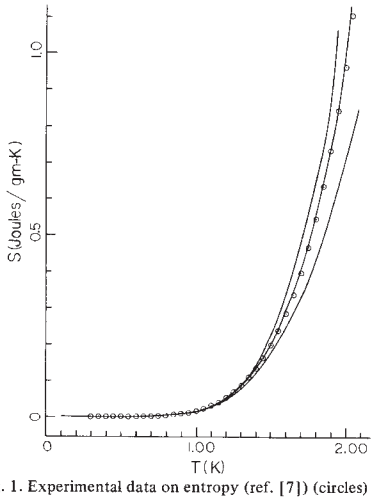


Fig. 1. Experimental data on entropy (ref. [7]) (circles) compared to eq. (9), middle curve. Other curves are discussed in the text.

$$\bar{S} = \frac{V}{2\pi^2}\int_0^\infty \left[k\ln(1+n) + \frac{n\epsilon}{T}\right] q^2\,\mathrm{d}q$$
$$= -\frac{V}{2\pi^2}\int_0^\infty \frac{\epsilon}{3T}\left(\frac{\partial n}{\partial q}\right)_{P,T} q^3\,\mathrm{d}q\,, \qquad (9)$$

where we have transformed to continuous variables and switched to variables P and T for comparison with experiments on He II. Integrating (9) over T for fixed P gives for the Gibbs free energy

$$\Phi = \Phi_0(P) - \frac{kTV}{2\pi^2}\int_0^\infty \ln(1+n)q^2\,\mathrm{d}q$$
$$-\frac{V}{2\pi^2}\int_0^T\int_0^\infty n\left(\frac{\partial\epsilon}{\partial T}\right)_{P,q}\mathrm{d}T q^2\,\mathrm{d}q\,. \qquad (10)$$

Writing $V = \Phi_0'$ for the molar volume at $T = 0$K we obtain the equation of state as an integral equation

$$V = V_0(P)$$
$$-\frac{1}{2\pi^2}\int_0^T \frac{V\,\mathrm{d}T}{T}\int_0^\infty \epsilon\left[\left(\frac{\partial n}{\partial P}\right)_{T,q} + \frac{\kappa q}{3}\left(\frac{\partial n}{\partial q}\right)_{P,T}\right] q^2\,\mathrm{d}q\,, \qquad (11)$$

where $\kappa = -V^{-1}(\partial V/\partial P)_T$ is the compressibility. The term in κ arises from the differentiation of the V in the integrand of (10). The remaining term in (11) follows on integration by parts over q, having made use of the fact that $n = n(\epsilon/T)$ so that

$$\left[\frac{\partial}{\partial P}\left(\frac{\epsilon}{T}\frac{\partial n}{\partial q}\right)\right]_T = \left[\frac{\partial}{\partial q}\left(\frac{\epsilon}{T}\frac{\partial n}{\partial P}\right)\right]_T\,. \qquad (12)$$

To test eq. (9) directly we use the neutron scattering data at the vapor pressure of Yarnell et al. [1] and Henshaw and Woods [2] and the model dispersion curve method of Brooks and Donnelly [3] to obtain the data of fig. 1 (middle curve). The classical expression amounts to omitting the final term of (10) and using the dispersion curve for $T = 0$K, the lower curve results. If finite temperature data for ϵ is used omitting the final term of eq. (11) the upper curve results. The correct formulae greatly extend the usefulness of the neutron data. Nevertheless, the neutron data of Dietrich et al. [4] shows that at high pressures and temperatures, the deconvolved energies do not yield accurate thermodynamic data, and further research is needed to discover how to find the relevant energy levels from neutron scattering when the line width is large.

The form (9) for $\bar{S}$ was first proposed by Bendt et al. [5] for He II assuming $\epsilon = \epsilon(q, T)$ alone. The present derivation is generally valid for all P and T.

Equation (4) may also be obtained by maximizing the entropy

$$S = k\sum_\alpha [(1+n_\alpha)\ln(1+n_\alpha) - n_\alpha \ln n_\alpha]\,, \qquad (13)$$

for given E, V and T; substituting eq. (4) into (13) recovers (5) and hence (9). The usefulness of eq. (13) in the presence of interactions suggests that as interactions between quasi particles become increasingly strong, the classification of states α remains intact, and hence the form of the entropy (13) is preserved. Some of our colleagues believe that this is an essentially weaker assumption than our basic assumption (2). In either case the entropy (9) is the starting point for deriving the thermodynamics and the resulting expressions will be used to generate tables of all the thermodynamic properties of He II [6].

We are grateful to Professors R.P. Feynman, D.L. Goodstein and R.M. Mazo for a number of discussions.

References

[1] J.L. Yarnell, G.P. Arnold, P.J. Bendt, and E.C. Kerr, Phys. Rev. 113 (1959) 1379.
[2] D.G. Henshaw and A.D.B. Woods, Phys. Rev. 121 (1961) 1266.
[3] J.S. Brooks and R.J. Donnelly, Phys. Lett. 46A (1973) 111;
see also J.S. Brooks, Ph.D. thesis, University of Oregon (1973).
[4] D.W. Dietrich, E.H. Graf, C.H. Huang and L. Passell, Phys. Rev. A5 (1972) 1377.
[5] P.J. Bendt, R.D. Cowan and J.L. Yarnell, Phys. Rev. 113 (1959) 1386.
[6] J.S. Brooks and R.J. Donnelly, J. Phys. Chem. Ref. Data (to be published).
[7] R.J. Donnelly, Experimental Superfluidity, University of Chicago Press, Chicago, 1967.

THE ELECTRON BUBBLE IN LIQUID HELIUM

P. H. Roberts
University of Newcastle upon Tyne,
Newcastle upon Tyne, NE1 7RU, England

ABSTRACT

The Bose condensate model of helium is used to examine the structure of the electron bubble in helium. The solution obtained makes use of the fact that the parameter $(am/\ell M)^{1/5}$ is small and m/M is negligible, where m is the electron mass, M is the boson mass, ℓ is the electron-boson scattering length, and a is the healing length. It is shown that, to leading order, the radius of the bubble is $b = (\pi M^2 a^2/m\rho_\infty)^{1/5}$, where ρ_∞ is the helium density. The effects of (quantum) surface tension and of polarization are discussed, and are shown to be small. Consideration is given to the effective mass and radius of the bubble, and the ellipticity induced in it by slow motion is given. The normal modes of pulsation of the bubble are found and the mobility of the ion is computed. The theory is compared with experiment.

1. INTRODUCTION

It has become increasingly apparent over the past decade that the deliberately introduced impurity can be a fruitful experimental probe of the structure and dynamics of helium II, the superfluid phase of helium. Of particular interest is the negative ion which consists of an electron that, through its zero point motion, carves out a soft bubble of about 16Å in radius in the surrounding helium (1Å = 10^{-8} cm.). The induced hydrodynamic mass of such a large structure is greatly in excess of its physical mass, and it therefore responds to applied forces as would a much more massive ion. The experimental situation has been reviewed by Donnelly[1], and more recently by Fetter[2].

The negative ion provides an interesting and, as we shall see, a sensitive testing ground for theories of helium II. We examine in this paper one particularly simple model of helium near absolute zero, the Bose condensate. The approach is expounded by, for example, Gross[3] and by Fetter and Walecka[4]. The theory is so simple to apply that most of the properties of the electron bubble can be calculated in an elementary way. We will present our arguments in a hydrodynamic framework originally proposed by Madelung[5]. Since this may be unfamiliar to the reader, it is developed in §2 for the simple single-particle Schrödinger equation. It is generalized in §3 to the Bose condensate. The two theories are brought together in §4, where the theory of the electron bubble is developed. The final section (§5) confronts the theory with experiment.

2. MADELUNG'S TRANSFORMATION

It appears to have been Madelung[5] who first realized that Schrödinger's equation could be cast into a fluid mechanical mold, by expressing the wave function, $\psi(\mathbf{x}, t)$ in terms of its amplitude, $f(\mathbf{x}, t)$, and phase, $\phi(\mathbf{x}, t)$. Consider a particle of mass m in a field

of fixed potential, $w(\mathbf{x})$, and therefore obeying

$$i\hbar\partial\psi/\partial t = -(\hbar^2/2m)\nabla^2\psi + w\psi\,. \tag{1}$$

By writing

$$\psi = f\exp(im\phi/\hbar), \tag{2}$$

where f and ϕ are real, we can divide (1) into

$$2\,\frac{\partial f}{\partial t} + f\nabla^2\phi + 2\boldsymbol{\nabla} f\cdot\boldsymbol{\nabla}\phi = 0\,, \tag{3}$$

$$\frac{\partial\phi}{\partial t} + \frac{1}{2}(\boldsymbol{\nabla}\phi)^2 + \frac{w}{m} - \frac{\hbar^2}{2m^2}\,\frac{\nabla^2 f}{f} = 0\,. \tag{4}$$

When we introduce the (probabilistic) mass density, ρ, and current $\mathbf{j}$, by writing

$$\rho = m|\psi|^2 = mf^2, \qquad \mathbf{j} = (\hbar/2i)(\psi^*\boldsymbol{\nabla}\psi - \psi\boldsymbol{\nabla}\psi^*) = \rho\boldsymbol{\nabla}\phi\,, \tag{5}$$

and define a velocity, $\mathbf{u}$, by their ratio

$$\mathbf{u} = \mathbf{j}/\rho = \boldsymbol{\nabla}\phi\,, \tag{6}$$

we recognize that (3) and (4) are the continuity and momentum equations governing the potential flow (6):

$$\frac{\partial\rho}{\partial t} + \boldsymbol{\nabla}\cdot(\rho\mathbf{u}) = 0\,, \tag{7}$$

$$\frac{\partial\phi}{\partial t} + \frac{1}{2}\mathbf{u}^2 + \frac{w}{m} + \mu = 0\,, \tag{8}$$

where

$$\mu = -(\hbar^2/2m^2)\;\rho^{-1/2}\nabla^2\rho^{1/2}\,. \tag{9}$$

There are three main differences from classical potential flow. First, the total quantity of 'fluid' is not only conserved by (7): it is fixed by

$$\int|\psi|^2 d\mathbf{x} = 1, \quad \text{or} \quad \int\rho d\mathbf{x} = m\,. \tag{10}$$

Second, even if through the presence of walls ($w = \infty$) or otherwise the fluid is confined to a certain multiply-connected domain, $\mathcal{R}$, ψ must remain single-valued. It follows from (2) that, round any contour Γ in $\mathcal{R}$ not reducible to a point by a continuous deformation, ϕ can change by a multiple of h/m only. The circulation round Γ cannot freely take any value: it is quantized by the Bohr-Sommerfeld condition

$$\oint_\Gamma \mathbf{u}\cdot d\mathbf{x} = \frac{nh}{m}, \quad (n = 0,\ \pm1,\ \pm2,\ldots)\,. \tag{11}$$

Third, a completely new term has appeared in the momentum equation (8), namely μ.

The term μ is often called 'the quantum pressure'. This is a misnomer for at least three reasons. First, its dimensionality is incorrect, and it would be better regarded as a chemical potential per unit mass. Second, since

$$\partial\mu/\partial x_i = -\rho^{-1}\partial\sigma_{ij}/\partial x_j\,, \tag{12}$$

where[6-8]

$$\sigma_{ij} = \frac{\hbar^2}{2m^2}\,\rho\,\frac{\partial^2}{\partial x_i \partial x_j}\,(\ln\,\rho^{1/2})\,, \tag{13}$$

a rival, and properly dimensioned, contender for the title of quantum pressure exists as part of the unusual and complicated stress tensor (13). Third, the word 'pressure' suggests a phenomenon that depends only on 'the local thermodynamic state' (here fixed by ρ), and the presence of derivatives in (9), or (13), shows that all neighboring points are involved in its definition. Despite these objections, we follow the common usage.

As may be seen by setting $\hbar = 0$ in (9), the quantum pressure is the essential ingredient that distinguishes our subject from the classical theory. The fluid dynamicist can gather experience of its effects by translating some of the elementary situations of quantum theory into their corresponding fluid mechanical statements.

'Hydrostatics' arises from the quantum mechanical bound states by writing

$$\phi = -Et\,, \tag{14}$$

where E, the energy of the state, is a constant. By (8) the quantum pressure balances $E - w$ everywhere. It is best to avoid the usual fluid mechanical practice of absorbing E into w, since some energy levels may be inaccessible. For example, when (8) is written as

$$-(\hbar^2/2m)\nabla^2 f = [E - w(\mathbf{x})]f\,, \tag{15}$$

and it is supposed that w increases indefinitely with distance, r, from some origin, O, one family of solutions to (15) is found that increase with r, so that the normalization integrals (10) do not converge. The condition that only the normalizable solutions of the other family are used transforms (15) into an eigenvalue problem that confines E to discrete levels. Of course a continuum of eigenvalues exists when w is bounded above.

A well-known application of (15), that is particularly relevant to the bubble, is the potential well for which

$$w = \begin{cases} w_I, & \text{in } r < b \ \text{('Region I')}, \\ w_{II}, & \text{in } r > b \ \text{('Region II')}, \end{cases} \tag{16}$$

where w_I and w_{II} $(> w_I)$ are constants. Writing

$$\lambda_I^2 = (2m/\hbar^2)(E - w_I), \qquad \lambda_{II}^2 = (2m/\hbar^2)(w_{II} - E)\,, \tag{17}$$

we see that, for $w_I < E < w_{II}$, (15) is obeyed by

$$f = f_I \equiv A j_\ell(\lambda_I r) Y_\ell(\theta, \chi), \qquad r \le b\,; \tag{18}$$

$$f = f_{II} \equiv Ak_\ell(\lambda_{II}r)[j_\ell(\lambda_I b)/k_\ell(\lambda_{II}b)]Y_\ell(\theta,\chi), \quad \text{in } r \geq b\,, \tag{19}$$

where $j_\ell(z)$ is the spherical Bessel function of the first kind, $k_\ell(z)$ is the modified spherical Bessel function of the second kind, and $Y_\ell(\theta,\chi)$ is a surface harmonic of integral degree, ℓ, in spherical coordinates (r,θ,χ). The exclusion of the other spherical Bessel functions ensures that (10) can be met for some choice of the constant A. Continuity of f has been realized, and ∇f is continuous provided

$$\lambda_I b j'_\ell(\lambda_I b)/j_\ell(\lambda_I b) = \lambda_{II} b k'_\ell(\lambda_{II}b)/k_\ell(\lambda_{II}b)\,. \tag{20}$$

This dispersion relationship determines a discrete spectrum of admissible E. It may be seen that, when $\Delta w \equiv w_{II} - w_I$ is large compared with h^2/mb^2, eigensolutions exist for which $E - w_I \ll w_{II} - E$. For these, (19) takes the approximate form

$$f_{II} \approx A\exp[-\lambda_{II}(r-b)]Y_\ell(\theta,\chi), \quad r \geq b\,. \tag{21}$$

The fluid is confined in region II to a boundary layer of thickness $1/\lambda_{II}$, or

$$a_m = \hbar(2m\Delta w)^{-1/2}\,. \tag{22}$$

This phenomenon is often called 'healing', the layer itself a 'healing layer', and a_m the 'healing length'. To the fluid mechanicist, the abrupt increase Δw in w at $r = b$ can only be hydrostatically balanced by an equally abrupt increase in quantum pressure. It is the non-local character of quantum pressure that causes the fluid to pass through the barrier at $r = b$ and, when lower potentials are available externally, permits it to seep out of region I.

In view of later developments, it is worth elaborating the situation just described. First we allow the well to have any shape, defining n to be a coordinate that measures distance from the discontinuity normally outwards from I to II. Second, we allow w_I and w_{II} to vary, and suppose that the large transition from w_I up to w_{II} does not occur abruptly as in (16), but continuously over some distance, a, comparable with a_m. Within this distance, there is evidently no unique way of defining 'the' surface, S, of the well. We note, however, that in the case of discontinuous w just considered, $j_\ell(\lambda_I b)$ is small, by (20). This suggests we should locate S on the surface of zero f_I.

To elucidate the healing layer structure, we introduce a stretched coordinate, ξ, and cast (15) into dimensionless form by writing

$$\xi = n/a, \qquad f = ax(\xi), \tag{23}$$

$$q = a/a_m, \qquad w = w_I + g(\xi)\Delta w, \tag{24}$$

where we suppose g is exponentially small at the inner ($\xi = -\infty$) edge of the boundary layer, and is unity at the outer ($\xi = +\infty$) edge. It should be realized that x, w_I, Δw and g will generally depend on position on S. To the first two orders, however, this dependence only occurs parametrically in the solution, and will therefore be suppressed. Writing

$$x = x_0(\xi) + ax_1(\xi) + \ldots\,, \tag{25}$$

substituting into (15), and equating like powers of a, we obtain

$$d^2x_0/d\xi^2 - q^2 g(\xi) x_0 = 0\,, \tag{26}$$

$$d^2x_1/d\xi^2 - q^2 g(\xi) x_1 = -(C_1^{-1} + C_2^{-1}) dx_0/d\xi\,, \tag{27}$$

where C_1 and C_2 are the principal radii of curvature of S at the point concerned. Since f_{II} is identically zero, the solutions to (26) and (27) must obey

$$x_0 \to 0, \qquad x_1 \to 0, \qquad \text{for} \quad \xi \to +\infty\,. \tag{28}$$

Successful matching to the interior solution, f_I, requires

$$x_0 \sim (\partial f_I/\partial n)_S \xi, \qquad x_1 \sim -\tfrac{1}{2}\,(C_1^{-1} + C_2^{-1})(\partial f_I/\partial n)_S \xi^2\,, \quad \text{for} \quad \xi \to -\infty\,. \tag{29}$$

Explicit solutions can generally be obtained only by numerical means. They obey integral conditions which we will later find useful (§4):

$$\frac{1}{q^2}\left(\frac{\partial f_I}{\partial n}\right)_S^2 = \int_{-\infty}^{\infty} x_0^2 \,\frac{dg}{d\xi}\, d\xi\,, \tag{30}$$

$$-\frac{2}{q^2}\left(\frac{1}{C_1} + \frac{1}{C_2}\right) \bar{\int}_{-\infty}^{\infty} \left(\frac{dx_0}{d\xi}\right)^2 d\xi = \int_{-\infty}^{\infty} 2x_0 x_1 \,\frac{dg}{d\xi}\, d\xi\,, \tag{31}$$

where the bar through the integral sign signifies that the convergent part of the integral is taken. Despite appearances, this integral is negative.

The integral relationships (30) and (31) may be interpreted in the light of the Grant tensor (13). The terms on the left give the main parts of the leap of σ_{nn} across the boundary layer; the terms on the right give the corresponding integrated effects of the external force $f^2\nabla w$ balancing them. The dominant term, given on the left of (30), arises from the 'pressure' of the particle trying to escape the well. The next largest term, shown on the left of (31), clearly has the form of a quantum surface stress, with

$$T_m = \frac{2a\Delta w}{q^2} \bar{\int}_{-\infty}^{\infty} \left(a\,\frac{dx_0}{d\xi}\right)^2 d\xi \tag{32}$$

as coefficient of surface tension. It is easy to make an estimate of T_m. If we take g to be a unit step function, we find from (26) that

$$T_m \approx -a_m^3 \Delta w\, (\partial f_I/\partial n)_S^2\,. \tag{33}$$

Passing from 'hydrostatics' to 'hydrodynamics' by abandoning (14), we see from (8) that another type of healing phenomenon will occur when u is large. A particularly significant case occurs at a vortex line where, by (11), u is of order $nh/m\varpi$, at small distances, ϖ, from the vortex axis. It follows from (8) and (9) that ρ is of order ϖ^{2n} for $\varpi \to 0$. The fact that ρ is zero on the axis itself means that a closed vortex ring, or a vortex

line terminating on boundaries ($w = \infty$), will transform an otherwise simply-connected container into a multiply-connected domain, $\mathcal{R}$, so justifying *a posteriori* the application of (11). Unlike the healing at a wall considered earlier, the depression of ρ at the vortex axis occurs over distances comparable with the scale of the container. The corresponding vortices in the condensate discussed in §3 have cores confined to much smaller distances from their axes.

Before concluding this section we make one remark, obvious perhaps, but relevant to §4. When the particle is trapped in a potential well with moving walls [$w = w(\mathbf{x}, t)$], ϕ is necessarily non-zero and ρ is time-dependent. Nevertheless, provided the time-scales over which w changes are large compared with the reciprocal of the quantum frequency h/mb^2, we can regard the fluid as being in a quasi-hydrostatic state, ignore the time derivative in (1), and treat t in w parametrically. In quantum language, the Born-Oppenheimer approximation is said to apply.

3. THE CONDENSATE MODEL

We now consider an assembly of N identical particles (bosons) of mass M in a potential field $W(\mathbf{x})$. If the particles did not interact, the wave function for the system could be written down as a symmetrized product of the N one-particle wave functions, $\psi(\mathbf{x}, t)$, obeying (11) with W and M replacing w and m. It would be probably more convenient, however, to replace the normalization condition (10) by

$$\int |\Psi|^2 d\mathbf{x} = N, \quad \text{or} \quad \int \rho d\mathbf{x} = \rho_\infty v\,, \tag{34}$$

where v is the volume of the system and $\rho_\infty = MN/v$. The resulting theory is well-understood, and contains features that fruitfully represent helium near absolute zero[8]. It may be seen from (18), however, that the ground state for, say, the potential well (16) would be one in which all the particles would be at the origin, with high probability. To eliminate this unphysical behavior, the imperfect Bose condensate has been devised. A short-range repulsive potential $V(\mathbf{x} - \mathbf{x}')$ is introduced in an *ad hoc* way, and to W the potential

$$\int V(\mathbf{x} - \mathbf{x}')|\Psi(\mathbf{x}')|^2 d\mathbf{x}' \tag{35}$$

is added which increases as the density of nearby bosons increases. The simplest case arises when V is taken to be

$$V(\mathbf{x} - \mathbf{x}') = V_0 \delta(\mathbf{x} - \mathbf{x}')\,. \tag{36}$$

Equation (1) is then replaced by

$$i\hbar \partial\Psi/\partial t = -(\hbar^2/2M)\nabla^2\Psi + (V_0|\Psi|^2 + W)\Psi\,. \tag{37}$$

A fuller and more satisfactory derivation of (37) may be found elsewhere[3,4]. It is of some interest that nonlinear Schrödinger equations of the form (37) have been the subject of close scrutiny in recent years in non-quantal contexts, particularly in theories of weakly nonlinear waves and stability[9]. The Madelung transformation

$$\Psi = F \exp(iM\Phi/\hbar)\,, \tag{38}$$

follows the course of §2 with minor changes. Most significant is the addition of a 'gas pressure',

$$p = (V_0/2M^2)\rho^2\,, \tag{39}$$

which (multiplied by δ_{ij}) should be included in the stress tensor (13). Thus (8) is replaced by

$$\frac{\partial\Phi}{\partial t} + \frac{1}{2}\mathbf{u}^2 + \frac{W}{M} + \frac{2p}{\rho} + \mu = 0\,. \tag{40}$$

The presence of the repulsion, V_0, and its associated gas pressure restores a number of physical effects absent in §2. The tendency towards condensation is eliminated for all sufficiently large systems. To see this, return to the hydrostatic theory of §2 and the potential well (16). The spherically symmetric ($\ell = 0$) ground state now obeys

$$\frac{d^2F}{dr^2} + \frac{2}{r}\frac{dF}{dr} = \frac{2M}{\hbar^2}(E - W)F - \frac{2MV_0}{\hbar^2}F^3\,. \tag{41}$$

If NV_0 and ΔW are both large compared with $\hbar^2/Mb^2$, (41) gives everywhere except near the surface, S, of the well

$$\rho = \rho_\infty = MF_I^2 = MN/v\,, \tag{42}$$

the corresponding one-particle energy being given by

$$E - W_I = \rho_\infty V_0/M\,. \tag{43}$$

The fluid is spread out uniformly in the well.

Near S the derivatives of F become large, and the constant solution (42) breaks down. We may follow the argument of §2. Introducing a new healing length

$$a = \hbar(2\rho_\infty V_0)^{-1/2}\,, \tag{44}$$

writing

$$\xi = n/a, \quad F = X(\xi), \quad W = W_I + (\rho_\infty V_0/M)G(\xi)\,, \tag{45}$$

where $G(\xi)$ is exponentially small for $\xi \to -\infty$, expanding X as

$$X(\xi) = X_0(\xi) + aX_1(\xi) + \dots\,, \tag{46}$$

substituting into (15), and equating like powers of a, we obtain

$$d^2X_0/d\xi^2 - [G(\xi) - 1 + (X_0/F_I)^2]X_0 = 0\,, \tag{47}$$

$$d^2X_1/d\xi^2 - [G(\xi) - 1 + 3(X_0/F_I)^2]X_1 = -(C_1^{-1} + C_2^{-1})dX_0/d\xi\,. \tag{48}$$

Matching at the edges of the boundary layer requires

$$X_0 \to 0, \qquad X_1 \to 0, \qquad \text{for} \quad \xi \to +\infty\,, \tag{49}$$

$$X_0 \to F_I, \qquad X_1 \to 0, \qquad \text{for} \quad \xi \to -\infty\,. \tag{50}$$

Again, explicit solutions generally require numerical integrations, though useful integral relations may be established, for instance

$$\frac{1}{2}F_I^2 = \int_{-\infty}^{\infty} X_0^2 \frac{dG}{d\xi}\, d\xi\,, \tag{51}$$

$$-2\left(\frac{1}{C_1}+\frac{1}{C_2}\right)\int_{-\infty}^{\infty}\left(\frac{dX_0}{d\xi}\right)^2 d\xi = \int_{-\infty}^{\infty} 2X_0X_1\frac{dG}{d\xi}\, d\xi\,. \tag{52}$$

Once more, the dominant contribution to the leap in stress across the healing layer arises from the interior solution, although now it is the gas pressure and not the quantum pressure that is mainly responsible. Again, the next largest term can be interpreted as a quantum surface tension, with positive coefficient

$$T_M = 2a\left(\frac{\rho_\infty V_0}{M}\right)\int_{-\infty}^{\infty}\left(\frac{dX_0}{d\xi}\right)^2 d\xi\,. \tag{53}$$

As before, T_M may be estimated[3] from a simple model of X. If we suppose that G is a step function of infinite height ($\Delta W = \infty$), (46) may be solved as $X_0 = F_I \tanh(-\xi/\sqrt{2})$, and (53) gives

$$T_M \approx \sqrt{2}\hbar^2\rho_\infty/3M^2a\,. \tag{54}$$

Passing again from 'hydrostatics' to 'hydrodynamics', we note that the gas pressure can supply the restoring force necessary for compressional waves. Perturbing about the static solution (41), we readily find that long wavelength sound propagates at the velocity

$$c = \sqrt{(dp/d\rho)} = \sqrt{(2p/\rho)} = \hbar/Ma\sqrt{2}\,. \tag{55}$$

At wavelengths of order a and smaller, the quantum pressure increases the phase speed, increases the group velocity, and introduces weak dispersion.

Vortex lines may be studied as in §2. Unlike their classical counterparts, the cores of these vortices do not have sharply defined surfaces separating regions of zero and non-zero vorticity. All the vorticity they contain is concentrated as δ-functions on their axes. Such a vortex, if classical, would have infinite self-energy. Here, however, the density decreases over the characteristic distance a as the axis is approached, so ensuring a finite tension. The depletion of fluid in the core makes the vortex resemble the classical hollow core model. The permanence of vortex rings implied by the Kelvin-Helmholtz theorem makes them excellent candidates for quasi-particle models, so reviving in a novel context the ideas underlying the vortex atoms proposed by Kelvin in the nineteenth century.

4. STRUCTURE OF THE NEGATIVE ION

It is possible[3,10–12] to account with relative ease for many features of the negative ion by combining the methods of §§2 and 3 above. We use the theory of §1 to represent the electron, regarding w as the potential created by the surrounding condensate; we apply the

formalism of §2 to the *exterior* of the bubble, taking for W the potential of the electron. More explicitly, we introduce the energy,

$$\iint U(\mathbf{x}-\mathbf{x}')|\psi(\mathbf{x})|^2|\Psi(\mathbf{x}')|^2 d\mathbf{x}d\mathbf{x}', \tag{56}$$

representing the repulsion of an electron at $\mathbf{x}$ and a boson at $\mathbf{x}'$. Taking again the simplest case of a $\delta-$function interaction

$$U(\mathbf{x}-\mathbf{x}') = U_0\delta(\mathbf{x}-\mathbf{x}'), \tag{57}$$

we then have

$$w(\mathbf{x}) = \int U(\mathbf{x}-\mathbf{x}')|\Psi(\mathbf{x}')|^2 d\mathbf{x}' = U_0|\Psi(\mathbf{x})|^2, \tag{58}$$

$$W(\mathbf{x}') = \int U(\mathbf{x}-\mathbf{x}')|\psi(\mathbf{x})|^2 d\mathbf{x} = U_0|\psi(\mathbf{x}')|^2. \tag{59}$$

By (10) and (58), $\Delta w = U_0\rho_\infty/M$ so that by (22), (23) and (44), $q^2 = mU_0/MV_0$. Equations (1) and (37) become coupled:

$$i\hbar\partial\psi/\partial t = -(\hbar^2/2m)\nabla^2\psi + U_0|\Psi|^2\psi, \tag{60}$$

$$i\hbar\partial\Psi/\partial t = -(\hbar^2/2M)\nabla^2\Psi + (V_0|\Psi|^2 + U_0|\psi|^2)\Psi. \tag{61}$$

As in §2, we define the electronic surface, S, of the bubble by the zero of ψ.

The key to a simple 'hydrostatic' solution of (60) and (61) lies in the fact, which we can verify *a posteriori*, that the radius, b, of the bubble is large compared with a and a_m, so that the boundary layer methods of §§2 and 3 can be used with minor emendations. We must not forget however that, since the roles of interior and exterior of the bubble have been exchanged for the condensate, the sign of ξ in §3 must be reversed. The mainstream value of F, denoted in §3 by F_I, is now written F_S.

To leading order, we set $g(\xi) = X_0^2/F_S^2$ and $G(\xi) = (a^2U_0/F_S^2V_0)x_0^2$ in (26) and (47), as (58) and (59) require. The integral relations (30) and (51) may then be combined to give

$$\frac{a^2}{q^2}\left(\frac{\partial f}{\partial n}\right)_S^2 - \frac{V_0}{2U_0}F_S^2 = \frac{a^2U_0}{F_S^2}[x_0^2X_0^2]_{-\infty}^{\infty} = 0, \tag{62}$$

where we have appealed to (28) and (50). To the next order, the forms of g and G require reconsideration[12]. In place of (27) and (48) we have

$$d^2x_1/d\xi^2 - (q/F_S)^2(X_0^2x_1 + 2x_0X_0X_1) = -(C_1^{-1} + C_2^{-1})dx_0/d\xi, \tag{63}$$

$$d^2X_1/d\xi^2 - (3X_0^2/F_S^2 - 1)X_1 - (a^2U_0/F_S^2V_0)(x_0^2X_1 + 2X_0x_0x_1) = -(C_1^{-1} + C_2^{-1})dX_0/d\xi. \tag{64}$$

The integral consequences (31) and (52) are modified accordingly, and the result (62) is altered to

$$\frac{a^2}{q^2}\left(\frac{\partial f}{\partial n}\right)_S^2 - \frac{V_0}{2U_0}F_S^2 = 2a\left(\frac{1}{C_1} + \frac{1}{C_2}\right)\int_{-\infty}^{\infty}\left[\frac{a^2}{q^2}\left(\frac{dx_0}{d\xi}\right)^2 + \frac{V_0}{U_0}\left(\frac{dX_0}{d\xi}\right)^2\right]d\xi, \tag{65}$$

which now includes the effects of interfacial tension.

The jump conditions (62) or (65) across the boundary layer suffice to match the mainstream electron solution to the mainstream condensate solution. Applied to the electron bubble, we have by (18)

$$f = A(\sin \lambda_I r)/\lambda_I r\,, \tag{66}$$

showing that $\lambda_I b = \pi$. By (10) the solution is normalized to the first two orders[12] in a/b, if $A^2 = \pi/2b^3$. By (10) and (11) we obtain

$$b = \left(\frac{\pi M^2 a^2}{m\rho_\infty}\right)^{1/5} - \frac{8M^2 a}{5\rho_\infty} \fint_{-\infty}^{\infty} \left[\frac{1}{m}\left(a\,\frac{dx_0}{d\xi}\right)^2 + \frac{1}{M}\left(\frac{dX_0}{d\xi}\right)^2\right] d\xi\,. \tag{67}$$

It is not at once clear whether b will be decreased by the positive surface tension (53) of the condensate or increased by the negative surface tension (32) of the electron. If we use estimates (33) and (54) however, we see that $|T_m|/T_M$ is of order q^{-1} and, since experiments indicate (§5) that $q < 1$, it appears that the bubble radius should be larger than $(\pi M^2 a^2/m\rho_\infty)^{1/5}$. Direct numerical integrations[12] of x_0 and X_0, and evaluation of the integral seen in (67), suggest that the difference is of the order of a.

The effects of polarization induced by the electron in the surrounding helium can be included by adding[12] to (56) the term

$$-(\tilde{\alpha}e^2/8\pi) \iint |\psi(\mathbf{x})|^2 |\Psi(\mathbf{x}')|^2 |\mathbf{x} - \mathbf{x}'|^{-4} d\mathbf{x} d\mathbf{x}'\,, \tag{68}$$

where $\tilde{\alpha}$ is the polarizability of the helium and e is the electronic charge. This has the effect of contracting the bubble by order $\tilde{\alpha}Me^2a^2/4\pi\hbar^2b^3$. A detailed theory[12] shows that the reduction is of order $a/3$ in the practically interesting cases.

Further complications arise when the dynamics of the bubble are considered, although the time-scales of interest are usually large enough compared with the electronic frequencies to justify the neglect of $\partial\psi/\partial t$ in (60); see §2. To evaluate the effective hydrodynamic mass, we consider the bubble in steady motion U, at small Mach numbers $\mathcal{M} = U/c$. The electronic radius, b, of the bubble is increased by about $5\mathcal{M}^2\%$ because of the pressure forces associated with the flow of condensate over its surface. It is also made slightly oblate, with an ellipticity close to $\mathcal{M}^2/2$. Ignoring this effect, it is found that the dipolar back-flow created by the ion coincides with that of a hard sphere whose radius, b_e, is less than the electronic radius, b, by one to two healing lengths:

$$b_e = b - (aM/\rho_\infty) \fint_{-\infty}^{\infty} X_0^2(\xi) d\xi\,. \tag{69}$$

It is this radius, rather than b, that determines the induced mass of the ion.

Further details of the calculations outlined above may be found in the paper by Roberts and Grant already cited[12]. We conclude this section by breaking new ground. We consider the oscillations of the bubble, their implications for phonon-ion collisions, and the mobility of negative ions at low temperatures. We again adopt the boundary layer methods

described above but, of course, retain the term $\partial\psi/\partial t$ in (61), so introducing a velocity potential, Φ, in the condensate. We retain only the dominant part of the boundary layer structure, excluding both surface tension and polarization effects. We write,

$$f = f_0 + \alpha a f' , \qquad F = F_0 + \alpha a F' , \qquad \Phi = \alpha \Phi' , \tag{70}$$

where the suffix 0 stands for the steady solution obtained earlier, and the terms in α represent time-dependent perturbations, where $0 < \alpha \ll 1$.

It is readily seen from (7) and (40) that Φ' and F' both obey the acoustic wave equation

$$\partial^2 \Phi'/\partial t^2 = c^2 \nabla^2 \Phi' . \tag{71}$$

We could, by following Celli, Cohen and Zuckerman[13], examine solutions in the form of outgoing waves. The eigenfrequencies would be complex, because of the reduction in oscillation amplitude at a point fixed in space as the energy of surface motion is radiated to infinity. The corresponding eigenfunction must tend to infinity with r since, the more distant a wave is, the earlier it must have left the surface, and the greater the amplitude of surface oscillation must then have been. We will not consider solutions of this type below. We will confine our attention to the scattering problem in which an incoming plane wave travelling in the $z-$direction

$$S'_{\mathrm{INC}} = \exp[i(kz - \omega t)] , \tag{72}$$

where k and $\omega = ck$ are real, is scattered by the bubble into a set of outgoing waves. We first aim to calculate the scattering amplitude, $h_\ell(k)$, of the ℓth partial wave[13]

$$h_\ell = \frac{1}{1 + iQ_\ell} , \qquad Q_\ell = \frac{kby_\ell(kb) + K_\ell y'_\ell(kb)}{kbj_\ell(kb) + K_\ell j'_\ell(kb)} , \tag{73}$$

where $y_\ell(z)$ is the spherical Bessel function of the second kind and K_ℓ is the spring constant of the bubble for this mode. We then use h_ℓ to compute the differential cross-section[14] of the bubble

$$\sigma(k, \omega) = k^{-2} \left| \sum_{\ell=0}^{\infty} (2\ell + 1) h_\ell(k) P_n(\cos\theta) \right|^2 , \tag{74}$$

and hence the momentum-transfer cross-section[14]

$$\sigma_T(k) = \iint \sigma(k, \theta)(1 - \cos\theta) \sin\theta d\theta d\chi . \tag{75}$$

From this we finally evaluate the mobility, μ_e, of the ion from[14]

$$\frac{e}{\mu_e} = -\frac{\hbar}{6\pi^2} \int_{-\infty}^{\infty} \sigma_T(k) \frac{\partial n(k)}{\partial k} k^4 dk , \tag{76}$$

where

$$n(k) = [\exp(\hbar ck/KT) - 1]^{-1} , \tag{77}$$

is the density of phonons (in $\mathbf{xk}-$space) at temperature T. Here K is Boltzmann's constant.

To determine the spring constants, K_ℓ, we have to match solutions of (71) across the boundary layer on S', the deformed electronic surface, to the quasi-static solutions (18) of the electron mainstream. We first consider the case $\ell \geq 1$. The fact that

$$\psi = A_0 j_0(\lambda_I r) + \alpha a A' j_\ell(\lambda_I r) Y_\ell(\theta, \chi)\,, \tag{78}$$

implies that S' has the equation

$$r = b(\theta, \chi, t) \equiv b_0 + \alpha a b'(t) Y_\ell(\theta, \chi)\,, \tag{79}$$

where (using $\lambda_I b_0 = \pi$)

$$b'/b_0 = j_\ell(\pi) A'/A_0\,. \tag{80}$$

We will continue to refer the boundary layer structure to the unperturbed position, S_0, of the electronic surface, and not to S'. We introduce x', X' and η', the boundary layer forms of f', F' and Φ', and expand these in ascending powers of a

$$x' = a^{-1}x'_0 + x'_1 + \cdots, \quad X' = a^{-1}X'_0 + X'_1 + \cdots, \quad \eta' = \eta'_0 + a\eta'_1 + \cdots\,, \tag{81}$$

where the coefficients shown depend on t and $\xi = (r - b_0)/a$ and parametrically on θ and χ. We substitute these into the boundary layer forms of (60) and (61) which, in the Madelung framework, give to the first two orders

$$\partial^2 x'/\partial\xi^2 - (q/F_S)^2(X_0^2 x' + 2x_0 X_0 X') = -(2a/b)\partial x'/\partial\xi\,, \tag{82}$$

$$\begin{aligned}\partial^2 X'/\partial\xi^2 - (3X_0^2/F_S^2 - 1)X' - (a^2 U_0/F_S^2 V_0)(x_0^2 X' + 2X_0 x_0 x') \\ = -(2a/b)\partial X'/\partial\xi + (M/F_S^2 V_0) X_0 \partial\eta'/\partial t\,,\end{aligned} \tag{83}$$

$$X_0\partial^2\eta'/\partial\xi^2 + 2(dX_0/d\xi)\partial\eta'/\partial\xi = -(2a/b)X_0\partial\eta'/\partial\xi - 2a^2\partial X'/\partial t\,. \tag{84}$$

It seems clear from (84) that $X_0^2\partial\eta'_0/\partial\xi$ is independent of ξ and, since there is no net condensate flow through the boundary layer at any point, that constant must be zero. Thus η'_0 takes throughout the boundary layer the mainstream value Φ'_S of Φ, evaluated on S_0. The right-hand side of (83) does not contribute to leading order in a, and (82) and (83) may be solved to give

$$x'_0 = \zeta dx_0/d\xi\,, \qquad X'_0 = \zeta dX_0/d\xi\,, \tag{85}$$

where ζ is independent of ξ. These forms represent a net displacement of the equilibrium boundary layer from S_0 to S', without change of form; we conclude that $\xi = -b'Y_\ell$.

In proceeding to the next order we note that, since the velocity of sound (55) is of order $1/a$, the time derivatives in (83) and (84) now contribute. In fact, excluding again a net flux of condensate through the boundary layer, (84) shows that $\partial\eta'_1/\partial\xi$ takes the value $-a^2\partial\zeta/\partial t$ throughout, and in particular on the outer edge ($\xi = \infty$) of the boundary layer. It follows that

$$a\partial b'/\partial t = (\partial\Phi'/\partial r)_S\,, \tag{86}$$

an equation with an obvious interpretation. The equations (82) and (83) again admit an integral, namely

$$\frac{MX_0^2}{2U_0F_S^2}\,\frac{\partial\eta_0'}{\partial t} = \frac{a^2}{q^2}\left[\frac{dx_0}{d\xi}\,\frac{\partial x_1'}{\partial\xi} - x_1'\,\frac{d^2x_0}{d\xi^2}\right]_{-\infty}^{\xi} + \frac{V_0}{U_0}\left[\frac{dX_0}{d\xi}\,\frac{\partial^2X_1'}{\partial\xi^2} - X_1'\,\frac{d^2X_0}{d\xi^2}\right]_{-\infty}^{\xi}. \quad (87)$$

On taking the limit $\xi \to +\infty$, and using (18) to evaluate the contributions from the lower limits, we find

$$\partial\Phi_S'/\partial t = c^2K_\ell ab'/b_0\,, \quad (88)$$

where

$$K_\ell = \begin{cases} 5/2, \text{ if } \ell = 0\,; \\ 1 - \ell + \pi j_{\ell-1}(\pi)/j_\ell(\pi), \text{ if } \ell \geq 1\,. \end{cases} \quad (89)$$

The numerical values of K_ℓ for the first 20 values of ℓ are given in Table I. That of K_0 was obtained from an analysis too similar in spirit to the one just described to be repeated here. It may be noted that K_1 is zero, representing the fact that the bubble is neutrally stable to a uniform displacement. Equations (86) and (88) are applied on S_0, and provide the boundary conditions to which solutions of (71) must be subjected.

We developed a program for an Hewlett-Packard 9820A desk computer to evaluate $\sigma_T(k)$ and μ_e from an arbitrary set of the spring constants. The results were tested with the values ($K_0 = 0.23474$; $K_1 = 0$; $K_2 = 0.45045$; $K_n = 0$, $n > 2$) used by Baym, Barrera and Pethick[14], and good agreement was obtained. The programme was then used to generate the results shown in Figures 1 and 2. The effect of truncating series (74) at $\ell = 2$ and $\ell = 19$ are shown in both cases. The prominent peak in $\sigma_T(k)$ seen in Figure 1 is due to a d-wave resonance ($\ell = 2$). A new minor peak is added every time ℓ is incremented by 1. The curve appears to approach its geometrical value[15] ($\sigma_T/4\pi b_0^2 \to \frac{1}{2}$) with an oscillation of amplitude $(kb_0)^{-2/3}$ and period $2^{-1/3}$. In Figure 2, we see μ_eT^3 plotted in units of L_0, as a function of T measured in units of T_0 where

$$T_0 = \hbar c/b_0K, \qquad L_0 = 3\pi e\hbar^2c^3/2b_0K\,. \quad (90)$$

TABLE 1
Spring Constants

ℓ	K_ℓ	ℓ	K_ℓ
0	2.500000000	10	11.56327795
1	0.000000000	11	12.59928101
2	2.289868134	12	13.62973970
3	3.771253431	13	14.65585350
4	5.032253885	14	15.67849696
5	6.198547165	15	16.69832328
6	7.314641577	16	17.71583055
7	8.400646541	17	18.73140537
8	9.467085072	18	19.74535255
9	10.520037400	19	20.75791573

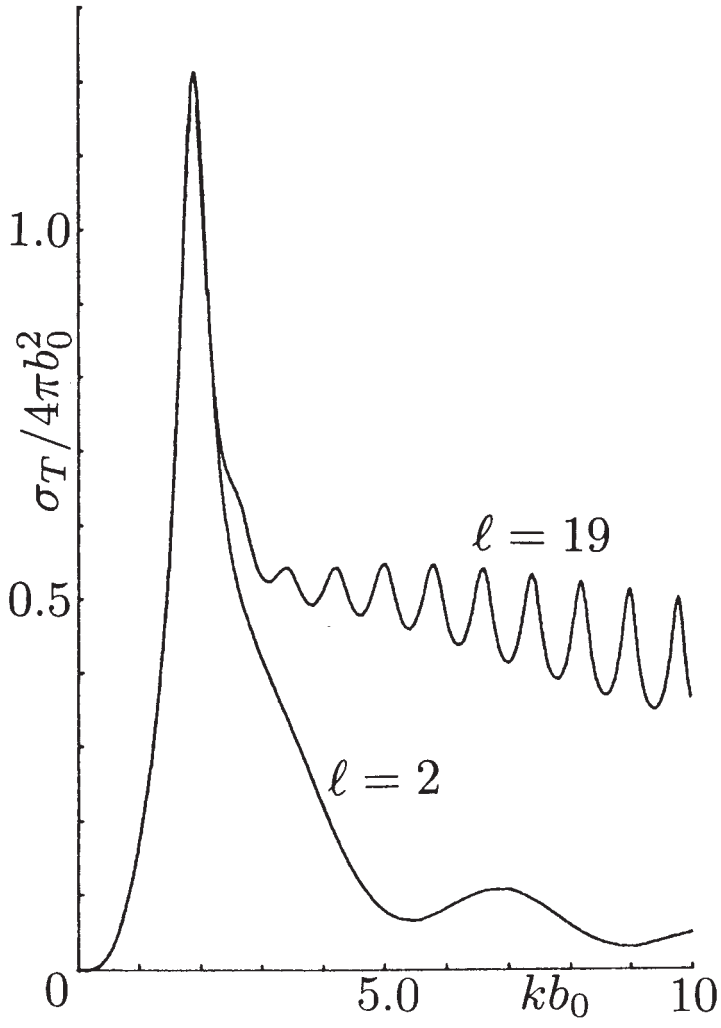


Fig. 1. The momentum-transfer cross-section, σ_T, as a function of wavenumber k. The effects of truncating (74) at $\ell = 2$ and $\ell = 19$ are shown.

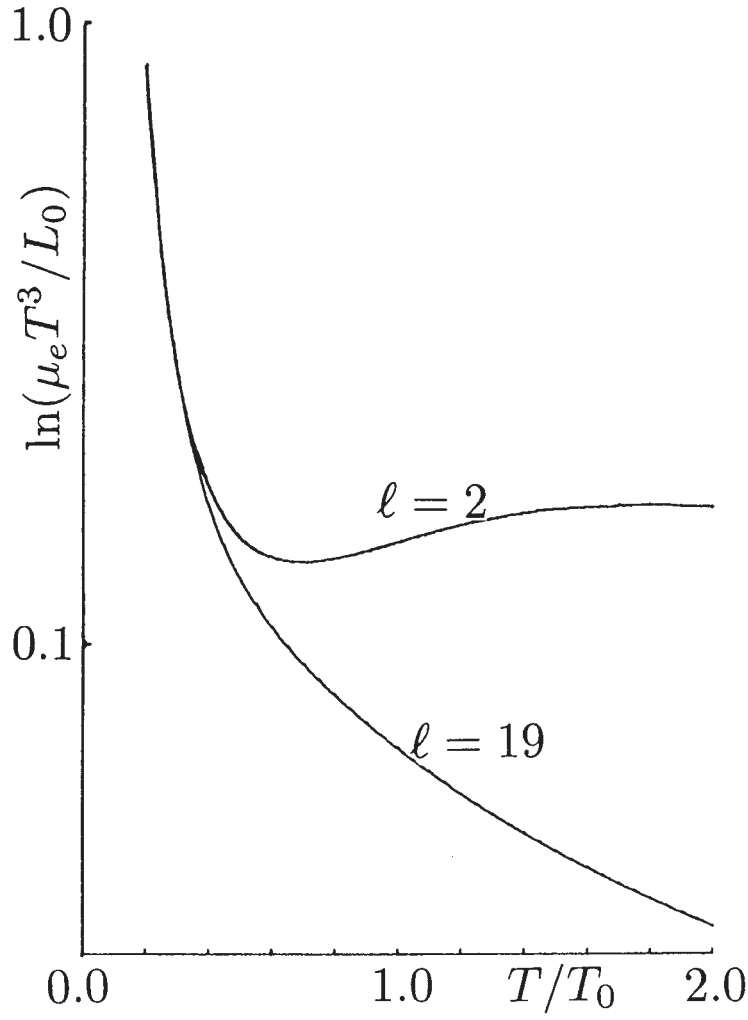


Fig. 2. The mobility, μ_e, as a function of temperature T. The function $\mu_e T^3$ is shown in units of L_0 as a function of T in units of T_0, where L_0 and T_0 are defined by (90). The effects of truncating (74) at $\ell = 2$ and $\ell = 19$ are shown.

5. EXPERIMENTAL COMPARISONS

The condensate model of helium II is essentially a theory having only one disposable parameter, namely the pseudo-potential, V_0, or equivalently the healing length, a. It is natural to seek to choose this so that theory and observation are in optimum accord. Clearly a choice of a made to fit one physical phenomenon well is likely to conflict with others, and an overall consistency with the experimental facts is not to be anticipated. One notes particularly that, since the condensate is a gas obeying the equation of state (39), we should not expect the theory to perform well at the vapor pressure.

One can obtain an estimate of $V_0 = 4\pi d\hbar^2/M$ from measurements of the atomic diameter, d, by α-particle scattering experiments. Values of d of about 2.7Å have been found. If ρ_∞ is 0.145g/cm^3, the healing length would be 0.82Å, leading to too small a velocity of sound. One popular procedure has been to extract a from accurate experimental determinations of the relation between the velocity and energy of circular vortex rings. This had led to estimates of $a \approx 1.28$Å, giving much too small a value of c. The reliability of the approach can, however, be questioned. One would have expected a to decrease with increasing pressure, but the reverse seems to be true[16]. It is now believed[17] that the core of a superfluid vortex is the seat of excitations (normal fluid), and that the surface of a vortex core marks the distance from the axis at which the Landau critical velocity is reached, rather than a quantum healing distance; such a belief is consistent with the increase of a with ρ_∞.

Perhaps the most satisfactory way of estimating a is through the velocity of sound (55). To give an example, if we take $c = 238$ m/s, we obtain $a \approx 0.47$Å and for $\rho_\infty = 0.145\text{g/cm}^3$ we find that $V_0 \approx 1.7\ 10^{-37}\text{g cm}^5/\text{s}^2$, a value admittedly three times larger than the scattering experiments suggest, and moreover one which will alter as c and ρ_∞ change through applied pressure. Nevertheless, by using c to determine a, we obtain a coefficient of surface tension, T_M, from estimate (54) of 0.37g/s^2, which is in good agreement with the experimental value of 0.34g/s^2 at low temperatures.

Turning to the bubble, we see that, in the first approximation, the theory does not require a knowledge of the pseudo-potential, U_0, either for its equilibrium structure or for its oscillation spectrum. The radius, $(\pi M^2 a^2/m\rho_\infty)^{1/5}$, predicted by the first approximation is somewhat small, 11.8Å using the values quoted above. Since c and ρ_∞ increase with increasing pressure, p, this radius decreases with increasing pressure, although somewhat more slowly than experiments indicate. The bubble radius is increased when the effects of surface tension are added. Unfortunately, Grant and Roberts[12] did not examine values of a as small as 0.47Å, so that the value of the integral appearing in (67) is not known. Using our earlier estimates, however, it appears that b will be increased to about 13.3Å by surface tension effects. Table 2 gives $\mu_e T^3$ for a few values of T for both the $\ell = 2$ and the $\ell = 19$ truncations, and for values of b of 11.8Å, 14.3Å and 16.0Å. At the $\ell = 2$ level of truncation, there is a clear tendency for $\mu_e T^3$ to approach a limiting value, of about $36\text{cm}^2\text{K}^3\text{Vs}$ in the case of the 16Å bubble, as T increases. The explanation of this behavior was provided by Baym, Barrera and Pethick[14] in terms of the shape of the d-wave resonance of Fig. 1. Not surprisingly in view of the very different form of σ_T obtained at $\ell = 19$ truncation, the constancy of $\mu_e T^3$ is not as marked at this level. Nevertheless, the values shown for $\ell = 19$ in Table 1 are not ridiculously far from the experimental value[18] of about $32.5\text{cm}^2\text{K}^3\text{Vs}$ in the range of T in which Baym, Barrera and Pethick measured the success of the work.

When we take the theory of the bubble to the second approximation, a new disposable

parameter enters, namely the psuedo-potential, U_0, or equivalently $q = a/a_m$, a relation we can also write as $q^2 = mU_0/MV_0$. Roughly speaking q, as the ratio of the two healing lengths, measures the relative penetration of the condensate wave function into the cavity to the penetration of the electron wave function into the condensate. If q were zero, it would be legitimate to treat the condensate as an abrupt edge and only consider the electronic boundary layer of §2. At first sight it might appear that, since q^2 is proportional to $m/M \doteq 1.37\ 10^{-4}$, it would be admissible to follow Celli, Cohen and Zuckerman[13] in taking this view. The indications are, however, that U_0/V_0 is large. Scattering experiments give an electron-helium scattering length, λ, of about 0.60Å, implying that $U_0 = 2\pi\lambda h^2/m$ is about $4.6\ 10^{-35}$g cm^5/s^2. Taken with the experimental value of $5.7\ 10^{-38}$g cm^5/s^2 for V_0, we obtain $U_0/V_0 \approx 810$ and $q \approx 0.33$. It would be interesting to see whether the effect of restoring q to the Celli-Cohen-Zuckerman theory would have serious repercussions. The indications are that it would not.

The neglect of q in the condensate theory described here would eliminate the condensate surface tension, T_M, and transform the interfacial boundary layer into the structure considered in §2. The associated negative surface tension, T_M, would tend to expand the bubble, an effect confirmed by the calculations of Grant and Roberts[12]. All influences of interfacial tension are, however, of second order in the condensate theory. In the approach of Celli, Cohen and Zuckerman[13], the interfacial tension, σ, is a first order effect. These authors regarded σ as a disposable parameter that could be legitimately chosen to fit the observed bubble radius at the applied pressure of interest. It is easy to verify that the interfacial tension they require is positive and, particularly at higher pressures, several times larger than the condensate surface tension, T_M, considered earlier. It may be wondered why, with this sign difference, the bubble radii they obtain are, being in perfect agreement with experiment, larger than those obtained from the condensate. The answer is to be found in (39). The condensate is a gas and, to obtain agreement with the observed helium densities, it is necessary to choose a large V_0, leading to pressure (39) of the order of 40 atmospheres. In contrast, the Celli, Cohen and Zuckerman theory treats the helium as a classical compressible fluid, not containing the pressure (39), and to avoid large bubble radii at the vapor pressure, a positive interfacial tension is needed. As we have stated above, we regard (39) as an artificial construct of the theory, not to be identified with the applied pressure, and base our comparison with experiment on density and velocity of sound data.

As Schwarz and Stark[18–19] observe, if the spring constants were regarded as disposable parameters, there would be no difficulty in reproducing any ion mobility data precisely. It appears that even the added flexibility given to the theory by the *ad hoc* interfacial tension, σ, already permits an excellent account of the mobilities. Schwarz and Stark[18–19] have shown that, for their spring constants, the constancy of $\mu_e T^3$ in the range of T of interest is not lost when the truncation level is increased as it is in ours. In comparing our theory with theirs, one must be perplexed by the substantial difference in the spring constants and in the shape of the mobility curve (labelled '$\ell = 19$' in Fig. 2). He must wonder if, in the disappointing form of that curve on the present theory, and in the sensitivity of the mobility itself to the healing length [as evinced by the c-dependence of (90)], the condensate theory has not met its most severe test to date. He may also speculate on the physical basis of the *ad hoc* interfacial tension required by the other approach to survive its trial by experiment, and also whether the effects of roton-ion collisions at the higher

temperatures have been underestimated.

TABLE 2
Ion Mobilities, μ_e
(T in °K, $\mu_e T^3$ in units of cm^2K^3Vs)

$\ell = 2$ Truncation						$\ell = 19$ Truncation					
$b_0 = 11.8$Å		$b_0 = 13.3$Å		$b_0 = 16.0$Å		$b_0 = 11.8$Å		$b_0 = 13.3$Å		$b_0 = 16.0$Å	
T	$\mu_e T^3$	T	$\mu_e T^3$	T	$\mu_e T^3$	T	$\mu_e T^3$	T	$\mu_e T^3$	T	$\mu_e T^3$
0.34	229.	0.37	117.	0.31	97.0	0.32	265.	0.33	158.	0.35	70.5
0.41	131.	0.45	77.2	0.43	51.8	0.37	178.	0.42	84.9	0.39	59.0
0.51	86.7	0.52	62.3	0.53	41.7	0.48	95.3	0.46	71.0	0.43	49.0
0.58	70.0	0.56	56.9	0.58	39.5	0.52	79.7	0.52	59.0	0.50	40.0
0.63	63.9	0.64	50.2	0.65	37.6	0.58	66.2	0.57	51.1	0.59	33.1
0.72	56.3	0.70	47.5	0.73	36.6	0.64	57.3	0.71	39.8	0.70	28.1
0.78	53.4	0.78	45.2	0.78	36.5	0.80	44.7	0.85	33.8	0.83	24.6
0.98	49.5	0.94	43.9	0.95	37.4	1.01	36.0	1.01	29.6	1.02	21.0
1.18	49.7	1.15	45.0	1.16	39.9	1.13	30.1	1.13	26.8	1.14	19.4

ACKNOWLEDGEMENT

I am extremely grateful to Professor Russell J. Donnelly for his criticism of this work, and particularly for his contributions to §5. The work described was supported by the Air Force Office of Scientific Research under Grant AF-AFOSR-71-1999 and by the National Science Foundation under Grant NSF GH 35898.

REFERENCES

1. R. J. Donnelly, "Experimental Superfluidity," University Press, Chicago (1967).
2. A. L. Fetter, "The Physics of Liquid and Solid Helium," ed. K. H. Bennemann and J. B. Ketterson, Wiley, New York (1975).
3. E. P. Gross, "Quantum Fluids," ed. D. F. Brewer, North Holland, Amsterdam (1966).
4. A. L. Fetter and J. D. Walacka, "Quantum Theory of Many Particle Systems," McGraw Hill, New York (1971).
5. E. Madelung, Z. für Phys. **40**, 322 (1927).
6. J. Grant, Ph.D. Thesis, University of Newcastle (1972).
7. J. Grant, J. Phys. A, **6**, L151 (1973).
8. S. Putterman, "Superfluid Hydrodynamics," North Holland, Amsterdam (1974).
9. L. M. Hocking and K. Stewartson, Mathematika **18**, 219 (1971).
10. R. C. Clark, Phys. Lett. **16**, 42 (1965).
11. R. C. Clark, "Superfluid Helium," ed. J. F. Allen, North Holland, Amsterdam (1966).
12. J. Grant and P. H. Roberts, J. Phys. A., **7**, 260 (1974).
13. V. Celli, M. H. Cohen and M. J. Zuckerman, Phys. Rev. **173**, 253 (1968).
14. G. Baym, R. E. Barrera and C. J. Pethick, Phys. Rev. Lett. **22**, 20 (1969).
15. N. F. Mott and H. S. W. Massey, "The Theory of Atomic Collisions," Third Edition, Clarendon Press, Oxford (1965).
16. M. Steingart and W. I. Glaberson, Phys. Rev. A., **5**, 985 (1972).
17. W. I. Glaberson, J. Low Temp. Phys. **4**, 289 (1969).
18. K. W. Schwarz and R. W. Stark, Phys. Rev. Lett. **21**, 967 (1968).
19. K. W. Schwarz, Phys. Rev. A., **6**, 1958 (1972).

CALCULATION OF THE STATIC HEALING LENGTH IN HELIUM II

P.H. ROBERTS [1], R.N. HILLS and R.J. DONNELLY
Institute of Theoretical Science and Department of Physics, University of Oregon, Eugene, OR 97403, USA

Received 28 November 1978

The healing length for stagnant liquid helium II at an infinite plane boundary is calculated from the Hills–Roberts equations. A new thermodynamic function needed to complete the calculation is derived from neutron scattering and thermodynamic data. Results for the healing length are obtained from absolute zero to about a tenth of a degree from the lambda point, and at all pressures. Comparison with experimental evidence is presented.

Experiments on the propagation of third sound in thin, unsaturated films of helium II show that the superfluid behaves as though it had a lower areal density than that given by the product of the superfluid density, ρ^s, and the film thickness at the same temperature, T. Rudnick and his collaborators [1,2] as well as other groups, have associated this reduction in density with "healing", the notion that the superfluid density decreases near a boundary. Hills and Roberts [3–5] have formulated a two-fluid theory to describe healing and relaxation which rests on accepted macroscopic balance laws for mass, momentum and energy together with a postulate for entropy growth. This theory is entirely hydrodynamical, valid over the whole temperature range and allows a constitutive dependence on density gradients. Near T_λ, these equations in a sense contain the Ginzburg–Pitaevskii order parameter model [7]. The boundary condition applicable to ρ^s is not known *a priori* but in this letter we explore the consequences of setting ρ^s equal to zero on an infinite plane boundary and compare the results with available data. The healing length we define is analogous to the "layer thickness" of boundary layer theory.

Consider stagnant helium in complete thermodynamic equilibrium, filling the half-space $z > 0$ above a plane boundary $z = 0$. The governing equations are (see eqs. (3.7)–(3.10), ref. [5])

$$A + \rho(\partial A/\partial\rho) = \Phi_0, \quad A = A(\rho, T, \rho^s), \tag{1}$$

$$\rho^2(\partial A/\partial\rho) + (\hbar^2/8m^2\rho^s)\,(\mathrm{d}\rho^s/\mathrm{d}z)^2 = P_0, \tag{2}$$

where A is a free energy function, Φ_0 and P_0 are the Gibbs free energy and pressure at great distances from the wall and m is the ^{4}He mass.

It is difficult to solve eqs. (1) and (2) simultaneously for general $A(\rho, T, \rho^s)$: the superfluid density ρ^s can range independently from zero to the bulk value, f ($f = f(\rho, T)$), at large distances from the wall; the total density ρ may itself vary near the wall. According to Brooks and Donnelly [6], however, A is dominated by the ground state free energy $A_G(\rho)$, which is ~ 15 J/g whereas the excitation part A_E is at most 0.4 J/g. This suggests an expansion in small A_E/A_G whereupon eqs. (1) and (2) give

$$(\hbar^2/8m^2\rho^s)\,(\mathrm{d}\rho^s/\mathrm{d}z)^2 - \rho_0 A_E(\rho_0, T, \rho^s) = -\rho_0 A_E(\rho_0, T, f), \tag{3}$$

where ρ_0 is the value of ρ far from the wall.

We must solve eq. (3) subject to the conditions

$$\rho^s = 0 \quad \text{on} \quad z = 0, \quad \rho^s \to f \quad \text{as} \quad z \to \infty, \tag{4}$$

the latter of which has in essence been incorporated in eq. (3). We obtain

[1] Permanent address: School of Mathematics, University of Newcastle upon Tyne, Newcastle upon Tyne, UK.

$$z = (\hbar^2 f/2\rho m^2)^{1/2} \int_0^R [A(fR^2) - A(f)]^{-1/2}\, \mathrm{d}R, \quad (5)$$

where we have suppressed the suffix on ρ_0, replaced ρ^s by fR^2, and written $A_E(\rho, T, \rho^s)$ as $A(fR^2)$, the other arguments (ρ_0 and T) being constants. We note that for small t, $t = T_\lambda - T$, it should be possible to write to a good approximation

$$A(\rho, T, \rho^s) = A_0(\rho, T) + A_1(\rho, T)[\rho^s - f(\rho, T)]^2/2\rho. \quad (6)$$

In this case eq. (5) yields a familiar solution obtained in the work of Ginzburg–Pitaevskii [7], viz:

$$\rho^s = f \tanh^2(z/D), \quad D = \hbar/m(A_1 f)^{1/2}. \quad (7)$$

There are various possible definitions for the term "healing length". For instance, it can be that value of z for which ρ^s attains 90% of its bulk value, that is $R = \sqrt{0.9}$. We shall use as our measure the "displacement thickness", δ, which is defined to be that distance for which $f\delta$ is the superfluid mass (per unit area of wall) "displaced" from the wall through healing: in other words, the hypothetical density distribution

$$\rho^s = 0, \quad z < \delta,$$
$$= f, \quad z \geqslant \delta, \quad (8)$$

should have the same net superfluid mass as the actual solution. For the density distribution (7) the displacement thickness is D.

For a general energy function A, eq. (5) can be evaluated only by numerical integration and ρ^s will not follow the tanh law (7) except for small t. The displacement thickness is

$$\delta = \int_0^\infty (1 - R^2)\mathrm{d}z = (\hbar^2 f/2\rho m^2)^{1/2} \times \int_0^1 (1 - R^2)[A(fR^2) - A(f)]^{-1/2}\, \mathrm{d}R. \quad (9)$$

No theory of fluid flow can be practically useful until the thermodynamic state functions are available. Neutron scattering and thermodynamic measurements properly analyzed give an excellent account of $A(\rho, T, f)$ but say virtually nothing about $A(\rho, T, \rho^s)$ when $\rho^s \neq f$. It is a principal aim of this letter to see whether one obvious theoretical method of computing $A(\rho, T, \rho^s)$ will yield useful results when used in conjunction with predictions such as eq. (9) of the Hills–Roberts theory.

It is well known that the state functions of the Landau theory depend not only on ρ and T, but also on w^2, where $\boldsymbol{w} = \boldsymbol{v}^n - \boldsymbol{v}^s$ is the relative velocity between components. Also, the normal fluid density $\rho^n (= \rho - \rho^s)$ can be obtained from the Helmholtz free energy $F(\rho, T, w^2)$ by differentiation,

$$\rho^n/2\rho = -\partial F/\partial w^2. \quad (10)$$

For sufficiently small T and density excitations $n(\boldsymbol{p})$, F_E can be accurately obtained by using the classical expression for a non-interacting Bose gas:

$$F_E = (kT/\rho h^3) \int \ln\{1 - \exp[-(\epsilon(p) - \boldsymbol{p}\cdot\boldsymbol{w})/kT]\}\mathrm{d}^3 p, \quad (11)$$

where $\epsilon(p)$ is the energy of a quasiparticle of momentum $\boldsymbol{p}$ in stagnant helium. Donnelly and Roberts [8] have shown that the equilibrium state functions such as $F(\rho, T)$ may be obtained to within a tenth degree of T_λ provided the dependence of $\epsilon(p)$ on T (and, of course, on ρ) is consistently incorporated. When t is sufficiently small or w^2 sufficiently large, the rise of n and the concomitant quasiparticle interactions make eq. (11) suspect.

The energy function A of the Hills–Roberts theory is given by the Legendre transformation

$$A = F_E - w^2 \partial F_E/\partial w^2, \quad (12)$$

and was numerically evaluated using the Brooks–Donnelly tables [6]. However, at low temperatures this scheme suffers a setback. When ρ^n is evaluated for $T \lesssim 0.6$ K, using eqs. (10) and (11), we find that ρ^n achieves its maximum, ρ_L^n, at the Landau velocity, w_L, but $\rho_L^n < \rho$ although $\partial\rho^n/\partial w^2 \to \infty$ as $w \to w_L^-$ (i.e. as $w \to w_L$ from below). The severity of this difficulty is extreme for small T, when eqs. (10) and (11) yield

$$\rho_L^n = p_L^2 (2\pi kT)^{3/2} \zeta(3/2)/(w_L^2 h^3 \mu_L^{1/2}), \quad \mu = \partial^2\epsilon/\partial p^2,$$

where ζ is the Riemann zeta function and w_L and p_L are obtained by solving, in the usual way, $w_L = \epsilon(p_L)/p_L = (\partial\epsilon/\partial p)_L$ for the Landau critical state. Evidently $\rho_L^n \to 0$ as $T \to 0$. To surmount this difficulty we recognize that w_L divides the states of thermal equilibrium, $w < w_L$, from the states $w > w_L$ at

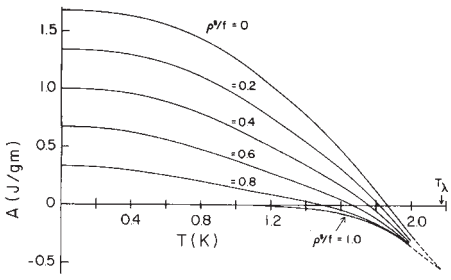


Fig. 1. The free energy function A at density $\rho = 0.1481$ g/cm^3 for six values of ρ^s/f.

which quasiparticle production rates are infinite and equilibrium is impossible. Thus, $w = w_L$ marks a point where any value of ρ^n between ρ_L^n and ∞ is admissable. To evaluate A for $\rho^n > \rho_L^n$ we may simply use the relation $\rho(\partial A/\partial\rho^s) = -w^2/2$ and integrate to get

$$A_E = A_L + w_L^2(\rho_L^s - \rho^s)/2\rho \quad (\rho^s < \rho_L^s). \tag{13}$$

This model allows us to treat the range $T \lesssim 0.6$ K. For $T = 0$ its consequences are particularly elementary. The solution of eqs. (3) and (4), though demanding a careful treatment too lengthy to report here, yields

$$\delta \to h/8mw_L, \quad T \to 0. \tag{14}$$

This remarkably simple result gives $\delta = 2.12$ Å at $T = 0$. Since $w_L \doteq \Delta/p_0$, it demonstrates that δ couples to the reciprocal of the gap (and therefore the structure factor) and not to the interatomic spacing.

Fig. 1 shows A as a function of T and it will be noticed that, for each ρ^s/f, the curve appears to converge on one value A_λ of A at T_λ. The behavior of δ as a function of T for several values of the pressure is shown in fig. 2 [‡1]. We find an increase of $\sim 27\%$ in δ on going from 0 to 25 atm. Steingart and Glaberson [9] have shown that the vortex core parameter

[‡1] It may be noticed that, according to eq. (2), it is the total pressure and not the kinetic pressure that is constant through the healing layer. The fact that we have used tabulations at constant P might therefore seem to be dangerous. It may be shown, however, that the resulting error in δ is only of order $A_E\delta/A_G$, and is no worse than the neglect of $(A_E/A_G)^2$ terms already discarded in the derivation of eq. (3).

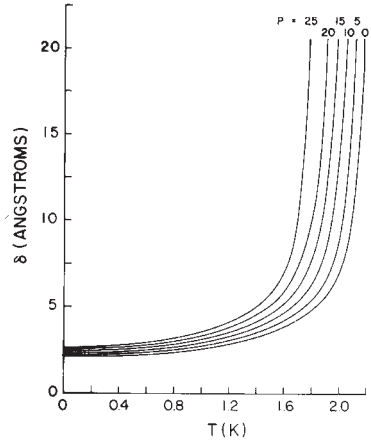


Fig. 2. The displacement thickness δ for six pressures P in atm.

increases by 27 ± 4% over the same pressure range. The vortex core parameter is a dynamic rather than a static healing length. Near $T = 0$, it can be evaluated in a manner analogous to the discussion of eqs. (13) and (14) and yields 0.91 Å or about 0.43δ. In fig. 3 we compare our results for δ with various experiments and the solid curve assumes that healing takes place at both the substrate and the free surface of

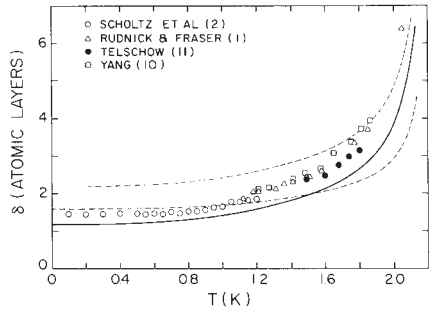


Fig. 3. The total displacement thickness at $P = 0$ compared to the results of various authors. The solid line corresponds to 2δ, i.e. healing on the substrate and the free surface. The lower dashed line corresponds to $1 + \delta$, i.e. healing at the edge of the solid layer and none at the free surface. The upper dashed line corresponds to $1 + 2\delta$, i.e. healing over the solid layer and at the free surface.

a film. The agreement is within 20% of the measurements throughout; our calculations, however, have no adjustable parameters.

There is considerable evidence that something of the order of one atomic layer on the substrata is solid. The dashed curves in fig. 3 show that two healing assumptions treating the solid layer as an infinite potential barrier less consistent without calculations than healing on the substrate and free surface. On the present theory it is the increase in ρ^n which reduces ρ^s. The data suggest that the solid acts in much the same way as excess normal fluid in reducing ρ^s.

We are grateful to Professor I.M. Khalatnikov for discussions of this problem. This research was supported under grants NSF ENG 76-07354, NSF DMR 76-21814 and AFOSR 76-2880.

References

[1] I. Rudnick and J.C. Fraser, J. Low Temp. Phys. 3 (1970) 225.
[2] J.H. Scholtz, E.O. McLean and I. Rudnick, Phys. Rev. Lett. 32 (1974) 147.
[3] R.N. Hills and P.H. Roberts, Int. J. Eng. Sci. 15 (1977) 305.
[4] R.N. Hills and P.H. Roberts, J. Low Temp. Phys. 30 (1978) 709.
[5] R.N. Hills and P.H. Roberts, J. Phys. C, to be published.
[6] J.S. Brooks and R.J. Donnelly, J. Phys. Chem. Ref. Data 6 (1977) 51.
[7] V.L. Ginzburg and L.P. Pitaevskii, Sov. Phys. JETP 7 (1958) 858 [Zh. Eksp. Teor. Fiz. 34 (1958) 1240].
[8] R.J. Donnelly and P.H. Roberts, J. Low Temp. Phys. 27 (1977) 687.
[9] M. Steingart and W.I. Glaberson, J. Low Temp. Phys. 8 (1972) 61.
[10] L.C. Yang, Ph.D. Thesis, Dept. of Phys. UCLA (1973).
[11] K. Telschow, quoted in ref. [2].

Superflow in Restricted Geometries

R. J. Donnelly, R. N. Hills, and P. H. Roberts[a]
Institute of Theoretical Science and Department of Physics, University of Oregon, Eugene, Oregon 97403
(Received 16 October 1978)

We propose a "competing barrier model" for nucleation of quantized vortices in small channels. The results are compared to experiments on decay of persistent currents, critical velocities, onset temperatures, and the effective superfluid density at onset.

For a number of years experimentalists have tried to use the Iordanskiĭ-Langer-Fisher[1,2] (ILF) theory of fluctuation dissipation for superflow in an unbounded region (intrinsic nucleation) to fit experiments in restricted geometries. Several years ago, two of us (DR)[3] pointed out the necessity of considering a permanent barrier ΔE for flow in a restricted geometry which is present, irrespective of any superflow. For a toroidal channel of circular cross section with zero circulation, DR point out that the probability of nucleating to the wall a vortex of one sign is equal to that for a vortex of opposite circulation. However, when there is a superflow present, the energy barriers for nucleation of the two types of vortex are not the same: For a large superflow the process associated with one type of vortex will completely dominate the other, while for slower flows the nucleation of both types of vortex have to be considered. This is the "competing-barrier model" of nucleation.

In this Letter we show that the competing-barrier model will qualitatively explain experiments on decay of persistent currents as reported, for example, by Hallock and co-workers[4] and Kojima *et al.*[5] We confine ourselves to the broad issues using a simplified one-dimensional nucleation model, neglecting the geometrical differences among various sorts of porous materials and thin films.

The nucleation probability, P, per unit time can be written

$$P = f \exp(-\Delta F/kT), \tag{1}$$

where f is the temperature-dependent attempt frequency discussed and tabulated by DR and ΔF is the barrier height. For the ILF model, ΔF is due to the superflow alone. Here, $\Delta F = \Delta E$ when the superflow $v_s = 0$ and the critical momentum is then p_c. With a flow, the $\vec{p}\cdot\vec{v}_s$ interaction shifts the barrier and (considering only one dimension) we have $\Delta F = \Delta E \pm p_c v_s$ so that (1) becomes

$$P = 2f \exp(-\Delta E/kT)\sinh(p_c v_s/kT). \tag{2}$$

Equation (2) shows that the dimensionless quantity $V\ (\equiv p_c v_s/kT)$, the ratio of ordered flow energy to fluctuation energy, is important.

Suppose the superflow takes place in a toroidal geometry containing n candidates per unit length for nucleation; then (cf. Ref. 3)

$$dv_s/dt = -nP\kappa, \tag{3}$$

where $\kappa \equiv h/m$. Equation (3) can be written nondimensionally as

$$dV/d\tau = -\sinh V \tag{4}$$

by introducing the dimensionless time $\tau = \nu_0 t$ and the nucleation rate ν_0:

$$\nu_0 = \nu \exp(-\Delta E/kT),\quad \nu = 2n\kappa f p_c/kT. \tag{5}$$

Equation (4) has the solution

$$V = \ln[(1+e^{-\tau}\tanh V_0/2)/(1-e^{-\tau}\tanh V_0/2)], \tag{6}$$

where V_0 is the value of V at $\tau = 0$. The character of (6) depends on the ranges of V_0 and τ.

For infinitesmal superflows, such as in "Nth sound" ($N = 2, 3, 4$), $V_0 \to 0$ and (6) gives for all τ

$$V \cong \ln[(1 + \tfrac{1}{2}V_0 e^{-\tau})/(1 - \tfrac{1}{2}V_0 e^{-\tau})] \cong V_0 e^{-\tau}. \quad (7)$$

Thus all small superflows decay exponentially.

For $V_0 \to \infty$ and $V \to \infty$ simultaneously, $e^{-V} \cong \exp(-V_0) + \tau/2$ so that

$$V \cong V_0 \text{ (small } \tau), \quad (8a)$$

$$V \cong \ln(2/\tau) \text{ (large } \tau), \quad (8b)$$

and the dividing case occurs at τ_L obtained by equating the two estimates in (8):

$$\tau_L = 2\exp(-V_0). \quad (9)$$

For fixed τ (> 0), V is independent of V_0 in the limit $V_0 \to \infty$:

$$V \cong \ln\{(1 + e^{-\tau})/(1 - e^{-\tau})\} = \ln\coth(\tfrac{1}{2}\tau),$$

so that

$$V \cong \ln(2/\tau) \text{ (small } \tau), \quad (10a)$$

$$V \cong 2e^{-\tau} \text{ (large } \tau), \quad (10b)$$

and the dividing case occurs at τ_E obtained by equating the two estimates in (10):

$$\tau_E \cong 1. \quad (11)$$

The dimensionless times τ_L and τ_E are fundamental to our discussion: τ_E is universal and sets the lifetime of all superflows, while τ_L depends on V_0. The flow observed depends on the magnitude of τ. For $\tau \ll \tau_L$ the flow is almost steady; for $\tau_L \ll \tau \ll \tau_E$ the flow shows logarithmic behavior; for $\tau \gg \tau_E$ the flow decays exponentially. The number of decades of logarithmic behavior is given by $\log(\tau_E/\tau_L) \cong 0.43V_0 - 0.37$.

Figure 1 shows several examples of (6), the time evolution of finite superflows. In particular, the flow for $V_0 = 15$ has $\log\tau_L = -6.2$. Larger initial flows are independent of V_0 at $\log\tau = -6.2$, and and experimenter observing at this τ would term $V_0 = 15$ the "saturated critical velocity." Hence, the condition at time τ for a saturated critical velocity is given by $\tau = \tau_L$ and thus the notion of a saturated critical velocity depends crucially on the time of observation.

An experimenter using Nth sound as a probe for evidence of superfluidity would observe nothing for times larger than τ_E. Thus the condition for "onset of superfluidity" is given by $\tau = \tau_E$.

In order to make numerical calculations, we need estimates of ΔE, n, and f. Imagine a channel containing packed powder with a mean open dimension d. A vortex stretched between two grains will have an energy $E_A = (\rho_s \kappa^2 d/4\pi)\ln(d/a)$, where a is the vortex core parameter. When the line is moved into a semicircle, it will just touch the next grain, and then its energy is $E_c = (\rho_s \kappa^2 d/4)\ln(d/a)$. The "barrier" is given by $\Delta E = E_c - E_A$ and $p_c = \frac{1}{2}\rho_s \kappa \pi d^2$. To order one, we adopt for simplicity

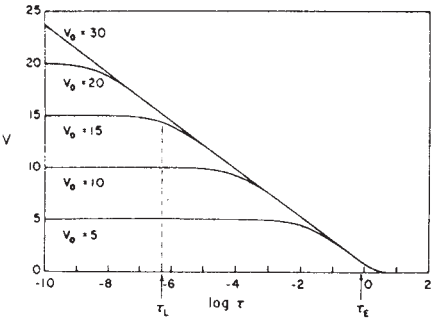


FIG. 1. Plot of Eq. (6) assuming $kT/p_c = 4$ and various values of the initial dimensionless velocity V_0. The dimensionless time showing the beginning of exponential decay τ_E is common to all curves. τ_L is shown for $V_0 = 15$. Initial flows greater than 15 have the same velocity near $\log\tau = -6$; initial flows less than 15 are "steady" for increasing periods as V_0 decreases.

$$\Delta E = (\rho_s \kappa^2 d/4\pi)\ln(d/a), \quad (12a)$$

$$p_c = \rho_s \kappa d^2. \quad (12b)$$

In the same spirit, since the preexponential factors are not important, we assume that n corresponds to one trapped vortex betwen each pair of grains, $n = 1/d$, while a constant value $f = 10^8$ sec^{-1} will suffice. For films we imagine the substrate contains a distribution of trapped vortex lines pinned between protuberances on the substrate. When a trapped line of length of order d moves into semicircle, it will just touch the free surface, forming a vortex of energy E_c as before. Thus the estimates of ΔE, f, and n may be retained for simplicity.

When a saturated current decays according to (8b), the slope of the decay is $dV/d\ln\tau = -1$, or

$$dv_s/d(\log t) = -kT\ln(10)/p_c. \quad (13)$$

Kojima *et al.*[5] observed a decay of 0.63% per decade of a saturated persistent current with $v_s = 67.7$ cm/sec at $t = 1$ sec, $T = 1.3$ K, and $d = 170$–

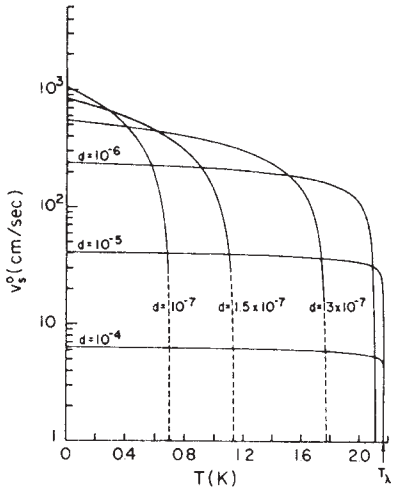


FIG. 2. Saturated velocities obtained from Eq. (14) at $t = 100$ sec for various channel sizes. Flows near onset will show decay as $V_0 \to 1$ (indicated by dashed lines).

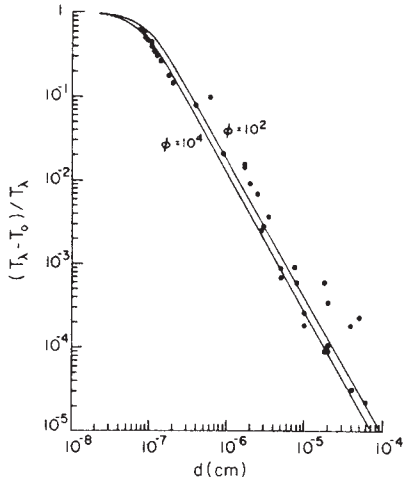


FIG. 3. Onset temperatures T_0 obtained from various channel sizes. Data from many sources.

325 Å. Thus $dv_s/d(\log t) = -6.3\times10^{-3}\times67.7$ or, by (13), $kT/p_c = 0.185$. For $d = 250$ Å we find, using (12b), that $kT/p_c = 0.208$ in good agreement with observation.

The condition that an initial velocity leads to a saturated flow is $\tau = \tau_L$, or $V_0 = \ln(2/\tau)$. Defining v_s^0 as the initial velocity corresponding to V_0, we find from (5) that $\tau = \tau_L$ becomes

$$v_s^0 = \Delta E/p_c - (kT/p_c)\ln(\nu t/2). \tag{14}$$

The first term $\Delta E/p_c = (\kappa/4\pi d)\ln(d/a)$ does not contain temperature explicitly and is a Feynman critical velocity.[6] By itself, it is appropriate only at very low temperatures, since the second term substracts from it, reducing the observed critical velocity. The results shown in Fig. 2 are in order-unity agreement with published results. They show the critical velocity approaching zero near the onset temperature $T_0 < T_\lambda$, to be discussed next.

The condition for the onset of superfluidity is $\tau = \tau_E$. If we express our observation time in terms of frequency $\varphi = t^{-1}$, this condition by (5) is $\nu_0 = \varphi\tau_E$ or

$$[\ln(d/a)/\ln(\nu/\varphi\tau_E)](\rho_s d/T_0) = 4\pi k/\kappa^2, \tag{15}$$

which must be solved by iteration since ν and a are themselves temperature dependent. We show in Fig. 3 the results of T_0 as a function of d compared with the results of many experiments, taking typical values of $\varphi = 10^2$–10^4 Hz. The general agreement is quite satisfactory—below about 0.6 K our results are questionable since the corresponding films are very thin. T_0 increases weakly with φ.

Recently Nelson and Kosterlitz[7] showed that for two-dimensional superfluids the ratio of superfluid mass per unit area near the transition temperature T_0 is given by the exact result,

$$\rho_s(T_0^-)/T_0 = 8\pi k/\kappa^2. \tag{16}$$

Bishop and Reppy[8] and Rudnick[9] have shown that (16) gives a good account of their experiments over a wide range of temperatures and thicknesses identifying $\rho_s(T_0^-)$ with $\bar\rho_s d$. We can write (15) in the form (16) by adopting an effective superfluid density at onset:

$$\bar\rho_s = \rho_s\{2\ln(d/a)/\ln(\nu/\varphi\tau_E)\}. \tag{17}$$

We illustrate the behavior of $\bar\rho_s$ in Fig. 4: It has the usual property of vanishing at $T = 0$ and $T = T_\lambda$. For comparison one can show data quoted by Rudnick derived from third-sound measurements on thin films[9] and by Bishop and Reppy from measurements of $\rho_s(T_0^-)$.[8] Values of $\bar\rho_s$ were deduced

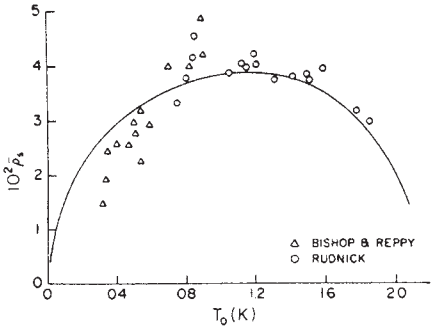


FIG. 4. The effective superfluid density at onset ρ_s calculated from Eq. (17) and compared with experiments of Rudnick (Ref. 9) and Bishop and Reppy (Ref. 8).

assuming the relationship between onset thickness d in atomic layers and T_0 is given by $T_0 = \beta(d - \alpha)$ with $\beta \simeq 1.5$ K per layer and $\alpha \simeq 1.4$ layers for $T_0 < 1$ K.

The works of Nelson and Kosterlitz,[7] Huberman, Myerson, and Doniach,[10] Ambegaokar, Halperin, Nelson, and Siggia,[11] and Myerson[12] address the problem of two-dimensional superfluidity with emphasis on the behavior near the transition. Reference 10 discusses a two-dimensional depairing model and Refs. 10 and 12 give a decay form which fits some of the film data quite well.[4] Reference 11 discusses a vortex depairing and recombination model and provides a detailed discussion of the experiments of Bishop and Reppy.[8] The present theory assumes three dimensions and requires a permanent barrier ΔE. It attempts to include behavior far from critical. Films so thin that the nucleation mechanism discussed above could not operate would have to be treated differently: In particular, the thickness of the film must be considerably greater than the healing length in the present model. We shall address these concerns more fully in a forthcoming article.

Evidence supporting the decay shape shown in Fig. 1 and the dependence of ν_0 and d and T will be presented elsewhere by D. Ekholm and R. B. Hallock, to whom we are indebted for experimental cooperation during the development of this theory. We are also grateful to J. Reppy and I. Rudnick for discussions of their experiments, and to B. Huberman, S. Doniach, E. Siggia, and V. Ambegaokar for the discussions of their theories.

This research was supported by National Science Foundation Grants No. NSF ENG 76-07354 and No. DMR 76-07354 and U. S. Air Force Office of Scientific Research Grant No. 76-2880.

---

[a] Permanent address: School of Mathematics, University of Newcastle upon Tyne, Newcastle upon Tyne, United Kingdom.

[1] S. V. Iordanskiĭ, Zh. Eksp. Teor. Fiz. 48, 708 (1965) [Sov. Phys. JETP 21, 467 (1965)].

[2] J. S. Langer and J. D. Reppy, in *Progress in Low Temperature Physics*, edited by C. J. Gorter (North-Holland, Amsterdam, 1970), Vol. VI, Chap. I.

[3] R. J. Donnelly and P. H. Roberts, Philos. Trans. Roy. Soc. A271, 41 (1970).

[4] K. L. Telschow and R. B. Hallock, Phys. Rev. Lett. 37, 1484 (1976); D. Ekholm and R. B. Hallock, J. Phys. (Paris), Colloq. 39, C6-306 (1978); R. B. Hallock, Bull. Am. Phys. Soc. 23, 536 (1978).

[5] H. Kojima, W. Veith, E. Guyon, and I. Rudnick, in *Low Temperature Physics, LT-13* (Plenum, New York, 1974), Vol. 1, p. 279.

[6] R. P. Feynman, *Progress in Low Temperature Physics*, edited by C. J. Gorter (North-Holland, Amsterdam, 1964), Vol. 1, p. 1.

[7] D. R. Nelson and J. M. Kosterlitz, Phys. Rev. Lett. 39, 1201 (1977).

[8] D. J. Bishop and J. D. Reppy, Phys. Rev. Lett. 40, 1727 (1978).

[9] I. Rudnick, Phys. Rev. Lett. 40, 1454 (1978).

[10] B. A. Huberman, R. J. Myerson, and S. Doniach, Phys. Rev. Lett. 40, 780 (1978).

[11] V. Ambegaokar, B. I. Halperin, D. R. Nelson, and E. D. Siggia, Phys. Rev. Lett. 40, 783 (1978).

[12] R. J. Myerson, Phys. Rev. B 18, 3204 (1978).

NON-LINEAR HYDRODYNAMICS AND A ONE FLUID THEORY OF SUPERFLUID He^4

S.J. PUTTERMAN[1] and P.H. ROBERTS
School of Mathematics, University of Newcastle upon Tyne, Newcastle upon Tyne, UK

Received 3 April 1982

At the fourth order of non-linearity classical hydrodynamics possesses a broken symmetry relating to the interaction of beams of sound waves. The non-linear evolution of these high-order properties of the acoustic field are identical with the two fluid hydrodynamics used to describe superfluid He^4.

At low temperatures the basic phenomenological or macroscopic description of superfluid helium is in terms of the two fluid hydrodynamic theories [1–4]. For ordinary fluids, as well as He^4 above $T_\lambda = 2.17$ K, five independent variables (density, specific entropy and local velocity; $\rho, s, \mathbf{v}$) are sufficient to characterize uniquely the state of flow. However, for the superfluid state an additional velocity field $\mathbf{v}_s$, the superfluid velocity, is required for a complete description even though the He^4 is a one component fluid (with the most symmetric interatomic potential).

A basis goal for a first principles or microscopic theory of He^4 must be to explain the two fluid phenomenon as well as the phonon–roton excitation spectrum. Generally the motion of the superfluid fraction has been regarded as the adiabatic transformation of a macroscopic ground state whereas the normal fluid flow has been viewed as a drift of excitations or quasi-particles. It is essential that there should be no drag between the super and normal components even when dissipative effects are included. The excitations are interpreted either as properties of the N particle hamiltonian [5] or as quantized states of the continuum [1] and the two fluid hydrodynamics results from a merging of these two independent motions. In another picture the superfluid velocity arises from the appearance of an off-diagonal long range order in the reduced density matrix [6].

[1] Permanent address: Physics Department, UCLA, Los Angeles, CA 90024, USA.

In contrast to these pictures we will show that the two fluid theory is a property of the classical hydrodynamics of a one component fluid. So starting from the equation of conservation of mass and Newton's law for a continuum:

$$\partial\rho/\partial t + \nabla\cdot\rho\mathbf{v} = 0, \tag{1}$$

$$\rho\,[\partial\mathbf{v}/\partial t + (\mathbf{v}\cdot\nabla)\mathbf{v}] = -\nabla p(\rho), \tag{2}$$

where p is the pressure, the two fluid hydrodynamics will be derived at the fourth order of non-linearity. Eqs. (1), (2) possess slowly varying solutions (which will be denoted by subscript "0" and which describe the background fluid) as well as quickly varying solutions which are sound waves of short wavelength. In addition to the interaction between the high frequency waves and the background the waves are continuously redistributing themselves in wave number (or k) space due to non-linear couplings. We will show that the high frequency waves obey the equation of motion of the normal fluid component and that the background obeys the equation of motion of $\mathbf{v}_s$.

As the distinction between normal and super components is based on the length scale of variation of ρ, $\mathbf{v}$ the problem is formulated in terms of geometric acoustics. The normal fluid motion will be the geometric acoustics of the high frequency solutions to (1), (2). Thus we expand the independent variables in the form

$$\rho = \rho_0 + \sum_{n=1}^{\infty} \int \rho_n(\boldsymbol{k},\boldsymbol{r},t)\exp[\mathrm{i}\theta(\boldsymbol{k},\boldsymbol{r},t)]\,\mathrm{d}^3k + \text{c.c.}, \tag{3}$$

0 031-9163/82/0000–0000/$02.75

where c.c. denotes the complex conjugate and we expand in a small parameter ϵ where ρ_n is of order ϵ^n but $\nabla\rho_n$ and $\partial\rho_n/\partial t$ are of order ϵ^{n+2}; θ satisfies the differential form $d\theta = \boldsymbol{k}\cdot d\boldsymbol{r} - \omega\, dt$ where $\boldsymbol{k}, \omega$ are zeroth order in ϵ but their derivatives are second order in ϵ. For simplicity of presentation, the expansion of ρ_0 in powers of ϵ is not explicitly indicated.

At second order in ϵ the contribution of a given wave to the acoustic energy density per volume d^3k of $\boldsymbol{k}$ space is

$$\epsilon_k = (c_0^2/\rho_0)|\rho_1|^2 + \rho_0|v_1|^2, \tag{4}$$

where $c_0^2(\boldsymbol{r}, t) = (\partial p/\partial\rho)_0$. It can be proved from (1), (2) that to order ϵ^4 the equation governing the wave action $n(\boldsymbol{k}, \boldsymbol{r}, t) \equiv \epsilon_k/\omega_0$ is

$$\left(\frac{\partial n}{\partial t}\right)_{k,r} + \left(\frac{\partial\omega}{\partial k_i}\right)_r \left(\frac{\partial n}{\partial r_i}\right)_k - \left(\frac{\partial\omega}{\partial r_i}\right)_k \left(\frac{\partial n}{\partial k_i}\right)_r = I(n), \tag{5}$$

where $\omega = \omega_0(k) + \boldsymbol{v}_0\cdot\boldsymbol{k}$ with $\omega_0 = c_0 k$. The left hand side of eq. (5) [7] describes the changes in the high frequency waves due to interactions with the background [geometric acoustics]. It is similar to the quantum quasi-particle Boltzmann equation used by Khalatnikov [2] for He^4, expect that here it is derived from the classical non-linear hydrodynamics i.e., from the same *macroscopic* equation which describes the background. The right hand side of (5) represents redistribution of sound wave energy due to non-linear collisions between the waves [8,9], which must be generalized to the anisotropic case considered here. In the absence of collisions n is conserved as should be expected since it is the adiabatic invariant of the motion for a sound field.

Before indicating how the normal fluid equation is to be derived from (5) we calculate how the sound field modifies the background flow. Placing the expansion (3) into (1), (2) and keeping terms as high as fourth order yields

$$\partial\rho_0/\partial t + \nabla\cdot\left[\rho_0\boldsymbol{v}_0 + \left(\int \rho_1\boldsymbol{v}_1^*\, d^3k + \text{c.c.}\right)\right] = 0, \tag{6}$$

$$\partial\boldsymbol{v}_0/\partial t + (\boldsymbol{v}_0\cdot\nabla)\boldsymbol{v}_0 = -\nabla[\Phi_0 + (U_n/c_0)\, dc_0/d\rho_0], \tag{7}$$

where

$$U_n = \int nc_0 k\, d^3k = \int \epsilon_k\, d^3k, \quad \Phi_0 = \int^{\rho_0} dp/\rho.$$

Eqs. (5), (6), (7) describe how the mixture of slow and fast solutions evolve in time. They reduce to the two fluid hydrodynamics when supplemented with a local equilibrium approximation. The collision integral has the property that $\boldsymbol{k}$, $|k|$ are additive conserved galilean invariant properties in an interaction of sound waves. Thus the stationary solution to $I(n) = 0$ will have the functional form $n = n_{eq}(\beta(c_0 k - \boldsymbol{k}\cdot\boldsymbol{w}))$ where $\boldsymbol{w}$ is a new galilean invariant vector and represents the broken symmetry. Next assume that off-equilibrium n is the same functional form but with β, c_0, $\boldsymbol{w}$ slowly varying functions of position and time. Multiplying (5) by $\boldsymbol{k}$ and summing over all $\boldsymbol{k}$ (all the high frequency waves) then yields:

$$\frac{\partial\rho_n w_i}{\partial t} + \frac{\partial}{\partial r_j}[\rho_n w_i(\boldsymbol{w} + \boldsymbol{v}_0)_j] + \rho_n w_j \frac{\partial}{\partial r_j}(\boldsymbol{w} + \boldsymbol{v}_0)_j + \rho s\nabla T = 0, \tag{8}$$

where we have set $\beta = 1/k_B T$,

$$\int n\boldsymbol{k}\, d^3k = \rho_n \boldsymbol{w}, \tag{9}$$

$$\rho s = \int S(n)\, d^3k, \tag{10}$$

with

$$\partial S(n)/\partial n = k_B\beta(c_0 k - \boldsymbol{w}\cdot\boldsymbol{k}). \tag{11}$$

From (10) and (11), (5) yields

$$\partial\rho s/\partial t + \nabla\cdot\rho s(\boldsymbol{w} + \boldsymbol{v}_0) = 0. \tag{12}$$

Eqs. (6) and (7) for the background flow become upon summing over $\boldsymbol{k}$:

$$\partial\rho_0/\partial t + \nabla\cdot(\rho_0\boldsymbol{v}_0 + \rho_n\boldsymbol{w}) = 0, \tag{13}$$

$$\partial\boldsymbol{v}_0/\partial t + (\boldsymbol{v}_0\cdot\nabla)\boldsymbol{v}_0 = -\nabla\mu, \tag{14}$$

where

$$\mu = \Phi_0 + (U_n/c_0)\, dc_0/d\rho_0. \tag{15}$$

That μ is the chemical potential is verified by considering the total energy density which to second order in ϵ is:

$$U = U_0(\rho_0) + \tfrac{1}{2}\rho_0 v_0^2 + U_n + \boldsymbol{v}_0\cdot\int(\rho_1^*\boldsymbol{v}_1 + \text{c.c.})\, d^3k,$$

and includes the background energy, the acoustic energy and the interaction energy between them. From

(9), (10), (11) one obtains

$$dU = \mu\, d\rho + T\, d\rho s + \boldsymbol{w}\cdot d\rho_n \boldsymbol{w} + d(\tfrac{1}{2}\rho_0 \boldsymbol{v}_0^2 + \boldsymbol{v}_0 \cdot \rho_n \boldsymbol{w}), \tag{16}$$

where μ is the same as in (15). The replacements

$$\rho_0, \boldsymbol{v}_0, \boldsymbol{w} \to \rho, \boldsymbol{v}_s, \boldsymbol{v}_n - \boldsymbol{v}_s$$

put equations (8), (12), (13), (14), (16) into a form identical with the two fluid thermohydrodynamics. In general $\int \boldsymbol{I}\boldsymbol{k}\, d^3k = 0$ so that an equation always exists for $\rho_n \boldsymbol{w}$. However, this equation will be closed in terms of the thermodynamic variables [i.e., $\rho s \nabla T$] only when the system of waves is sufficiently close to equilibrium that the local equilibrium form for n is valid. At low temperatures when the mean free path of sound waves becomes longer than the wavelength, local equilibrium will not be a valid approximation. Although (8) will not be correct in this case the proper description will still be given by (5), (6), (7). Eq. (5) is amenable to a Chapman–Enskog expansion around the local equilibrium state. This procedure will yield values for the off-equilibrium kinetic coefficients and one also finds that even in the presence of entropy production there is no friction between super and normal components. The important quantity $n\boldsymbol{k}$ often referred to as the Stokes drift of the sound field is the normal fluid momentum.

There remains the question of the precise functional form of n and whether this approach can incorporate dispersive systems. The above analysis [including the collision integral $I(n)$] has been extended to systems where the internal energy is an arbitrary combined function of the gradients of the density and a non-local potential. In particular consider a hydrodynamic energy density which may model He^4:

$$U(\rho) = \tfrac{1}{2}\rho v^2 + (\gamma/\rho)(\nabla\rho)^2 + \frac{1}{2m^2}\int \rho(\boldsymbol{r}) V(\boldsymbol{r}-\boldsymbol{r}')\rho(\boldsymbol{r}')\, d\boldsymbol{r}', \tag{17}$$

where γ is a constant, m the mass of a He^4 atom and $V(\boldsymbol{r}-\boldsymbol{r}')$ the two-particle potential. Such a mean field approach might be valid near $T = 0$ K since the exact N particle ground state wave function has no nodes even in the presence of sharp repulsive forces. The dispersion law implied by (17) is [10]:

$$\omega_0^2(k) = 2\gamma k^4 + (2\pi)^3 \rho_0 V(k) k^2/m^2, \tag{18}$$

where $V(\boldsymbol{r}) = \int \exp(i\boldsymbol{k}\cdot\boldsymbol{r}) V(\boldsymbol{k})\, d^3k$. Using $\gamma = \hbar^2/8m^2$ gives (17) the form of the Gross–Pitaevskii equation [11] and taking $V(r)$ as a hard sphere interaction $V(r) = V_0\delta(r-a)/4\pi a^2$ yields a phonon–roton like spectrum [12]. Eq. (5), including the symmetry properties of $I(n)$ remain valid, though (18) must be used for $\omega_0(k)$. Viewed in this fashion the continuum mechanics will allow for a calculation of the interaction of phonons and "rotons" and might provide a route to a microscopic theory of He^4.

Finally we consider the collision integral $I(n)$ which is needed to determine the local equilibrium form for n. It will contain direct collisions which remove energy from a given channel $\boldsymbol{k}$ as well as restituting terms. Along the lines of ref. [9] we will also include the possibility of a scattering from zero temperature noise which is renormalized so that the non-linearities cannot be used to remove energy from this noise. The anisotropic version of the collision term then is [from (1), (2)]

$$\begin{aligned} I(n) = a_0 \int k k_1 k_2 \{ & \delta(\omega - \omega_1 - \omega_2)\delta(\boldsymbol{k} - \boldsymbol{k}_1 - \boldsymbol{k}_2) \\ & \times [n_1 n_2 - n(n_2 + \tfrac{1}{2}\alpha) - n(n_1 + \tfrac{1}{2}\alpha)] \\ & + 2\delta(\omega + \omega_1 - \omega_2)\delta(\boldsymbol{k} + \boldsymbol{k}_1 - \boldsymbol{k}_2) \\ & \times [(n_1 + \tfrac{1}{2}\alpha)n_2 + (n + \tfrac{1}{2}\alpha)n_2 - n n_1]\}\, d^3k_1\, d^3k_2, \end{aligned} \tag{19}$$

where $n_i \equiv n(k_i)$, $\omega_i = \omega(k_i)$, $a_0 = (\pi c/2\rho)[1 + (\rho/c)\, dc/d\rho]^2$. In the framework of this approach α is an otherwise undetermined constant that characterizes the renormalized scattering from the zero temperature noise. Experimentally its value is determined as $\hbar$. A stationary solution to eq. (19) is [+1]

$$n = \alpha/\{\exp[\beta\alpha(\omega_0 - \boldsymbol{w}\cdot\boldsymbol{k})] - 1\}. \tag{20}$$

Substitution of (20) into (15), (9), (11) determines the equations of state.

In the limit $\alpha \to 0$ (20) becomes

$$n = k_B T/[\omega_0 - \boldsymbol{w}\cdot\boldsymbol{k}], \tag{21}$$

[+1] For the case where ω_0 is linear in k the interactions are not sufficient to create an isotropic equilibrium distribution as all rays interact in colinear fashion. In a forthcoming paper we show how higher order processes or dispersion bring about the complete equilibrium state.

which leads to an infinite value for U_n and other quantities unless a cut-off is imposed. The Planck distribution (20) does not present this problem. In principle eq. (8) could apply when (21) gives the dominant contribution to n. However in that limit it is hard to imagine a system for which the obvious requirement $\rho_n < \rho$ is satisfied.

We wish to thank Dr. C.A. Jones for bringing the results of refs. [10,12] to our attention and Dr. A.M. Soward for valuable discussions. We wish to thank UK Science and Engineering Research Council for support under Grant GR/B64215. One of us (S.P.) also acknowledges support of the ONR.

References

[1] L.D. Landau, J. Phys. (Moscow) 71 (1941).
[2] I.M. Khalatnikov, An introduction to the theory of superfluidity (Benjamin, New York, 1965).
[3] R.J. Donnelly, Experimental superfluidity (University of Chicago Press, Chicago, 1967).
[4] S.J. Putterman, Superfluid hydrodynamics (North-Holland, Amsterdam, 1974).
[5] A. Bijl, Physica 7 (1940) 869;
R.P. Feynman, in: Progress in low temperature physics, Vol. 1, ed. C.J. Gorter (North-Holland, Amsterdam, 1955) p. 36.
[6] O. Penrose, Philos. Mag. 42 (1951) 1373;
C.N. Yang, Rev. Mod. Phys. 34 (1962) 694.
[7] F.P. Bretherton and C.J.R. Garrett, Proc. R. Soc. A302 (1969) 529;
A.M. Soward, J. Fluid Mech. 69 (1975) 145.
[8] A.C. Newell and P.J. Aucoin, J. Fluid Mech. 49 (1971) 593;
D.J. Benney and P.G. Saffman, Proc. R. Soc. A289 (1966) 301;
P.J. Westervelt, J. Acoust. Soc. Am. 59 (1976) 760.
[9] M. Gabot and S. Putterman, Phys. Lett. 83A (1981) 91.
[10] N.N. Bogoliubov and D.N. Zubarev, Sov. Phys. JETP 1 (1955) 83.
[11] V.L. Ginzburg and L.P. Pitaevskii, Sov. Phys. JETP 34 (1958) 858;
E.P. Gross, Nuovo Cimento 20 (1961) 454; J. Math. Phys. 4 (1963) 195.
[12] K.A. Brueckner and K. Sawada, Phys. Rev. 106 (1957) 1117, 1128.

WHAT HAPPENS TO VORTEX RINGS THAT SHRINK?

C.A. Jones and P.H. Roberts

School of Mathematics, The University, Newcastle upon Tyne, UK

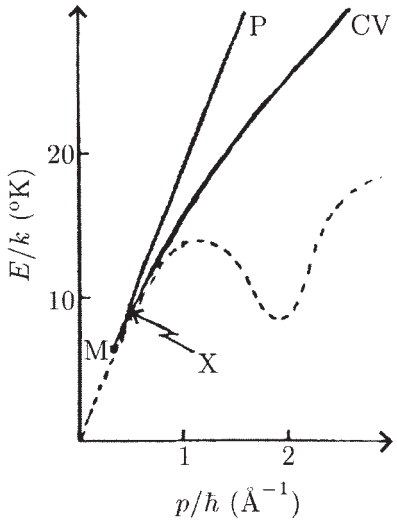
It has been known for more than two decades that a moving ion can generate a disturbance in a superfluid, strongly resembling a classical vortex ring, which attaches itself to the ion (Rayfield and Reif, 1964). Such vortex rings are crucial for theoretical understanding of nucleation processes and the decay of persistent currents (Iordanskiĭ, 1965; Langer and Fisher, 1967, Donnelly and Roberts, 1971). Take a reference frame fixed to the superfluid. If the radius, R, of the vortex ring is large enough, its vorticity is less than that of the normal fluid and, if parallel, the quasiparticles will predominantly hit the ring "from behind", so providing the energy and momentum necessary for its further expansion. Ultimately, the ring expands to the walls of the container, so reducing the counter-flow velocity. Conversely, "head on" collisions tend to remove energy and momentum from the ring, causing R to decrease. In thermal equilibrium, a statistical distribution of ring sizes and orientations is set up, in which the number of nucleating rings is tiny. For the theory to succeed it is essential that the population from which they are drawn is extremely large, comparable with the roton density. But what are these candidates for nucleation? As the scale R of the ring diminishes, quantum effects become increasingly important, and the disturbance progressively departs from the classical vortex ring. Donnelly and Roberts (1971), reviving the old picture of Onsager and Feynman that the roton is "the ghost of a vanishing vortex ring" (see Donnelly, 1974), suggested the rotons were the population from which nucleating rings are created.

To examine this question further, we took a simple model of helium at absolute zero, namely the Bose condensate with $\delta-$function atomic repulsion. We obtained, by numerical integration, the surprising results sketched below. Seeking an axisymmetric solitary wave that preserved its form as it moved with velocity v, we obtained, in the $v \to 0$ limit, classical vortex rings (CV in figure). On increasing v, the energy E and (pseudo-)momentum p diminished as did the scale of the disturbance, the cores at opposite ends of a diameter tending to merge, so creating a region of depleted density surrounding the entire ring. Further increase of v led to loss of vorticity (X on figure), but the sequence did not end there! As v increased a cusp was reached (M) where E and p reached simultaneous minima. Thereafter E and p increased as did the scale and magnitude of the disturbance. Ultimately, as v approached the speed of sound c, the disturbance extended, as a rarefaction pulse (P on diagram), to $\mathrm{O}(a/\epsilon)$ distances along the symmetry axis and to $\mathrm{O}(a/\epsilon^2)$ distances transverse to it, where a is the healing length and $\epsilon^2 = 1 - v/c$. The fluid velocity associated with the wave was $\mathrm{O}(\epsilon^2 c)$ and the disturbed density was $\mathrm{O}(\epsilon^2 \rho_0)$, where ρ_0 is the equilibrium density. Thus $E \sim cp = \mathrm{O}(\rho_0 c^2 a^3/\epsilon)$, as $\epsilon \to 0$. More exactly

$$E \sim cp - K^2 \frac{\rho_0^2 c^3 a^6}{p} + \mathrm{O}\left(\frac{1}{p^3}\right), \qquad p \to \infty, \tag{1}$$

where K is a positive constant. Since $v = \partial E/\partial p$, we have $\epsilon = K\rho_0 c a^3/p$. This

rarefaction pulse, which is the 3-dimensonal analogue of the Tsuzuki (1971) soliton, has also been discussed by Iordanskiĭ and Smirnov (1978), who obtained the same asymptotic theory for it. They make no attempt, however, to connect with the lower (vortex) branch.



Whether the branches remain connected for a more a realistic model of helium is an open question. One can hardly doubt that the minimum, M, would be moved. Assuming $c = 238$ m s^{-1} and $\rho_0 = 0.145$ g cm^{-3}, and choosing a to give the quantum of circulation, κ, correctly, we obtained

$$E_M/k = 6.4\text{K}, \qquad p_M/\hbar = 0.34\text{Å}^{-1}. \tag{2}$$

We have also looked at the analogous problem for a vortex pair. In this case the energy and (pseudo-)momentum per unit length tend to zero as $v \to c$.

REFERENCES

Donnelly, R.J., 1974, in *Quantum Statitical Mechanics in the Natural Sciences*, **4**, ed S.L. Mintz and S.M. Widmayer (New York: Plenum), 359.
Donnelly, R.J. and Roberts, P.H., 1971, *Phil. Trans. R. Soc. Lond. A*, **271**, 41.
Iordanskiĭ, S.V., 1965, *Sov. Phys. JETP*, **21**, 467.
Iordanskiĭ, S.V. and Smirnov, A.V., 1978, *JETP Lett.*, **27**, 535.
Jones C.A. and Roberts, P.H., 1982, *J. Phys. A*, **15**, 2599.
Langer, J.S. and Fisher, M.E., 1967, *Phys. Rev. Lett.* **19**, 560.
Rayfield, G.W. and Reif, F., 1964, *Phys. Rev.* **136***A*, 1194.
Tsuzuki, T., 1971, J. Low Temp. Phys. **4**, 441.

Part 4

MIXED PHASE REGIONS

SAACM - Vol. 1, n. 1, 1991

On the life cycle of a mush

R.N. HILLS

Mathematics Department
Heriot-Watt University
Edinburgh U.K.

P.H. ROBERTS

Department of Mathematics
University of California
Los Angeles, CA 90024, U.S.A.

with an appendix by

A.M. SOWARD

Department of Applied Mathematics
Newcastle University
U.K.

Abstract. *A layer of stagnant water and ice containing sources of heat is thermally insulated at one end, and is maintained at a subzero temperature at the other. The evolution of an arbitrary initial state to the final equilibrium state, in which one side of the layer is water (or in some circumstances ice) and the other side is (necessarily) ice, is studied in the limit of large Stefan number. It is shown that, depending on the degree of heating and the initial state of the system, the system may pass through a transient phase in which an internal region is filled with mush, i.e. is a mixture of both ice and water. The manner in which this slushy region is born, evolves and eventually dies is studied both numerically and analytically.*

1. INTRODUCTION

It is well known that melting/freezing processes in alloys and aqueous solutions will sometimes produce an interface between the solid and liquid phases which is so thin that it is usually modelled as a sharp transition surface. Seemingly equally possible is a state in which the two phases may be separated by a finite region of mixed phase, often termed a mush or mushy zone. Which

of these two situations actually arises in practice depends upon material parameter ranges and delicate thermal balances. For pure substances such as an ice-water system the sharp interface is much more prevalent, so much so that there is perhaps the feeling that it is the only possibility. We shall show, however, that sharp interfaces are not inevitable in this case either.

To illustrate we consider in this paper the melting evolution of an ice-water system contained in a one-dimensional finite region which is being warmed by volumetric heating. For definiteness, we assume that one boundary of the system is maintained at constant temperature while the other is thermally isolated. This is, of course, a Stefan problem and the case when the associated Stefan number is large was previously considered by Ockendon (1975). His initial state was one in which the system was originally all ice. Here we shall include more general initial data.

The analysis of this paper confirms that, for a certain range of positions of the initial phase interface, a sharp surface will persist throughout the whole evolution of the two-phase system: on one side of the interface the ice temperature is sub-zero (on the Centigrade scale) while on the other, the liquid has a positive temperature distribution. More interesting is the range of positions of the initial interface for which this does not happen for all time. For these cases it is possible to find a mathematical solution in which, after a certain onset-time, the temperature of the ice rises above the melting temperature without any release of latent heat (see Ockendon (1975), Lacey and Shillor (1983)). It is sometimes said that the ice has taken up a superheated state that is in quasi-equilibrium (i.e. metastable).

Statistical mechanics offers a compelling explanation for the existence of quasi-equilibrium states such as supercooled water, and provides methods (in terms of the "escape over a barrier" from the quasi-equilibrium state to the nearby state of complete thermodynamic equilibrium) by which the lifetime of a quasi-equilibrium state can be estimated. Supercooled water occurs because the surface energy required to nucleate a grain of ice increases with the radius, R, of the grain. Only if R exceeds some critical radius, R_c, will the grain increase in size and solidification occur; otherwise the grain will dissolve back into the fluid to be replaced by other grains that attempt to escape back over the barrier in turn. The top of the barrier itself is defined by the energy required to produce a single grain of radius R_c, and this is so large that escape over the barrier is a statistical rarity, thereby explaining the longevity of the supercooled state.

There is a complete theoretical symmetry between nucleation of a small ice grain in supercooled water and the nucleation of a small water cavity in superheated ice, both being quasi-equilibrium states that owe their existences to interfacial energy between phases. The practical difficulties in creating super-

heated ice appear to be the greater (see, for instance, Landau and Lifshitz (1969), §150), although there have been reports that such states have been observed even existing long enough to be experimentally studied (Chalmers (1964); Woodruff (1980)). In the context of the present one-dimensional problem, neither supercooled water nor superheated ice are possible states, since both rely on nonzero surface energies between phases, and these are conspicuously absent when those interfaces are planar!

Thus, instead of the appearance of superheated ice we show that a more physically acceptable solution has a transitory region of mixed phase. An earlier treatment of a mushy zone has been presented by Atthey (1974). Here we are particularly concerned with the manner in which the mush is born and the fashion of its eventual demise. There can be no question that it persists for all time: the continual warming by volumetric heating would systematically deplete the mass fraction of the solid phase in the mush in contradiction to the assumption that the solution was steady. Thus the system ultimately moves to a configuration with a sharp interface in which the heat being supplied by volumetric sources is in balance with the heat flux out of the non-insulated end. Governing the evolution to this final state there are two intrinsic time scales, one associated with latent heat and the other with thermal diffusion. The assumption of large latent heat (and hence large Stefan number) means that its associated time scale is slow, the diffusion time being short.

For the cases when a mush does occur its life cycle is intimately bound up with these characteristic rates. There are two main scenarios. For definiteness let us assume that ice lies to the left (smaller x). The mush is always formed by an ice/mush interface appearing and moving left on the fast time scale towards a critical position that is a function of the rate of volumetric heating. As this edge moves across it effectively lays down a distribution of solid and liquid phase in the mush behind it. The morphology of the mush continually changes: the volumetric heating works to diminish the proportion of solid present. Although the leading edge initially moves quickly, it actually approaches the critical position with increasing lethargy. Ultimately the mush vanishes when the trailing (mush/liquid) edge overhauls the leading edge in the neighbourhood of the critical position. The essential difference between the scenarios lies in the way in which the trailing edge follows across. When the initial configuration is an ice-water system, the trailing edge immediately follows across but only on the *slow* time scale. Since the leading adge approaches the critical position moving slower and slower, there will come a time when it is moving as slow as the trailing edge. Thereafter the mush/liquid interface is actually the faster moving and so the width of the mush must decrease and eventually vanish. The other type of mush evolution occurs when the initial phase configuration has only ice present.

In this case the Stefan condition that applies at phase interfaces at first dictates that the trailing edge does not move at all. It has to remain motionless until all the solid at the right hand end of the mush has melted. Since the solid in the mush is eroded at a constant rate and the extreme right hand end is born first, it must be obliterated first due to the heating. Once this happens the trailing edge becomes free to move but, in contrast to the case above, it does so on the *fast* time scale. This again follows from the Stefan condition and reflects the fact that, in the period of time required to melt all the solid at the stationary trailing edge, the solid fraction becomes very small elsewhere (of the order of the reciprocal of the Stefan number) and an $O(1)$ velocity for the mush/liquid interface is needed in order to satisfy the Stefan condition.

The mathematical definition of the problem discussed above is stated in Section 2. In particular we show why a mush is born and in the following two sections discuss aspects of the mush evolution for the first of the cases above. Section 5 is concerned with the evolution of a system that is initially all ice and so is pertinent to the second scenario.

2. OCKENDON'S PROBLEM

The domain $0 \leqslant x \leqslant 1$ (dimensionless units) is filled with H_2O either in its liquid phase (water) or its solid phase (ice) and the pressure is everywhere maintained constant. Uniformly distributed heat sources would cause the temperature, $u(x, t)$, to rise uniformly everywhere at the rate r (> 0), were is not for heat conduction which removes heat from the plane $x = 0$ which is maintained at temperature $u = -1$, the zero of u being the melting temperature; the plane $x = 1$ is a thermal insulator. The thermal diffusivity is everywhere 1 and the Stefan number is denoted by $\lambda = L/c_p T_m$ where L is the latent heat, c_p the specific heat and T_m a reference temperature on the absolute scale. The difference in the densities of ice and water is ignored so that questions of convection and of volume preservation do not arise. Given the initial u and the initial distribution of the phases, it is required to determine these at all subsequent times.

In mathematical terms, this one dimensional Stefan problem may, in the first instance, be stated as follows. Solve

$$u_t = u_{xx} + r, \tag{2.1}$$

subject to

$$u(0, t) = -1, \quad u_x(1, t) = 0, \tag{2.2,3}$$

and at any zero, $x = s(t)$, of u,

$$\lambda \dot{s} = [u_x]^{\text{solid}}_{\text{liquid}} \tag{2.4}$$

Here a superposed dot denotes time differentiation, $u_t = \partial u/\partial t$, $u_x = \partial u/\partial x$ and $[u_x]$ denotes the discontinuity in u_x in the limit $x \to s$ from the two sides. Implicit in this first approach is the assumption that u vanishes only at isolated x, and not over intervals of x, in $[0, 1]$. This is assumed also to be true of the initial distribution,

$$u_i(x) \equiv u(x, 0), \tag{2.5}$$

which is otherwise arbitrary.

Depending on the choice of u_i, there may be many distinct regions of ice ($u < 0$) and water ($u > 0$) in $[0, 1]$, but in the final steady state defined by (2.1) - (2.3), there is at most one region of each. We have

$$u_\infty(x) = u(x, \infty) \equiv -1 + rx - \frac{1}{2} rx^2. \tag{2.6}$$

Thus, if $r < 2$ the material is ultimately solid everywhere. We shall confine attention to the case $r \geqslant 2$, and will write (2.6) as

$$u_\infty = -\frac{1}{2} r(x - s_\infty)(x - \bar{s}_\infty), \tag{2.7}$$

where $s_\infty \leqslant 1, \bar{s}_\infty \geqslant 1$ and

$$\left.\begin{matrix} s_\infty \\ \bar{s}_\infty \end{matrix}\right\} = 1 \mp \left(1 - \frac{2}{r}\right)^{1/2}. \tag{2.8}$$

We see from (2.7) that ice ultimately fills $0 < x < s_\infty$ and water fills $s_\infty < x < 1$, and that $s_\infty = s(\infty)$.

To avoid the complications of following the evolution and demise of many regions of ice and water, we shall concentrate on simple initial states that have at most one of each. Typical of such an initial state is

$$u_i(x) = -1 + 2\alpha x - \alpha x^2. \tag{2.9}$$

If $\alpha < 1$, ice is initially present everywhere. If $\alpha \geqslant 1$, we may write (2.9) as

$$u_i = -\alpha(x - s_i)(x - \bar{s}_i), \tag{2.10}$$

where $s_i \leqslant 1, \bar{s}_i \geqslant 1$ and

$$\left.\begin{matrix} s_i \\ \bar{s}_i \end{matrix}\right\} = 1 \mp \left(1 - \frac{1}{\alpha}\right)^{1/2}. \tag{2.11}$$

Then ice initially fills $0 < x < s_i$ and water fills $s_i < x < 1$, where $s_i = s(0)$.

We should emphasize that the picture that emerges from (2.9) is robust, i.e. other simple choices of u_i, that have one or no zeros, evolve in a way similar to (2.9). Ockendon (1975) focussed on an initial state for which $u \leqslant 0$ and $s_i = 1$; this can be simulated by taking $\alpha = 1$ in (2.9).

One other 'critical' value of x will be significant in what follows, namely $x = s_c$, where

$$s_c = \sqrt{2/r}. \tag{2.12}$$

In the cases $r \geqslant 2$ of greatest interest, s_c falls within $[0, 1]$. It is clear that $s_c > s_\infty$.

Suppose that, for all t during the evolution of u, from (2.9) to (2.6), there exists a single interface $x = s(t)$ between ice in $[0, s)$ and water in $(s, 1]$. Denote u in these intervals by u_1 and u_2 respectively and write (2.4) as

$$\lambda \dot{s}(t) = u_{1x}(s, t) - u_{2x}(s, t) \tag{2.13}$$

Ockendon (1975) noted that Stefan problems of the present type are easily solved in the limit $\lambda \to \infty$. There are then two timescales, the "fast" $O(1)$ timescale of thermal diffusion and the "slow" $O(\lambda)$ timescale of latent heat processes. According to (2.13) the interface moves on the slow timescale, and during this time heat conduction processes are effectively instantaneous, i.e. on the slow timescale, (2.1) gives, to leading order

$$0 = u_{xx} + r, \quad r = 2/s_c^2. \tag{2.14}$$

Let us attempt to solve Ockendon's problem in the limit $\lambda \to \infty$. By (2.2), (2.3) and (2.14) we have, to leading order,

$$u_1 = -(x - s)(x - \tilde{s})/s_c^2, \tag{2.15}$$

$$u_2 = (x - s)(2 - x - s)/s_c^2, \tag{2.16}$$

where $\tilde{s} = s_c^2/s$. It may be noted that, even taking $s = s_i$, (2.15) and (2.16) do not coincide with (2.9). This is because there is a fast initial phase, $O(1)$ in duration, during which (2.9) evolves to (2.15) and (2.16) by heat conduction; this initial phase cannot be monitored by the slow timescale approximation (2.14) of (2.1).

Using (2.15) and (2.16), we see that the interface condition (2.13) is

$$\lambda \dot{s} = (s - s_\infty)(s - \tilde{s}_\infty)/s s_c^2. \tag{2.17}$$

Thus, if $s_i < s_\infty$, then s increases until ultimately $s = s_\infty$; if $s_i > s_\infty$, then $\dot{s} < 0$ and s decreases monotonically to s_∞. Equation (2.17) may be integrated to give

$$(2.18)\quad t = \frac{\lambda s_c^2}{(\bar{s}_\infty - s_\infty)}\left[\bar{s}_\infty \ln\left(\frac{\bar{s}_\infty - s}{\bar{s}_\infty - s_i}\right) - s_\infty \ln\left(\frac{s - s_\infty}{s_i - s_\infty}\right)\right],$$

so that

$$(2.19)\qquad s \sim s_\infty + (s_i - s_\infty)\exp\left[-\frac{(\bar{s}_\infty - s_\infty)}{\lambda s_c^2}\, t\right], \quad \text{as } t \to \infty.$$

This analysis provides a physically uncontentious solution if $s_i \leqslant s_c$. When however $s_i > s_c$, the second zero, $x = \tilde{s}$, of (2.15) falls within the interval $0 \leqslant x < s$ in which (2.15) is supposed to hold. It remains in that interval until s has decreased to s_c; thereafter is exceeds s_c, i.e. is then irrelevant. Thus, during the time in which $s > s_c$, the temperature in $\tilde{s}(t) < x < s(t)$ exceeds zero but, since the latent heat condition (2.4) is not obeyed at $x = \tilde{s}$ (u_x is in fact continuous there), it does *not* reflect a change of phase. It is for this reason that the material in such a region has been called "superheated ice" (Ockendon (1975); Lacey and Shillor (1983); Lacey and Tayler (1983)). For the physical reasons adumbrated in the introduction we prefer to investigate the appearance of a mushy region between the pure phases.

3. THE MUSH: OVERVIEW OF A LIFE CYCLE AND DETAILS OF THE BIRTH

In this and the following section we develop a new solution in which a mushy zone makes a temporary appearance. The initial position of the (sharp) phase interface, $x = s_i$, occurs at an interior point of [0, 1] and this corresponds to $\alpha > 1$ in (2.9). The case $\alpha \leqslant 1$ is considered in detail in Section 5. The morphology of the mixed phase region is characterized by the mass (or volume) fraction, $\phi(x, t)$, of ice at point x at time t. On the microscopic level, ϕ can only be 1 (within an ice grain) or 0 (in the water surrounding the grains). In the description of the mush used here, we assume that an average over many grains surrounding x at time t has been taken to obtain a smoothly varying $\phi(x, t)$ that can take any value in [0, 1], i.e. we describe the mush using the approach of mixture theory.

Before becoming enmeshed in the details of the phase evolution we first give an outline of the life cycle for the case $\alpha > 1$. Four distinct stages can be identified in our solution.

(1) **The gestation and birth of the mush** *(described in this section).*

During $0 \leqslant t \leqslant t_1 = O(1)$, the initial state (2.9) evolves until at $t = t_1$ (say)

$$(3.1)\qquad u_{1x}(s, t_1) = 0,$$

for the first time. Because $t_1 = O(1)$ and $\lambda \gg 1$, we have $s(t_1) = s_i$, to leading order. To the same accuracy, we may replace s in (3.1) by s_i.

(2) **Growth of the mush** *(Section 4).*

During $t_1 < t \leqslant t_2 = O(1)$, $s_1(t)$ moves towards s_c and the mush spreads quickly from s_i, and fills $s_1(t) < x < s_2(t)$, where s_1 and s_2 are to be determined. The solution now consists of three distinct regions

(i) **ice**. In $0 < x < s_1(t)$ we have

$$u = u_1(x, t) < 0, \quad \phi \equiv 1;$$

(ii) **mush**. In $s_1(t) < x < s_2(t)$ we have

$$u = u_3(x, t) \equiv 0, \quad 0 < \phi \equiv \phi_3(x, t) < 1;$$

(iii) **water**. In $s_2(t) < x < 1$ we have

$$u = u_2(x, t) > 0, \quad \phi \equiv 0.$$

We may not assume, *a priori*, that ϕ is continuous at s_1 and s_2, and we define

$$\phi_1 = \lim_{x \to s_1+} \phi_3, \qquad \phi_2 = \lim_{x \to s_2-} \phi_3,$$

to be the values of ϕ at the edges of the mush. Since domain (ii) is everywhere in phase equilibrium, $u_3 \equiv 0$, so that (2.13) is replaced by the demands

$$\lambda(1 - \phi_1)\dot{s}_1 = u_{1x}(s_1, t), \tag{3.2}$$

$$\lambda\phi_2\dot{s}_2 = -u_{2x}(s_2, t). \tag{3.3}$$

Because $t_2 = O(1)$ and $\lambda \gg 1$, the latter shows that $s_2(t_2) = s_i$ to leading order. The rapid motion of s_1 occurs, despite the enormous size of λ, because the system satisfies (3.2) by [cf. (3.1)]

$$\phi_1 = 1, \qquad u_{1x}(s_1, t) = 0. \tag{3.4,5}$$

(3) **Death of the mush** *(Section 4).*

During $t_2 < t \leqslant t_3 = O(\lambda)$, the mush contracts slowly in the following manner. While s_1 continues to approach s_c it does so ever more slowly. The trailing edge, s_2, has meanwhile been moving towards s_c on the slow time scale. Eventually there comes a time, t_2, when s_1 in approaching s_c is travelling as slowly as, and later even more slowly than, s_2. In a very definite sense, stage 2 is a part of stage 3, rather than being anterior to it.

(4) **Post-mush evolution** *(Section 4).*

During the final phase $t > t_3$, the solution at last evolves to the state (2.6) on the $O(\lambda)$ timescale and s becomes s_∞. The solution essentially coincides with that obtained in Section 2 for $s_i \leqslant s_c$.

In the remainder of this section we develop the solution for stage (1), that is, we solve (2.1) – (2.3) subject to (2.5) and (2.9). Since $\lambda \gg 1$, we replace (2.13) by

$$s = s_i,$$

so that for this period of the evolution, to leading order, the interface remains at the initial position. We are therefore effectively seeking the temperature distribution on a fixed interval and this can be obtained using standard Laplace transforms. Since $s_i < 1$ an appropriate representation is

$$(3.7)\quad u_1(x,t) = -\frac{(x-s_i)(x-s_c^2/s_i)}{s_c^2} + \Lambda s_i^2 \sum_{n=0}^{\infty} \frac{1}{(2n+1)^3} \sin\left(\frac{(2n+1)\pi x}{s_i}\right) \exp\left[-\frac{(2n+1)^2\pi^2 t}{s_i^2}\right] \qquad 0 < x < s_i,$$

$$(3.8)\quad u_2(x,t) = \frac{(x-s_i)(2-s_i-x)}{s_c^2} + 4\Lambda(1-s_i)^2 \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)^3} \times \cos\left(\frac{(2n+1)\pi(1-x)}{2(1-s_i)}\right) \exp\left[-\frac{(2n+1)^2\pi^2 t}{4(1-s_i)^2}\right], \qquad s_i < x < 1,$$

where $\Lambda = 8(\alpha s_i^2 - 1)/\pi^3 s_c^2$. This solution holds until (3.1) is obeyed for the first time at $t = t_1$. This onset time for the appearance of the mush can be easily found from (3.7) to be the solution of

$$(3.9)\quad \sum_{n=0}^{\infty} \frac{1}{(2n+1)^2} \exp\left[-\frac{(2n+1)^2\pi^2 t_1}{s_i^2}\right] = \frac{\pi^2(s_i^2 - s_c^2)}{8s_i^2(1-\alpha s_c^2)}$$

In Table I we list onset times obtained by numerical solution of (3.9) for

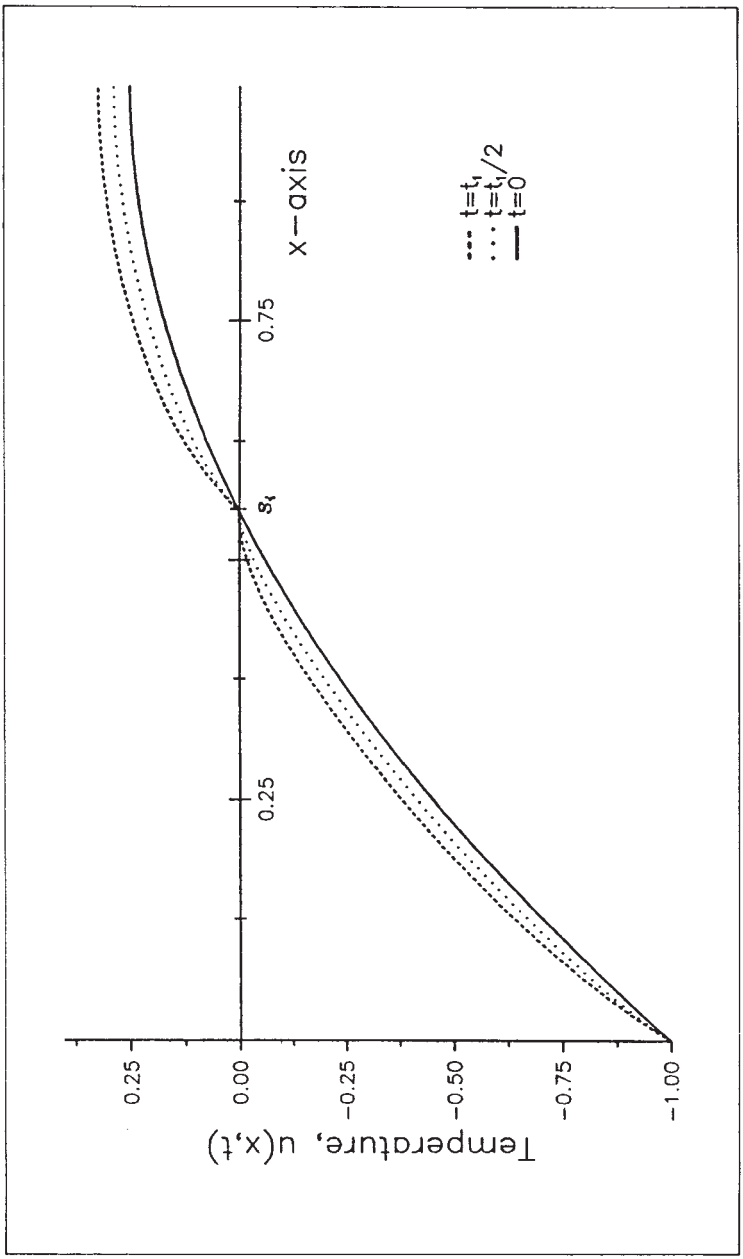


Figure 1. The evolution of the temperature distribution $u(x, t)$ for $r = 15$, $\alpha = 1.25$, from an initial state to just prior to the onset of a mush at $t = t_1$ (= 6.62832e-03). Since $\lambda \gg 1$, the position of the ice/water interface at $x = s_1$ (= 0.5528) does not move to leading order and at onset t_1, $u_{1x}(s_1, t_1) = 0$.

Table I. Non-dimensional onset times for the appearance of a mushy zone for various values of the volumetric heating, r, and the parameter α, that characterizes the initial state of the temperature field [see (2.9)].

	$\alpha = 1 + p$				
r	p=1/64	p=1/16	p=1/4	p=1/2	p=1
5	5.6566e-03	2.5264e-02	NO MUSH	NO MUSH	NO MUSH
10	7.8499e-04	3.3640e-03	1.7533e-02	NO MUSH	NO MUSH
15	2.9632e-04	1.2585e-03	6.2832e-03	1.6994e-02	NO MUSH
20	1.5429e-04	6.5291e-04	3.2057e-03	8.1630e-03	NO MUSH
50	2.1480e-05	9.0927e-05	4.3511e-04	1.0666e-03	2.9699e-03
100	4.9584e-06	2.1636e-05	1.0324e-04	2.5040e-04	6.8176e-04
150	2.0172e-06	9.3877e-06	4.5076e-05	1.0902e-04	2.9477e-04

various ranges of r and α and in Figure 1 we show stages of a temperature distribution evolution up to the onset time t_1.

4. GROWTH AND DEATH OF THE MUSH

As indicated in Section 3, at the onset time t_1, the leading edge of the mush $x = s_1(t)$ moves to the left towards $x = s_c$. The position of this solid/mush interface and the temperature distribution in the solid phase are determined by solving on $(x, t) \in [0, s_1(t)] \times [t_1, t)$

$$u_{1t} = u_{1xx} + r, \quad r = 2/s_c^2, \tag{4.1}$$

subject to

$$u_1(0, t) = -1, \quad u_1(s_1, t) = 0, \quad u_{1x}(s_1, t) = 0. \tag{4.2,3,4}$$

The initial data, $u_1(x, t_1)$, is given by (3.7) and, according to (3.2) and (4.4),

$$\phi_1 = 1. \tag{4.5}$$

These equations have been solved for $r = 15$, $\alpha = 1.25$. This typical case provides our principal example throughout this section. Standard numerical techniques were used. In Figure 2 we illustrate the manner in which the interface moves across and show the development of the temperature field.

The rate at which s_1 approaches s_c decreases with $s_1 - s_c$. If s_1 were not eventually overtaken by the trailing edge, $x = s_2(t)$ at time t_3 (say) with the concomitant loss of the mush, the evolution of s_1 to s_c of the solution would take an infinite time to complete. To investigate this point, we develop the

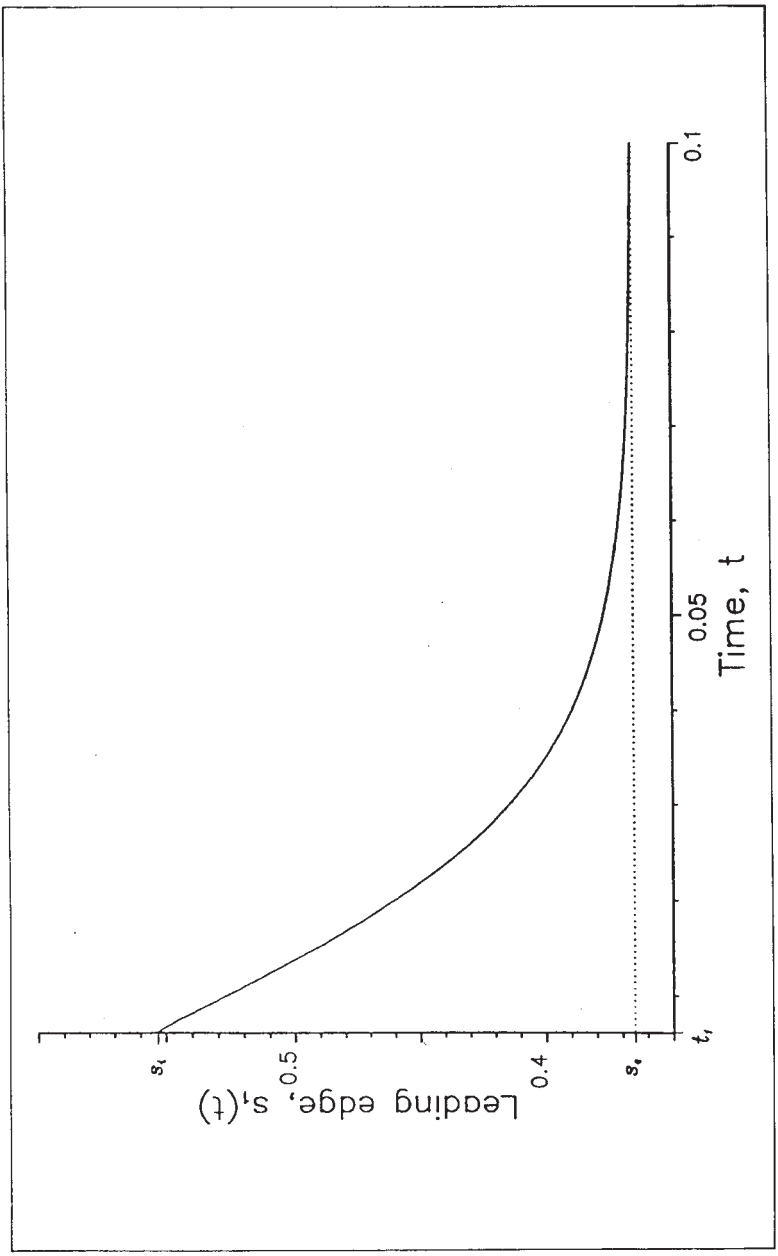


Figure 2. (a) The evolution of the leading edge $x = s_1(t)$ of the mushy zone for the case $r = 15$, $\alpha = 1.25$, showing that the edge approaches the critical position $x = s_c$ (= 0.3652) increasingly slowly; see (4.14).

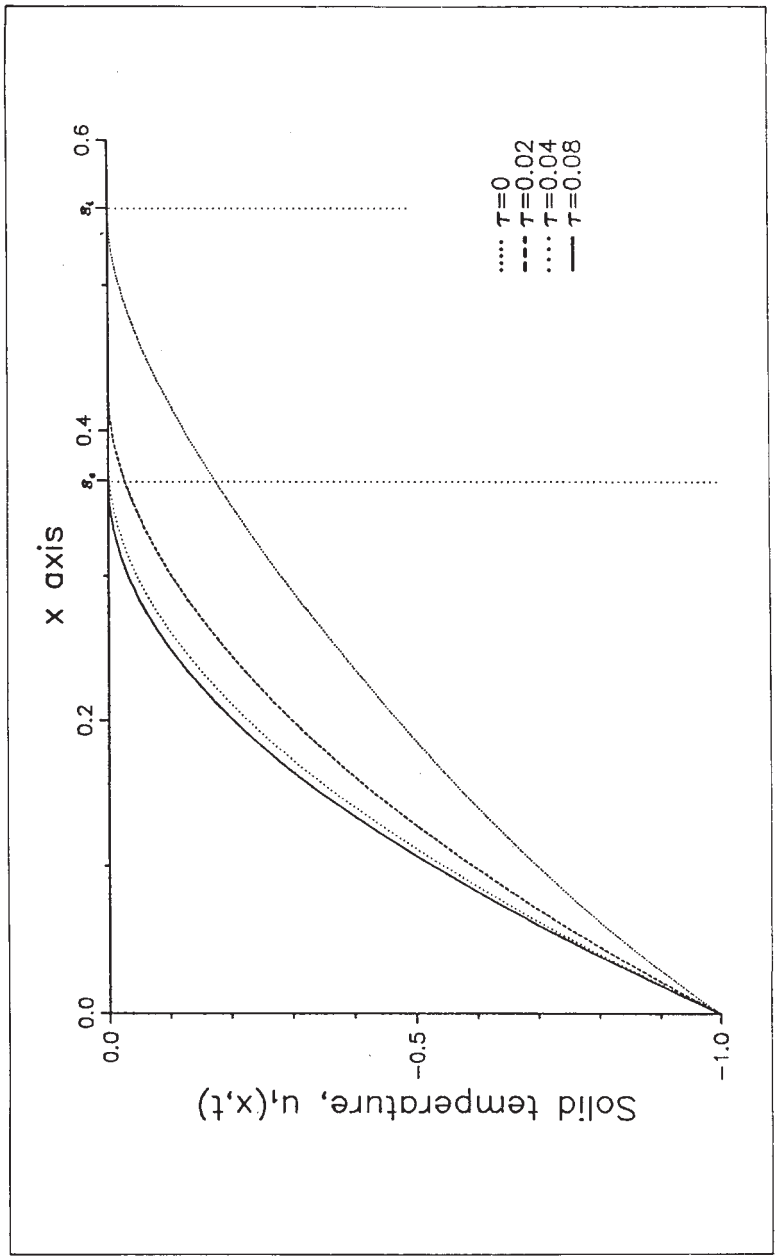


Figure 2. (b) The temperature distribution $u_1(x, t)$ within the solid for $r = 15$, $\alpha = 1.25$ showing how the solid/mush interface moves towards the critical position $x = s_c$ (= 0.3652) for times $\tau = t - t_1$ subsequent to the onset time t_1.

solution to (4.1)-(4.4) for small $s_1 - s_c$ and small $\dot{s}_1$. The solution we seek is asymptotic to the steady solution that exists when $s_1 = s_c$, namely

$$u_1(x) = -(1 - x/s_c)^2. \tag{4.6}$$

For small $s_1 - s_c$, we seek solutions of the form

$$u_1(x, t) = -(1 - x/s_c)^2 - v(x)\, e^{-k^2 t}, \tag{4.7}$$

where k and v are to be determined and v is of order $s_1 - s_c$. Substituting (4.7) into (4.1) and (4.2), we obtain ($' = d/dx$)

$$v'' + k^2 v = 0. \tag{4.8}$$

$$v(0) = 0. \tag{4.9}$$

We expand conditions (4.3) and (4.4) about $s_1 = s_c$, to obtain

$$0 = u_1(s_1, t) = u_1(s_c, t) + (s_1 - s_c)\, u_{1x}(s_c, t),$$

$$0 = u_{1x}(s_1, t) = u_{1x}(s_c, t) + (s_1 - s_c)\, u_{1xx}(s_c, t),$$

with $O(s_1 - s_c)^2$ errors. After substitution from (4.7) we find, to the same accuracy

$$v(s_c) = 0. \tag{4.10}$$

$$s_1 - s_c = -\frac{1}{2} s_c^2\, v'(s_c)\, e^{-k^2 t} \tag{4.11}$$

By (4.8) - (4.10) we have

$$v(x) = A \sin kx, \qquad k = \pi/s_c, \tag{4.12,13}$$

for some constant A, so that by (4.11)

$$s_1 - s_c = \frac{1}{2} \pi s_c\, A e^{-(\pi/s_c)^2 t}. \tag{4.14}$$

In short, we now have

$$u_1(x, t) = -\left(1 - \frac{x}{s_c}\right)^2 - \frac{2}{\pi s_c} \sin\left(\frac{\pi x}{s_c}\right)(s_1 - s_c), \tag{4.15}$$

$$\dot{s}_1 = -(\pi/s_c)^2\, (s_1 - s_c), \tag{4.16}$$

both expressions being in error at order $(s_1 - s_c)^2$. For the illustrative case $r = 15$, $\alpha = 1.25$, if we determine for $s_1 - s_c = 0.01$ the constant A by matching the asymptotic solution, s_{1a}, of (4.14) with the numerical solution, s_1, then

the error satisfies $\| (s_{1a} - s_1)/s_1 \| \simeq 10^{-4}$. Matching nearer to s_c gives concomitantly reduced errors.

In the mush itself, that is for $s_1(t) < x < s_2(t)$, there is no heat conduction, as $u_3 \equiv 0$, so the heat sources act only to melt the ice. Thus

$$\phi_{3t} = -r/\lambda = -2/\lambda s_c^2, \tag{4.17}$$

which gives

$$\phi_3(x, t) = \phi_0(x) - 2t/\lambda s_c^2. \tag{4.18}$$

Here $\phi_0(x)$ must be such that $\phi_3(s_1(t), t) = 1$ for all relevant t, i.e. such that (4.5) is obeyed. We therefore introduce the function, $T(s_1)$ say, that is the inverse of $s_1(t)$, so that $s_1(T(x)) \equiv x$. (We note that s_1 is a monotonically decreasing function of t for $t > t_1$ so that the inverse exists). The statement that $\phi_3(s_1(t), t) = 1$ for all relevant t can then be restated as $\phi_3(x, T(x)) = 1$ for all relevant x. Then, by (4.18),

$$\phi_0(x) = 1 + 2T(x)/\lambda s_c^2. \tag{4.19}$$

Since by (4.18) and (4.19) $T(x)$ is monotonically decreasing for $s_c \leqslant x \leqslant s_i$, ϕ_0 and ϕ_3 are also monotonically decreasing functions of x, and

$$\phi_3(x, t) = 1 - 2[t - T(x)]/\lambda s_c^2. \tag{4.20}$$

In Figure 3, for $r = 15$, $\alpha = 1.25$, we graph ϕ_3 versus x for various times.

Finally, in this section we consider the trailing edge of the mush, that is the mush/liquid interface at $x = s_2(t)$. Throughout it moves slowly much in the manner of the persistent, sharp interface of the case $s_i < s_c$ discussed in Section 2. The explanation of this follows from (3.3): both $u_{2x}(s_2, t)$ and ϕ_2 are nonzero so that $\dot{s}_2 = O(\lambda^{-1})$. Compared with this slow evolution, diffusion processes are effectively instantaneous and it follows that (cf. (2.16))

$$u_2(x, t) = (x - s_2)(2 - x - s_2)/s_c^2. \tag{4.21}$$

By (3.3) we have

$$\lambda\phi_2\dot{s}_2 = -2(1 - s_2)/s_c^2, \tag{4.22}$$

and from (4.20) we have

$$\lambda\{1 - 2[t - T(s_2)]/\lambda s_c^2\}\dot{s}_2 = -2(1 - s_2)/s_c^2. \tag{4.23}$$

The numerical solution of (4.23) for the illustrative case used in this paper is represented in Figure 4 and shows that $t_3 = 0.534$.

The mush having disappeared at $t = t_3$, the subsequent evolution from $s = s_c$ is given by the marginal case $s_i = s_c$ of the $s_i \leqslant s_c$ theory developed in Section 2.

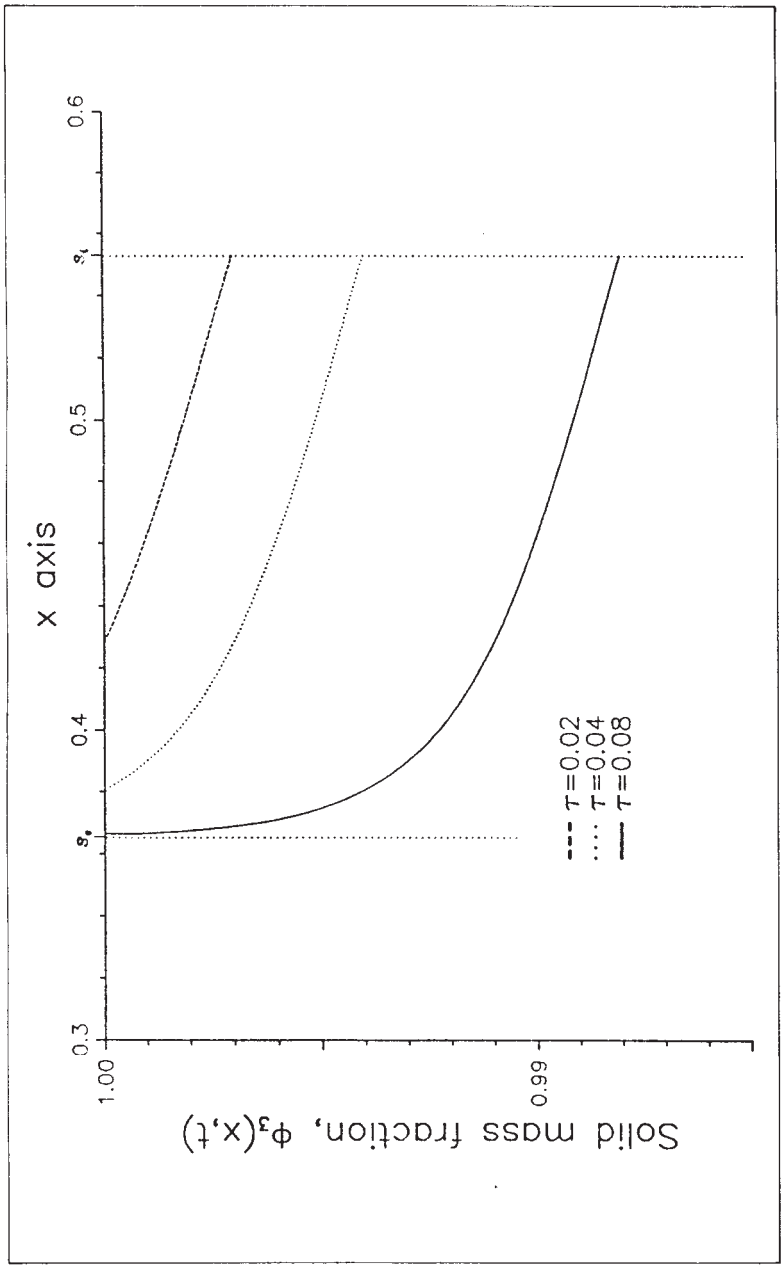


Figure 3. The distribution of the mass fraction of the solid, $\phi_3(x, t)$, in the two-phase mush at various times $\tau = t - t_1$ (> 0), where t_1 is the onset time.

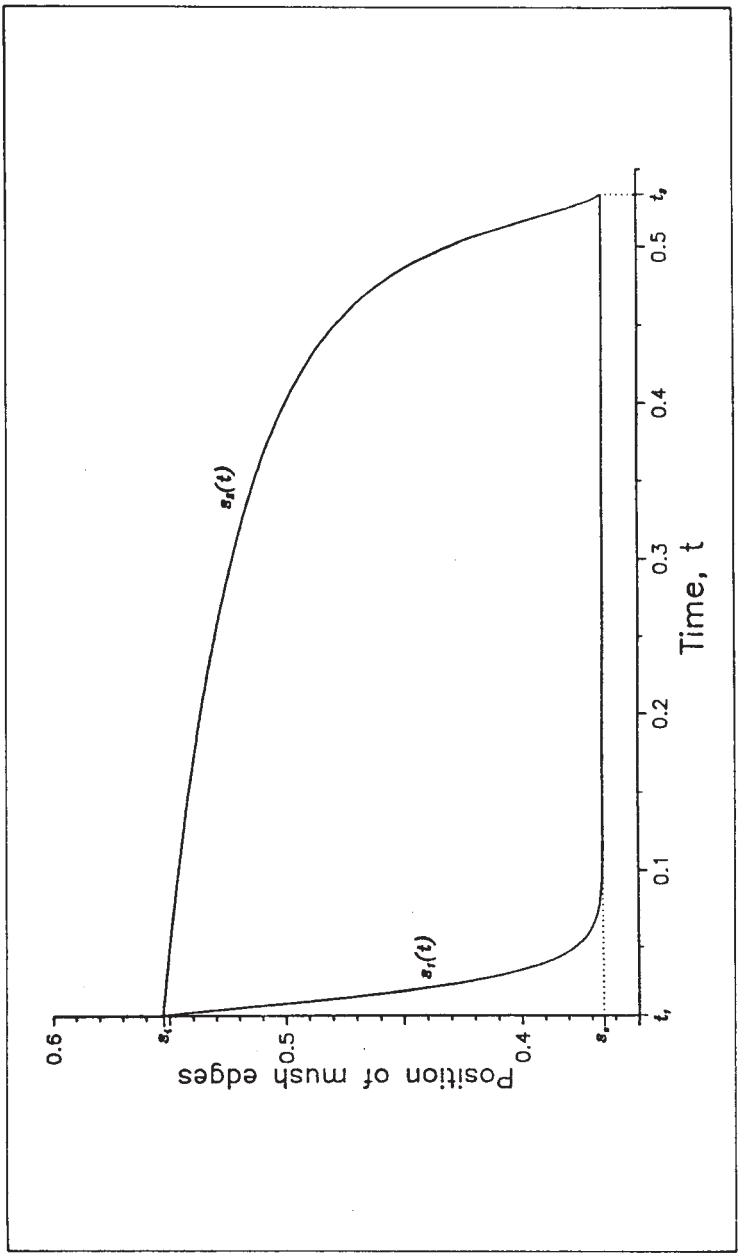


Figure 4. The evolution of the position of the edges of the mushy region for the case $r = 15$, $\alpha = 1.25$. The region quickly forms at time t_1 by the leading edge $x = s_1(x, t)$ moving across towards $x = s_c$ and the trailing edge moves across on the slow time scale with the mush disappearing at $t = t_3$. Thereafter the system has a sharp phase boundary that moves towards the equilibrium position $x = s_\infty$ (= 0.06905).

The relevant solution to (2.17) is now, instead of (2.18),

$$t = t_3 + \frac{\lambda s_c^2}{2(\bar{s}_\infty - s_\infty)}\left[\bar{s}_\infty \ln\left(\frac{\bar{s}_\infty - s}{\bar{s}_\infty - s_c}\right) - s_\infty \ln\left(\frac{s - s_\infty}{s_c - s_\infty}\right)\right]. \tag{4.24}$$

5. THE CASE $\alpha \leqslant 1$

Suppose that water is absent initially, that is u_i is everywhere negative, but nevertheless still satisfies the endpoint conditions (2.2) and (2.3). A typical example is provided by (2.9) with $\alpha \leqslant 1$. If $\alpha < 1$, the first stage of the evolution is again on the fast time scale with u increasing until, at time $t = t_1$, u first becomes zero at $x = 1$. Thereafter a mush rapidly develops as in the previous section but the third and fourth stages dealing with the diminution and demise of the mush are markedly different.

The central observation is that, from (4.22), since s_2 is initially 1, $\dot{s}_2 = 0$ until ϕ_2, the mass fraction of solid at $x = 1$, vanishes. Hence the trailing edge must languish motionless until the volumetric heating has dissolved all the ice at $x = 1$. According to (4.20), since $T(1) = t_1$, ϕ_2 will vanish at $t = t_2$, where

$$t_2 = t_1 + \frac{1}{2}\lambda s_c^2, \tag{5.1}$$

and

$$\phi_3(x, t_2) = 2[T(x) - t_1]/\lambda s_c^2. \tag{5.2}$$

Since $T(x)$ is monotonically decreasing and $T(1) = t_1$, it follows from (5.2) that $\phi_3 \geqslant 0$ at $t = t_2$, within the mush, equality occurring only at $x = 1$. Clearly, from (5.2), $\phi_3 = O(1/\lambda s_c^2)$ at $t = t_2$ and the left-hand side of (3.3) is, therefore, at most of order $2\dot{s}_2/s_c^2$. This means that s_2 will envolve on the fast time scale in contrast to the situation of the previous section, where s_2 always moved on the slow scale. Because the evolution is rapid, we cannot adopt the approximation (2.14) but must retain the u_{2t} term in (2.1). Thus, for $t > t_2$ we must solve

$$u_{2\hat{t}} = u_{2xx} + r, \quad r = 2/s_c^2, \tag{5.3}$$

$$u_2(s_2, \hat{t}) = 0, \quad u_{2x}(1, \hat{t}) = 0, \tag{5.4,5}$$

$$2[T(s_2) - t_1 - \hat{t}]\,\dot{s}_2/s_c^2 = -u_{2x}(s_2, \hat{t}), \tag{5.6}$$

where $\hat{t} = t - t_2$. We observe that both sides of equation (5.6) vanish when $\hat{t} = 0$. Consequently, some care will be needed to determine the manner in which the trailing edge starts moving. It is clear that if s_2 moves off at $\hat{t} = 0$

with a speed that is less than or equal to the speed of s_1 initially, then s_2 must exactly mimic the journey of s_1 since ϕ_2 will always then vanish at the trailing edge. This is not the case and to illustrate we concentrate on the case $\alpha = 1$ so that stage 1 does not occur and the onset time, t_1, is zero. To investigate the solution for $1 - s_2 \ll 1$, $\hat{t} \ll 1$, we assume the first few terms of a double power series solution with

$$u_2 \doteq 2a_1 t - a_2(1-x)^2, \tag{5.7}$$

where a_1, a_2 are constants to be determined. Condition (5.5) is identically satisfied and (5.3,4,6) give

$$2a_1 = r - 2a_2, \quad 2a_1\hat{t} = a_2(1-s_2)^2, \quad [T(s_2) - \hat{t}]\,\dot{s}_2 = -s_c^2 a_2(1-s_2). \tag{5.8}$$

To make further progress, we need information regarding $T(x)$ which is related to the motion of the leading edge. In the appendix to this paper A. M. Soward has shown that when $\alpha = 1$ is possible to construct a similarity solution for the leading edge by using a boundary layer analysis. His result shows that, for $s_2 \to 1$,

$$T(s_2) \sim (1-s_2)^2/4\Gamma^2, \tag{5.9}$$

where Γ is the solution of the transcendental equation

$$s_c^2\,\Gamma \operatorname{erfc}\Gamma + (1-s_c^2)[\Gamma \operatorname{erfc}\Gamma - e^{-\Gamma^2}/\sqrt{\pi}] = 0. \tag{5.10}$$

From (5.8) we find that

$$a_2 = [(1 + 20\Gamma^2 + 4\Gamma^4)^{1/2} - (1 + 2\Gamma^2)]/8\Gamma^2 s_c^2. \tag{5.11}$$

Equation $(5.8)_2$ shows that

$$s_2(t) \sim 1 - (2a_1\hat{t}/a_2)^{1/2}. \tag{5.12}$$

By comparing with the result (5.9), which implies that

$$s_1(t) \sim 1 - 2\Gamma\sqrt{\hat{t}}, \tag{5.13}$$

we see that, since $a_1/2a_2\Gamma^2 = 2s_c^2 a_2 + 1 > 1$, the trailing edge, when it does start to move, moves faster than the leading edge initially did. The same inequality shows that, for $s_2 \sim 1$, $\phi_2 > 2a_2^2(1-s_2)^2/\lambda a_1 > 0$, i.e. as stated above, ϕ_2 is nonzero except at $\hat{t} = 0$.

The asymptotic results (5.7), $(5.8)_1$, (5.11), (5.12) can be used as an initial condition when solving the system (5.3) - (5.6) numerically. In Figure 5 we illustrate the evolution of the edges of the mush for $r = 15$, $\alpha = 1$.

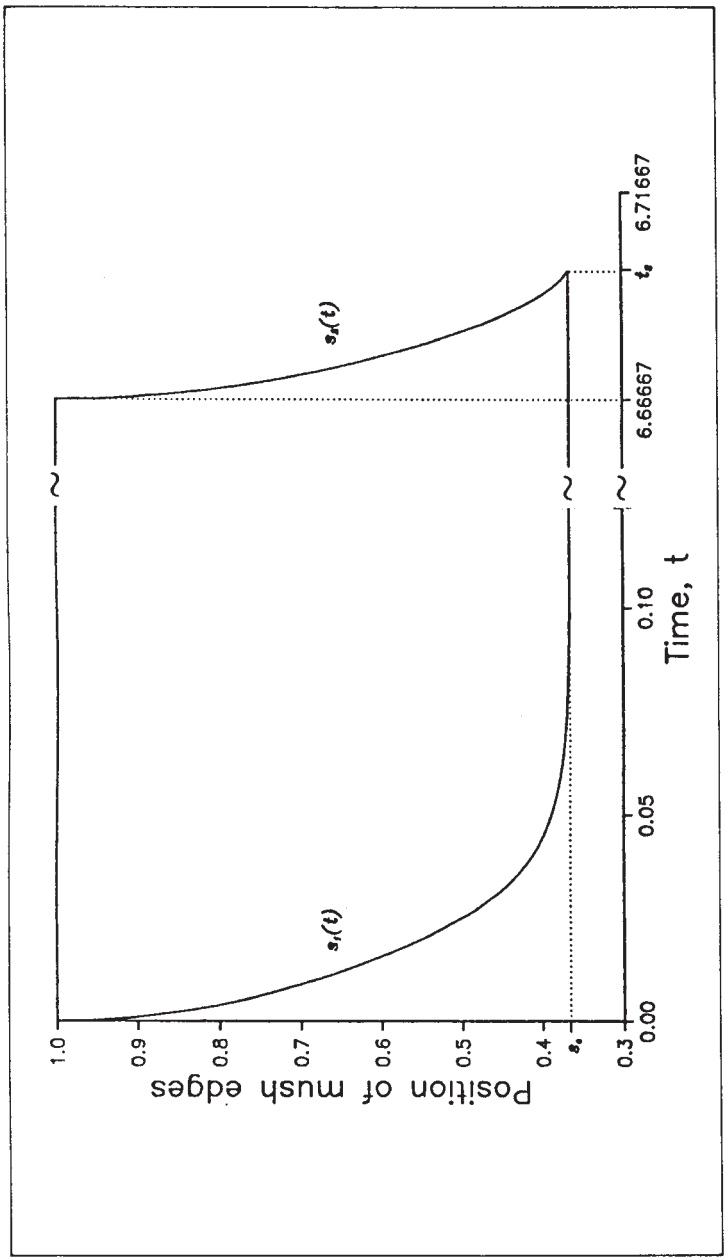


Figure 5. The evolution of the position of the edges of the mushy region for the case $r = 15$, $\alpha = 1$. For this case the onset time $t_1 = 0$ and the region forms by the leading edge $x = s_1(x, t)$ moving across towards $x = s_c$. In order to satisfy the Stefan condition (4.22), the trailing edge must remain motionless at $x = 1$ until the solid fraction there has been melted almost completely by volumetric heating. Thereafter the trailing edge $x = s_2(t)$ moves to $x = s_c$ on the fast time scale.

ACKNOWLEDGEMENTS

The research of one of the authors (P.H.R.) is sponsored by the U.S. Office of Naval Research under contract N00014-86-K-0691 with the University of California, Los Angeles.

REFERENCES

Atthey, D.R. 1974, A finite difference scheme for melting problems. *J. Inst. Maths. Applics.* **30**, 353.

Chalmers, B. 1964 *Principles of Solidification.* John Wiley, New York.

Lacey, A.A. and Shillor, M. 1983, The existence and stability of regions with superheating in the classical two-phase one-dimensional Stefan problem with heat sources. *I.M.A. J. Appl. Math.* **30**, 215.

Lacey, A.A. and Tayler A.B. 1983 A mushy region in a Stefan problem. *I.M.A. J. Appl. Math.* **30**, 303.

Landau, L.D. and Lifshitz, E.M. 1969, *Statistical Physics. Vol. 5 Course of Theoretical Physics,* (2nd edition). Pergamon Press, Oxford.

Ockendon, J.R. 1975, Techniques of analysis in *Moving Boundary Problems in Heat Flow and Diffusion,* (ed. J.R. Ockendon and W.R. Hodgkins). Clarendon Press, Oxford.

Woodruff, D.P. 1980, *The Solid-Liquid Interface.* C.U.P. Cambridge.

APPENDIX

A.M. SOWARD
Department of Applied Mathematics, Newcastle University, U.K.

Early time solution for $\alpha = 1$

This appendix aims to re-inforce the analysis of the case $\alpha = 1$ described in Section 5, but only minor modifications are required to make the analysis applicable also to the more general case $\alpha > 1$ considered in Sections 2-4.

According to (2.9) for $\alpha = 1$, the initial state,

(A1) $$u_i(x) = -(1 - x)^2,$$

is one in which ice is present everywhere, but in which melting is about to be initiated at $x = 1$.

By writing

(A2) $$u_1(x, t) = u_i(x) + v(x, t),$$

we introduce the departure, $v(x, t)$, of the temperature from (A2) at later times; evidently

(A3) $$v(x, 0) = 0.$$

By (4.1) and the boundary conditions (4.2)-(4.4), v must be the solution of

(A4) $$v_t = v_{xx} + 2Q, \qquad 0 < x < s_1(t)$$

that satisfies, for $t > 0$,

(A5) $$v(0, t) = 0,$$

(A6) $$v(s_1, t) = (1 - s_1)^2,$$

(A7) $$v_x(s_1, t) = -2(1 - s_1),$$

where

(A8) $$Q = s_c^{-2} - 1, \quad r = 2/s_c^2.$$

For sufficiently small times, the solution in the interior of the region $0 < x < s_1(t)$ away from the end points is

(A9) $$v = V(t) = 2Qt.$$

This does not satisfy the boundary conditions at $x = 0$ and $x = s_1(t)$, and in their immediate neighbourhoods boundary layers form, whose widths are of

order $t^{1/2}$ and in which solution (A9) adjusts to conditions (A5)-(A7). So at $x = 0$, for example, the solution takes the similarity form

(A10) $$v = V(t)[1 - \psi(\xi)],$$

where $\xi = x/2t^{1/2}$. By (A4), $\psi(\xi)$ satisfies

(A11) $$\psi'' + 2\xi\psi' - 4\psi = 0,$$

where the prime denotes differentiation with respect to ξ. Condition (A4) and matching to (A9) require

(A12) $$\psi(0) = 1,$$

(A13) $$\psi(\xi) \to 0, \quad \text{as} \quad \xi \to \infty.$$

The solution to (A11) - (A13) is

(A14) $$\psi(\xi) = (1 + 2\xi^2)\,\text{erfc}\,\xi - 2\pi^{-1/2}\xi e^{-\xi^2},$$

where erfc denotes the complementary error function:

$$\text{erfc}\,\xi = \frac{2}{\pi^{1/2}} \int_{\xi}^{\infty} e^{-x^2}\,dx.$$

Of greater interest is the nature of the boundary layer at the moving end point, $x = s_1(t)$. To understand the structure of this layer, it is helpful to consider the comparison problem defined on the *fixed* interval $0 < x < 1$, and to solve (A4) subject to the boundary condition (A5) with an *evolving* value for $v(1, t)$. That value is determined by the demand that the resulting solution obeys the two conditions (A6) and (A7) at the interior point $x = s_1(t)$.

When

(A15) $$v(1, t) = V(t)\,(1 - \mu),$$

where μ is a constant as yet unknown, a similarity solution of (A4) can again be constructed which obeys (A15) and matches to (A9):

(A16) $$v = V(t)\,[1 - \mu\psi(\xi)],$$

where ψ is again given by (A14) but now $\xi = (1 - x)/2t^{1/2}$. Fortuitously, (A16) can also satisfy both (A6) and (A7) at an interior point

(A17) $$s_1(t) = 1 - 2\Gamma t^{1/2}$$

where Γ is another constant to be determined. Specifically, (A6) and (A7) are met when respectively

$$1 - \mu\psi(\Gamma) = 2\Gamma^2/Q, \tag{A18}$$

$$-\mu\psi'(\Gamma) = 4\Gamma/Q. \tag{A19}$$

By (A14) we therefore have

$$Q = \frac{\Gamma \operatorname{erfc} \Gamma}{\pi^{-1/2} e^{-\Gamma^2} - \Gamma \operatorname{erfc} \Gamma}, \tag{A20}$$

$$\mu = 1/\operatorname{erfc} \Gamma. \tag{A21}$$

The value of Γ is obtained by solving (A20), and then μ is derived from (A21).

We will examine the nature of the solution for the two extremes: (i) $Q \ll 1$ and (ii) $Q \gg 1$. In both cases the results will be valid only during the time in which the boundary layers at $x = 0$ and $x = s_1$ are non-interacting, which is the case when t is sufficiently small:

$$t \ll 1. \tag{A22}$$

(i) The Case $Q \ll 1$

When Q is small, it follows that

$$0 < 1 - s_c \ll 1, \tag{A23}$$

so that the initial state $u_i(x)$ is very close to the state (4.2) to which the temperature in the ice tends during the growth of the mush, namely $u_1 = (1 - x/s_c)^2$. The solutions of (A20) and (A21) are

$$\Gamma = \pi^{-1/2} Q + O(Q^2), \tag{A24}$$

$$\mu = 1 + 2\pi^{-1} Q + O(Q^2). \tag{A25}$$

Since μ is close to unity, the boundary layer at $x = s_1(t)$ is approximately the same as that at the stationary boundary $x = 0$. Furthermore since Γ is small the displacement,

$$1 - s_1 \sim 2Q(t/\pi)^{1/2}, \tag{A26}$$

of the ice/mush interface away from $x = 1$ is ever small compared with the boundary layer thickness $t^{1/2}$.

(ii) The Case $Q \gg 1$

When Q is large, it follows that

$$0 < s_c \ll 1, \tag{A27}$$

so that the ice/mush interface eventually moves to a location far from its initial position at $x = 1$. The asymptotic solutions of (A20) and (A21) are

(A28) $$\Gamma \sim \left(\frac{1}{2} Q\right)^{1/2} + O(Q^{-1/2}),$$

(A29) $$\mu \sim \left[\left(\frac{1}{2}\pi Q\right)^{1/2} + O(Q^{-1/2})\right] e^{Q/2}.$$

In contrast to the case of small Q, the displacement,

(A30) $$1 - s_1 \sim (2Qt)^{1/2},$$

of the ice/mush interface is large compared with the boundary layer thickness. Only the large ξ asymptotic behaviour is relevant to the interval $0 < x < s_1(t)$, and that gives

(A31) $$v \sim V(t)\,[1 - 2e^{-\xi}/Q],$$

where

(A32) $$\xi = (Q/2t)^{1/2}\,(s_1 - x).$$

Since the interface $x = s_1(t)$ moves very rapidly, the boundary layer in front of it is very thin, of order $(t/Q)^{1/2}$ in width. Indeed, the location of the moving interface can be found by ignoring the boundary layer entirely, and simply applying the boundary condition (A6) directly to the mainstream, so obtaining

(A33) $$(1 - s_1)^2 = V(t).$$

This gives the result (A30) directly. The boundary layer is passive, and is only required so that condition (A7) on $v_x(s_1, t)$ can be met.

Clearly the solution (A30) for $s_1(t)$ is valid only for the very short times for which Qt is small. At later times we can make further progress by ignoring the boundary layer (A31) generated by the moving interface. Thus we apply condition (A6) directly to the mainstream solution, as modificed by the boundary layer at the stationary wall, $x = 0$. We therefore substitute (A10) into (A6), set

(A34, A35) $$s_1 = 2\eta t^{1/2} \qquad \tau = 2Qt,$$

and solve for τ in terms of η, so obtaining

(A36) $$\tau^{-1/2} = (2/Q)^{1/2}\ \eta + [1 - \psi(\eta)]^{1/2}.$$

When η is large ($\psi(\eta) \ll 1$) and the moving interface $x = s_1$ has not penetrated the boundary layer significantly, we recover solution (A33) in the form

(A37) $$\tau^{-1/2} - 1 = (2/Q)^{1/2}\ \eta,$$

valid provided that

(A38) $$Q^{1/2}(1-\tau) \gg 1.$$

An interesting feature of the result (A36) is that the advance of the rapidly moving interface is arrested sharply in the tail of the boundary layer (A10), close to $s_1 = 2\eta_e t_e^{1/2}$ at time $t_e = \tau_e/2Q$, where η_e is the solution of

(A39) $$\eta_e^2 \exp(\eta_e^2) = Q^{1/2} \quad \eta_e \gg 1,$$

and τ_e is the corresponding solution of (A36). To see how this is achieved, we note that

$$\psi(\eta) \sim \pi^{-1/2}\, \eta^{-3}\, e^{-\eta^2}, \quad \text{as} \quad \eta \to \infty,$$

and set

(A40) $$\eta = \eta_e + \frac{1}{2}\rho.$$

The lowest order approximation of (A36) then becomes

(A41) $$(2Q)^{1/2}\, \eta_e \left(\frac{1}{\tau^{1/2}} - \frac{1}{\tau_e^{1/2}} \right) = \eta_e \rho + \frac{1}{(2\pi)^{1/2}} \left(1 - e^{-\eta_e \rho}\right),$$

where, correct to leading order,

(A42) $$\tau_e^{-1/2} - 1 = (2/Q)^{1/2} \eta_e,$$

as in (A37). The asymptotic solutions of (A41) for positive and negative ρ are given by

(A43) $$2(\eta - \eta_e) \sim \begin{cases} \left(\dfrac{Q}{2\tau_e}\right)^{1/2} \left(1 - \dfrac{\tau}{\tau_e}\right) & Q^{-1/2} \ll \tau_e - \tau \ll 1, \\ -\dfrac{2}{\eta_e} \ln\left[\left(\dfrac{\pi Q}{\tau_e}\right)^{1/2} \left(\dfrac{\tau}{\tau_e} - 1\right) \eta_e\right], & Q^{-1/2} \ll \tau - \tau_e \ll 1. \end{cases}$$

This result highlights the nature of the transition that occurs at the time $t = t_e$.

Manuscript received: December 19 1989

SAACM - Vol. 2, n. 3, 1992

Relaxation effects and the evolution of a mush

R.N. HILLS

Department of Mathematics
Heriot-Watt University
Edinburgh, U.K.

P.H. ROBERTS

Department of Mathematics
University of California
Los Angeles, CA 90024, U.S.A.

Abstract. *An earlier paper (Hills, Roberts and Soward, 1991) was concerned with the one-dimensional evolution of a layer of stagnant water and ice that contains sources of heat. One end of the layer is thermally insulated while the other is maintained at a sub-zero temperature. In Hills et al. (1991) it was assumed that the melting/freezing processes occur instantaneously and it was shown that under certain circumstances, depending on the degree of heating and the position of the initial interface, a transient region of mixed phase develops. This paper is concerned with a non-zero relaxation time for the phase change. When the relaxation time is nonzero, even the sharp front that appears in classical theory (where the relaxation time is zero) becomes fuzzy or «blurred». The structure of this blurred front is analyzed for two choices of the law of relaxation. It is shown that the effect of the relaxation time is passive in the sense that the pattern of phase evolution (Hills et al., 1991) remains substantially unchanged.*

1. GENERAL APPROACH TO EVOLVING SYSTEMS OF MIXED PHASE

The dynamics and thermodynamics of an evolving system of mixed phase raises formi- dable theoretical difficulties even in the simplest case (considered here) in which the material is a pure substance and alloying is absent. The development of the theory has followed a natural evolution, and is still incomplete.

At first, situations were studied in which the phases are separated discontinuously at a sharp interface. This may move through the material as a freezing or a melting front, depending on whether the evolution requires a net outflux of heat from, or influx of heat to, the front. The resulting Stefan condition at the front suffices to determine uniquely the mathematical solutions to Stefan problems of this type, and many such problems have been solved; see for example Rubenstein (1971). The key parameter in the theory

is the Stefan number:

$$\lambda = L/c_p T_M, \tag{1.1}$$

which is the ratio of the time scale, $\tau_L = \ell/U$, of latent heat release, where $U = \kappa T_M/\rho L\ell$ is a typical speed of front motion, and the conduction time scale, $\tau_\kappa = \rho c_p \ell^2/\kappa$:

$$\lambda = \tau_L/\tau_\kappa. \tag{1.2}$$

Here T_M is the melting temperature, L is the latent heat per unit mass and ℓ is a characteristic length over which T varies. For simplicity we assume the density ρ and specific heat at constant pressure c_p are the same for solid and liquid.

It was during this stage of theory development that the power and utility of asymptotic methods in the subject became apparent for the first time. Solutions for $\lambda \to \infty$ could be far more readily obtained than solutions for arbitrary λ and they provided a reliable qualitative guide to the latter. When λ is large, the front moves so slowly that the temperatures in the phases on either side of the front adjust almost instantaneously to the changes in the imposed conditions (heating or cooling at walls; volumetric heating or cooling) that cause the front to move. Asymptotic methods also proved useful later; see below. But in subsequent developments, it was realized that, when the material contains sufficiently powerful sources or sinks of heat, regions of mixed phase would be created. This is easily imagined: if a source provides heat to the solid phase more rapidly than thermal conduction can remove it, the difference will cause a region to be formed in which the material is neither totally solid nor totally liquid.

Three theoretical approaches to systems containing mushy regions have been advocated. First, there is the so-called H-method based upon the enthalpy of the whole system. If u is the non-dimensional temperature, with its zero corresponding to the melting temperature, the enthalpy for the classical Stefan problem is written as

$$H(u) = u + \frac{1}{2} L\hat{\phi}, \tag{1.3}$$

where $\hat{\phi}$ is a phase index with $\hat{\phi} = -1$ corresponding to solid and $\hat{\phi} = +1$ liquid. The enthalpy can be shown to satisfy the energy equation

$$\partial_t H(u) = \kappa \nabla^2 u, \tag{1.4}$$

where κ is the thermal diffusivity. Equations (1.3), (1.4) are then equivalent, in a weak sense, to the classical Stefan problem [see Oleinik (1960)]. If H^S and H^L are the enthalpies of the pure phases, a mushy zone is signalled by $u = 0$ but with the system

enthalpy satisfying $H^S < H < H^L$. The enthalpy method formed the basis of the numerical investigation of Atthey (1974) and many subsequent studies.

The second method represents an extension of the approach so successfully used for systems lacking a mush [see Meirmanov (1981), Primicerio (1982)]. The pure solid and liquid regions are dealt with as before but in the intermediate region the mass fraction of the solid will consist of solid crystals surrounded by liquid. In a first approximation, nothing can be said about the distribution of the sizes and shapes of the crystals, only their mass fractions are known. The theory therefore treats the mush as a special kind of mixture, but one that is amenable to precise mathematical treatment:

$$u = 0, \qquad L\partial_t \phi = -r, \tag{1.5}$$

where r is the rate per unit mass at which the source injects heat and ϕ the mass fraction of the solid phase. The single phase regions are connected across the mushy zone by *two* Stefan conditions, one at the fluid-mush interface and the other at the solid-mush interface. The mathematical well-posedness of this formulation has been extensively studied [see Fasano and Primicerio (1985)]. In particular, solutions to Stefan problems of this type are uniquely determined. We demonstrated this recently by giving an explicit solution to Ockendon's problem in the limit of large λ (Hills, Roberts and Soward, 1991). In non-dimensional variables, with an x suffix denoting $\partial/\partial x$, this one-dimensional problem is

Ockendon's Problem *[see Ockendon (1975)].*

Water and/or ice (assumed to be stationary and of the same density) exists at every point of domain $0 < x < 1$, *in which there is a uniform distribution of heat sources that, in the absence of heat conduction and latent heat, would cause the temperature,* u, *to rise everywhere at the constant rate,* r, *where* $r > 0$. *The wall* $x = 1$ *is insulating* ($u_x = 0$) *and the wall* $x = 0$ *is held at a freezing temperature:* $u = -1$. *Given the distribution of ice* ($\phi = 1$) *and water* ($\phi = 0$) *together with* u *at all* x *in* $0 < x \leq 1$ *for* $t = 0$, *find* u *and the mass fraction* ϕ *of ice for all* x *in* $0 < x \leq 1$ *at all later times,* $t > 0$.

The third method is the so-called «phase-field model» [see Caginalp and Fife (1987), Penrose and Fife (1990) and references therein]. It is possible to view this approach as an outgrowth of the enthalpy method in which the step-function form for the phase index is «rounded» into a (steep) continuous function. Thus, (1) $\widehat{\phi} = \widehat{\phi}(\mathbf{x}, t)$ and the value

(1) Some ambiguity exists in the literature concerning the relationship between the phase index $\widehat{\phi}$ and the mass of fraction. Some authors refer to the phase index as «the order parameter» and make no connection to the concentration. Of course, it is not inevitable that the order parameter is

at any x and t reveals which phase is present at that place and time. A pair of coupled partial differential equations are postulated governing u and $\hat{\phi}$:

$$u_t = \nabla^2 u - \frac{1}{2}\lambda \hat{\phi}_t + r, \tag{1.6}$$

$$\alpha \hat{\phi}_t = \nabla^2 \hat{\phi} - \epsilon^{-2}\left[f\left(\hat{\phi}\right) - u\right], \tag{1.7}$$

where α and ϵ are positive constants. The last term of (1.4) represents the $\hat{\phi}$-derivative of a coarse-grained free energy function, $F(u, \hat{\phi})$, expanded about $u = 0$: f is «N-shaped», that is $f(-1) = f(1) = 0$, $\mathrm{d}f/\mathrm{d}\hat{\phi}$ is positive at $\hat{\phi} = -1$, 1, and f has in general one further zero in $-1 < \hat{\phi} < 1$. When $\hat{\phi}$ is interpreted as the solid mass fraction, equation (1.6) is the usual heat conduction equation, in which the latent heat of phase conversion and heating due to external sources are included. Equation (1.7) is appropriate when $\hat{\phi}$ is not a conserved quantity but if $\hat{\phi}$ is conserved then the equation

$$\alpha \hat{\phi}_t = -\nabla^2\left[\nabla^2 \hat{\phi} - F_{\hat{\phi}}\right] \tag{1.8}$$

has been proposed [see Langer (1971), Hohenberg and Halperin (1977), Binder (1987)]. The core of the phase-field model is, however, the steady-state part of the theory, and on this the two theories essentially agree. It is in the treatment of the kinetic effects that (1.7) and (1.8) differ.

The aim and hope of the phase-field approach is to solve for $\hat{\phi}$ and u without introducing sharp fronts and Stefan conditions. Because $\epsilon \ll 1$, solutions develop in which $\hat{\phi}$ is close to -1 or to 1 nearly everywhere, and in which $\hat{\phi}$ changes rapidly from one to the other, in an $O(\epsilon)$ distance. Such a «blurred front» (as we shall call it) mimics the sharp front of earlier theories separating solid and liquid phases, and can be analyzed by asymptotic methods. A mush manifests itself in phase-field theory through the occurrence of an instability that in turn leads to chaotic regions formed by a corporation of many blurred fronts. The index $\hat{\phi}$ provides a detailed description of the solid grains of the mush and the surrounding liquid, although it is difficult, in numerical computations, to model systems in which the grain size is small compared with the dimensions

a mass fraction: in superfluid helium the order parameter is complex and does not coincide with the superfluid density. But when $\hat{\phi}$ is an imprecise order parameter it is difficult to know how to interpret $\hat{\phi}$ when it is neither -1 nor 1, as is the case within the blurred front described below. It is also hard to interpret the so-called «homogeneous solution» which has been analyzed in the phase-field literature and in which $f(\hat{\phi}) = u =$ constant, an equation that in general requires $\hat{\phi}$ to be neither -1 nor 1. We follow Binder (1987) in supposing the phase index $\hat{\phi}$ prescribes the solid mass fraction as $\frac{1}{2}(1 - \hat{\phi})$.

of the mushy zone. But is it required in many practical applications? What is often most valuable is a description of the *average* evolution of systems containing mushy zones.

There are, to be sure, two avenues open. First, many numerical integrations of the phase-field model can be performed, starting from initial states that differ from one another only by a small random «jitter» in $\widehat{\phi}$. The expectation is that, because of the spinodal energy range, an instability will develop to herald the appearance of mush in any one of the ensemble of solutions, and a domain of chaotic $\widehat{\phi}$ will develop. For example, in Ockendon's problem it should appear in $\tilde{s}_1(t) < x < \tilde{s}_2(t)$ with $\tilde{s}_i(t) \simeq s_i(t)$ where $s_1(t) < x < s_2(t)$ is the mushy zone located by Hills *et al.* (1991). Of course, because of the instability and the jitter, the distribution of $\widehat{\phi}$ within the mushy zone in each realization of the ensemble will be completely dissimilar, and the boundaries, $\tilde{s}_i(t)$, will also be located at slightly different places. After averaging over the ensemble, a mean description of the evolving system will be obtained.

Clearly, even for Ockendon's one-dimensional problem, an immense amount of numerical work is required, and it is natural to seek an alternative. The second avenue referred to above is that of developing a physically well-based mean field theory, from which the average evolution of the system can be derived directly. Such was the approach of Hills *et al.* (1991); such is that applied below. A similar philosophy motivated our theory of coarsening (Hills and Roberts, 1990): the theory of Lifshitz and Slezov (1959) followed the evolution of the distribution function of grain sizes; ours sought only to monitor the average grain size, but it correctly recovered the $t^{1/3}$ law of Lifshitz and Slezov. Of course, something is lost when one abandons the detailed «microscopic» description of the mush studied by Lifshitz and Slezov and by phase-field theory, but something is also gained: a much greater tractibility and generalizability. The historical precedents for the second avenue of approach are legion; for example, to determine a fluid flow one does not generally try to solve Boltzmann's equations!

Although the mean field (or «mixture») theory of a mush is less ambitious than the phase-field theory, since it does not attempt to describe the microscale, it is in one sense far more general; it can readily be developed to include the dynamics of moving phases and mush, something to which (1.6) and (1.7) manifestly cannot be applied. Such indeed is the theory developed by us previously (Hills and Roberts, 1987a, b), which is a refinement of the simple theory on which our solution to Ockendon's problem was based (Hills *et al.*, 1991). In a stationary medium, it leads to (1.6) but only employs the kinetic relaxation process inherent in (1.7), which is modelled by

$$\tau\phi_t + u = 0. \tag{1.9}$$

Thus, the ∇^2 term in (1.7) plays no part, since our averaged ϕ does not vary over intergrain length scales; evidently our τ corresponds to $2\alpha\epsilon^2$ in (1.7) and $\tau \ll 1$ since $\varepsilon \ll 1$. The time delay introduced by τ also leads to blurred fronts between solid and

liquid phases [see Hills and Roberts, (1987a) and below], and, as we shall show, these are amenable to analysis by asymptotic methods. We shall not consider questions of mathematical well-posedness. Our theory resembles the phase-field model in that one can solve for ϕ throughout the system and later interpret the resulting solution in terms of solid ($\phi = 1$), liquid ($\phi = 0$), and mush ($0 < \phi < 1$). This possibility depends on a different treatment of the kinetic term, involving a more sophisticated choice of τ than the naïve constant value assumed in phase-field theory. In § 4 below, we adopt

$$\tau = \frac{\tau_0}{\phi(1-\phi)}, \qquad \tau_0 = \text{constant}. \tag{1.10}$$

This choice is motivated by the obvious physical fact that once the material reaches phase equilibrium, whether it be $\phi = 0$ or $\phi = 1$, it cannot evolve further, that is ϕ cannot pass from the range $0 \leq \phi \leq 1$; moreover, it makes physical sense that the rate, τ^{-1}, at which it approaches equilibrium should decrease as the departure from equilibrium diminishes.

The structure of the solution to our model in the case of small τ parallels that of the phase field model in the case of small ϵ. It appears however to be more amenable to analytic treatment. We are able, for example, to obtain in § 4 a solution to Ockendon's problem that parallels the sharp front solution of Hills *et al.* (1991). This solution works even in cases in which a mush arises. Our solution is complicated by the nonlinearity of (1.10), and in § 2 we present the solution in the analytically simpler case of constant τ. Because of the need to keep ϕ within the range $0 \leq \phi \leq 1$ however, more attention has to be paid to matching the regions in which $\phi = 0$ or 1 to those in which $0 < \phi < 1$, that is the blurred front and the mush. The connection to Hills *et al.* (1991) is closer and simpler than in the case of § 4, and we therefore analyze the case $\tau =$ constant first.

Although our theory, dealing as it does with an averaged ϕ, differs from the phase-field theory in a very basic way, we realize that it may be perceived as an attack on the phase-field model. This was not our intention. Our motive in undertaking the research presented here was to demonstrate that the theory developed by Hills and Roberts (1987a) leads to mathematically and physically robust solutions to a particular problem (that of Ockendon) even in cases where mushy zones exist. It would of course be quite improper to use the successes of either model to discredit the other. Only direct comparison with experiment can be the arbiter, and up to the present time such observational tests are lacking.

2. A CONSTANT MELTING/FREEZING RELAXATION TIME

In this section we consider the Ockendon problem posed in the introduction when the melting/freezing processes have an associated finite relaxation time, τ. In the theory developed by Hills and Roberts (1987a, b) the rate at which the phase conversion takes

place is linearly proportional to the difference in Gibbs energies for the two phases. The authors demonstrated by means of a very simple similarity solution that phase transitions that were «sharp» in the conventional ($\tau = 0$) theory were «blurred» when $\tau \neq 0$. In other words, the transition from pure water to ice occurred within a front whose thickness increased monotonically with τ. The morphology of the front is characterised by the mass fraction $\phi(\mathbf{x}, t)$ of ice at the point $\mathbf{x}$ at time t and is modelled by a mixture theory in which, at each point, there is simultaneously both phases. Since ϕ lies between 0 and 1 within the front, this region was previously designated as a mush but here we prefer the terminology «blurred front» to distinguish it from the extended mushy zones of Hills *et al.* (1991). Thus, within the region $0 \leq x \leq 1$ the solid and liquid phases of H_2O are separated by a thin blurred front of mixed phase that occupies $[s_1(t), s_2(t)]$. The ice lies to the left of this region (smaller x) while on its right there is pure water. Throughout $[0, 1]$ a uniform distribution of heat sources is present that, in the absence of heat conduction and latent heat, would increase the temperature, u, at the rate $\tau > 0$. The difference in the densities of ice and water is ignored so that questions of convection and of volume preservation do not arise. The plane $x = 1$ is thermally isolated while at $x = 0$ the temperature is maintained at $u = -1$. The temperature in the pure phase regions ($1 \equiv$ ice, $2 \equiv$ water) satisfies

$$u_{\alpha t} = u_{\alpha xx} + \tau, \quad \tau = 2/s_c^2 \quad (\alpha = 1, 2), \tag{2.1}$$

subject to

$$u_1(0, t) = -1, \quad u_{2x}(1, t) = 0. \tag{2.2, 3}$$

In the blurred front, denoted by the subscript 3, [see Hills and Roberts (1987a)]

$$u_{3t} = u_{3xx} + \lambda \phi_t + \tau, \tag{2.4}$$

$$\tau \phi_t + u_3 = 0. \tag{2.5}$$

In (2.4), λ denotes the Stefan number (1.1). In what follows we shall assume that $\lambda \gg 1$ and that the dimensionless characteristic time of the melting/freezing process, τ, satisfies $\tau \ll 1$. Within the mixed phase region of thickness $\delta(t)$,

$$\delta(t) = s_2(t) - s_1(t), \tag{2.6}$$

the mass fraction ϕ varies but it is continuous at the edges so that there is no absorption or emission of latent heat there. Thus

$$[[u]] = 0, \quad [[u_x]] = 0, \quad \text{at} \quad x = s_1(t), s_2(t), \tag{2.7, 8}$$

(2.9, 10) $\phi(s_1,t) = 1, \quad \phi(s_2,t) = 0.$

Here $[[\psi]]$ denotes the jump at the front, e.g. $[[\psi]] = \psi_3 - \psi_1$ at $x = s_1$. Finally we make the assumption that the temperature is zero on the leading edge of the front. Thus, if the front moves to the right as a freezing front, we have

(2.11f) $u(s_2,t) = 0, \quad \text{if} \quad \dot{s}_2 > 0,$

while if it moves to the left as a melting front

(2.11m) $u(s_1,t) = 0, \quad \text{if} \quad \dot{s}_1 < 0.$

These leave unsettled what to assume if $\dot{s}_1 > 0$ and $\dot{s}_2 < 0$, and how to deal with the apparent over-determination that arises if $\dot{s}_1 < 0$ and $\dot{s}_2 > 0$. We regard these as bizarre cases not allowed by the theory, although we cannot prove this. Certainly they do not arise below, where $\dot{s}_1$ and $\dot{s}_2$ invariably have the same sign.

Although latent heat is not released or removed at the boundaries of the mixed phase region, equivalent amounts appear or disappear volumetrically throughout its interior as reflected by the term $\lambda\phi_t$ in (2.4). The conditions (2.7)-(2.11) replace the Stefan jump condition of the conventional theory [see Hills *et al.* (1991), equation (2.4)].

Associated with the volumetric heating there is a critical point, $x = s_c$, where

(2.12) $s_c = \sqrt{2/\tau}.$

The case $\tau > 2$ is of greatest interest for then s_c falls within $[0,1]$ and the final steady state has a sharp phase interface at $x = s_\infty$ with the temperature distribution [see Hills *et al.* (1991), § 2]

(2.13) $u_\infty = -(x - s_\infty)(x - \bar{s}_\infty)/s_c^2,$

where

(2.14) $\left.\begin{matrix} s_\infty \\ \bar{s}_\infty \end{matrix}\right\} = 1 \mp (1 - s_c^2)^{1/2},$

and so $s_c > s_\infty$, $s_\infty \leq 1$, $\bar{s}_\infty > 1$.

We shall assume that initially the ice/blurred front boundary is situated at $s_1(0) = s_i$ with the front thickness being $\delta(0) = O(n^{-1})$, where $n^2 = \lambda/\tau \gg 1$. We shall consider only the limit $\lambda \to \infty$. The system then possesses the «fast», $O(1)$, intrinsic time scale associated with thermal diffusion, a fast relaxation time scale $O(\tau)$, and the «slow», $O(\lambda)$, intrinsic scale of the latent heat processes. On the slow scale, heat

conduction (and relaxation) processes are effectively instantaneous. We shall show that across the front the temperature change is $O(n^{-1})$. Then, to leading order, an initial temperature distribution in the pure phase regions quickly evolves on the $O(1)$ time scale to

$$u_1 = -(x - s_1)(x - \tilde{s}_1)/s_c^2, \quad u_2 = (x - s_2)(2 - x - s_2)/s_c^2, \tag{2.15, 16}$$

where $\tilde{s}_1 = s_c^2/s_1$ and thereafter changes in these regions take place on the slow $O(\lambda)$ time scale.

We shall reconsider three situations discussed in Hills *et al.* (1991):

(a) Case $\tau > 2$ and $s_i < s_\infty$ (or $\tau < 2$ and $s_i < 1$).

In the $\tau = 0$ theory, no mushy zone develops in either of these cases. The sharp phase surface, $x = s(t)$, moves slowly towards $x = s_\infty$ as a freezing front; $s(t)$ increases monotonically to $x = 1$ (if $\tau < 2$) or to s_∞ (if $\tau > 2$). The motion is governed by the differential equation

$$\lambda \dot{s} = \Delta(s) \equiv (s - s_\infty)(s - \bar{s}_\infty)/s s_c^2, \tag{2.17}$$

that arises from the Stefan condition. When $\tau \neq 0$, the sharp interface is replaced by a blurred front and from (2.4) and (2.5) we find to leading order that

$$\phi_{3txx} = n^2 \phi_{3t} + 2/\tau s_c^2. \tag{2.18}$$

The general solution that satisfies (2.9) and (2.10) is

$$\phi_3(x,t) = f(x) - \frac{2t}{\lambda s_c^2} + a(t)\frac{\sinh n(s_2 - x)}{\sinh n\delta} + b(t)\frac{\sinh n(x - s_1)}{\sinh n\delta}, \tag{2.19}$$

where

$$a(t) = 1 - f(s_1) + 2t/\lambda s_c^2, \quad b(t) = -f(s_2) + 2t/\lambda s_c^2, \tag{2.20}$$

and $f(x)$ is an arbitrary function that still has to be determined. Since $n\delta = O(1)$, to leading order the temperature distribution follows from (2.5) as

$$u_3 = -\frac{\lambda}{n}\left[a(t)\frac{\cosh n(s_2 - x)}{\sinh n\delta}\dot{s}_2 - b(t)\frac{\cosh n(x - s_1)}{\sinh n\delta}\dot{s}_1\right]. \tag{2.21}$$

From condition (2.8) and the pure phase temperature distributions (2.15), (2.16), we have

$$\text{(2.22, 23)} \qquad \lambda a(t)\dot{s}_2 = \left(s_c^2 - s_1^2\right)/s_1 s_c^2, \quad \lambda b(t)\dot{s}_1 = 2\left(1 - s_2\right)/s_c^2,$$

but since $s_2 - s_1 = O(n^{-1})$ we deduce from these [and (2.20)] that the leading edge s_2 satisfies (2.17) to leading order (as indeed does the trailing edge s_1). Thus the blurred region evolves in much the same way as the sharp interface in the $\tau \neq 0$ case. From (2.20) and (2.23) we have

$$\text{(2.24)} \qquad f\left(s_2\right) = \frac{2t}{\lambda s_c^2} - \frac{2\left(1 - s_2\right)}{s_c^2 \Delta\left(s_2\right)}.$$

Noting that, since s_2 is a monotonically increasing function, we may introduce the inverse function, $T(s_2)$, to s_2, i.e. such that $s_2(T(x)) \equiv x$. Then (2.24) gives

$$\text{(2.25)} \qquad f(x) = \frac{2}{\lambda s_c^2} T(x) - \frac{2\left(1 - x\right)}{s_c^2 \Delta(x)}.$$

Next we turn to condition (2.11f). From (2.21)-(2.23), to leading order, we find

$$\text{(2.26f)} \qquad 2 s_1 \left(1 - s_2\right) \cosh n\delta = s_c^2 - s_1^2,$$

and, given $s_2(t)$ satisfies (2.17), this equation will determine the position of the trailing edge, $s_1(t) = s_2 - \delta$, of the blurred front: since $s_2 - s_1 = O(\delta)$ we have

$$s_1 = s_2 - \frac{1}{n} \cosh^{-1}\left[\frac{s_c^2 - s_2^2}{2 s_2 \left(1 - s_2\right)}\right].$$

We see that $s_1 = s_2 = s_\infty$ is a solution expressing the fact that the blurred front becomes sharp in the final equilibrium state.

Using (2.26f), we see that the temperature field (2.21) can be written, to leading order, as

$$\text{(2.27f)} \qquad u_3 = -\frac{2\left(1 - s_2\right)}{n s_c^2} \sinh n\left(s_2 - x\right).$$

As expected $u_3 < 0$ within the front. It only remains to satisfy (2.7) at $x = s_1$. This can be accomplished by making a small $O(n^{-1})$ correction to the outer solution (2.15).

(b) Case $s_\infty < s_i < s_c$.

In the $\tau = 0$ theory, a sharp phase surface $x = s(t)$ moves to the left as a melting front; $s(t)$ decreases monotonically to s_∞. The analysis of this case parallels that of

(a) above with the labels 1 and 2 of the leading and trailing edges interchanged. In place of (2.26f) we have

$$(2.26m) \qquad \left(s_c^2 - s_1^2\right) \cosh n\delta = 2s_1\left(1 - s_2\right),$$

so that $s_2 = s_1 + n^{-1} \cosh^{-1}[2s_1(1 - s_1)/(s_c^2 - s_1^2)]$. In place of (2.27f)

$$(2.27m) \qquad u_3 = \frac{\left(s_c^2 - s_1^2\right)}{ns_1 s_c^2} \sinh n\left(x - s_1\right).$$

As expected, $u_3 > 0$ within the front. The addition of a small $O(n^{-1})$ constant to u_2 takes care of (2.7) at $x = s_2$.

(c) Case $s_i > s_c$.

In the $\tau = 0$ theory, the initially sharp phase interface moves rapidly to the left until the condition

$$(2.28) \qquad u_{1x}\left(s_1, t\right) = 0$$

is satisfied for the first time. Then a mushy zone forms by the leading edge moving left towards $x = s_c$ with condition (2.28) holding continually as it does so. Indeed, the detailed motion of the leading edge is determined by solving (2.1), (2.2) and (2.28). This is done in Hills *et al.* (1991) using numerical methods. At the beginning the leading edge moves on the fast time scale but, as the critical point $s = s_c$ is approached, the front moves increasingly slowly. For the case when $s_i < 1$, the trailing edge proceeds left on the slow time scale and eventually the mush disappears when the trailing edge overhauls the leading edge in the neighbourhood of the critical point $x = s_c$. Then the last stage of the evolution from $x = s_c$ to $x = s_\infty$ proceeds as for the sharp front described in case (b).

For $\tau \neq 0$ the picture is substantially the same. The sharp front of the first stage of the $\tau = 0$ evolution is replaced by a thin blurred front. The mush begins to evolve only when the condition (2.28) is first satisfied. Since the motion of the ice/mush interface is essentially determined by the ice region contracting so that (2.11m) and (2.28) are satisfied, the leading edge moves as in the $\tau = 0$ case. The point here is that (2.11) and (2.28) are already obeyed at $x = s_1$ by the $\tau = 0$ solution given in Hills *et al.* (1991).

For the blurred front of case (b) we sought a solution in which $n \gg 1$ but $n\delta = O(1)$, but now with a finite mushy zone developing we should properly look for a solution for which $n \gg 1$ but $\delta = O(1)$. We shall therefore treat all terms that are of order $\exp(-n\delta)$ as zero. The mass fraction is again determined by (2.18) and the general solution satisfying the end conditions (2.9) and (2.10) is [cf. (2.19)]

$$(2.29) \qquad \phi_3(x,t) = f(x) - \frac{2t}{\lambda s_c^2} + a(t)e^{-n(x - s_1)} + b(t)e^{-n(s_2 - x)},$$

with $a(t)$ and $b(t)$ being given by (2.20). The temperature distribution follows from (2.5) as [cf. (2.21)]

$$u_3(x,t) = -\frac{\lambda}{n}\left[a(t)e^{-n(x-s_1)}\dot{s}_1 - b(t)e^{-n(s_2-x)}\dot{s}_2\right] + \frac{2}{n^2 s_c^2}, \tag{2.30}$$

where we have omitted $O(n^{-2})$ terms that are exponentially small away from $x = s_1$ or $x = s_2$. To leading order u_3 is exponentially small across the zone and, in particular, condition (2.7) is satisfied to dominant order. From the condition (2.8) at $x = s_1$ we find that $a(t) = 0$ so that, by again introducing the inverse function $T(s_1)$, we have

$$f(x) = 1 + 2T(x)/\lambda s_c^2. \tag{2.31}$$

At the trailing edge the boundary condition (2.8) gives

$$\lambda\left\{1 + \frac{2}{\lambda s_c^2}\left[T(s_2) - t\right]\right\}\dot{s}_2 = -\frac{2(1-s_2)}{s_c^2}, \tag{2.32}$$

which is precisely the equation governing the evolution of the trailing edge in the $\tau = 0$ case.

The final forms for the temperature and mass fraction in the mushy zone become

$$u_3 = -\frac{\lambda}{n}\left\{1 - \frac{2}{\lambda s_c^2}\left[t - T(s_2)\right]\right\}\dot{s}_2 e^{-n(s_2-x)} + \frac{2}{n^2 s_c^2}, \tag{2.33}$$

$$\phi_3 = 1 - \frac{2}{\lambda s_c^2}\left[t - T(x)\right] - \left\{1 - \frac{2}{\lambda s_c^2}\left[t - T(s_2)\right]\right\}e^{-n(s_2-x)}. \tag{2.34}$$

Thus, while the fine detail of these fields are altered for the $\tau \neq 0$ case, the motion of the leading and trailing edges of the mushy zone are substantially the same as for the non-relaxing case.

3. STRUCTURE OF THE GENERAL BLURRED FRONT

Guided by cases (a) and (b) of Section 2, we can now understand better the structure and motion of a blurred front in the general, three-dimensional case. The equations to be solved are [cf. (2.1), (2.4), (2.5), (2.7)-(2.11)]

$$u_{\alpha t} = \nabla^2 u_\alpha + \tau, \quad (\alpha = 1, 2), \tag{3.1}$$

$$u_{3t} = \nabla^2 u_3 + \lambda\phi_t + \tau, \tag{3.2}$$

(3.3) $$\tau\phi_t + u_3 = 0,$$

where (1,2,3) again refer to (ice, water,mush), together with

(3.4, 5) $$[[u]] = 0; \quad [[\mathbf{n}_\alpha \cdot \nabla u]] = 0, \quad \text{on} \quad S_1, S_2,$$

(3.6, 7) $$\phi = 1, \quad \text{on} \quad S_1; \quad \phi = 0, \quad \text{on} \quad S_2,$$

and

(3.8f, m) $$u = 0 \begin{cases} \text{on } S_1, & \text{if } U_1 < 0, \\ \text{on } S_2, & \text{if } U_2 > 0. \end{cases}$$

Here S_1 and S_2 denote the edges of the blurred front adjacent to solid and liquid, $\mathbf{n}_1$ and $\mathbf{n}_2$ are normals to those edges in the direction from solid to liquid, and U_1 and U_2 are the velocities of those edges along the normals.

The analysis hinges on the assumption

(3.9) $$n \to \infty, \quad \delta \to 0, \quad n\delta = O(1).$$

The leading and trailing edges of the front are then close together and are asymptotically coincident in the limit $n \to \infty$. To leading order $\mathbf{n}_1$ and $\mathbf{n}_2$ are the same ($= \mathbf{n}$, say); let it define the $O\xi$ axis of a locally Euclidean coordinate system $O\eta\zeta\xi$ which meets S_α at P_α ($\alpha = 1,2$). Although $\delta(\eta,\zeta)$ is a function of position we may, in the neighbourhood of P_1 and P_2, treat S_1 and S_2 as flat, and δ as a constant, $\delta(0,0)$. In the limit (3.9),

$$\frac{\partial}{\partial\xi} \gg \frac{\partial}{\partial\eta}, \quad \frac{\partial}{\partial\zeta},$$

so that (3.1), (3.2) and (3.5) assume forms analogous to (2.1), (2.4) and (2.8) with $\partial/\partial x$ replaced by $\partial/\partial\xi$, and with S_1 and S_2 given locally by $\xi = s_1(t)$ and $\xi = s_2(t)$.

There is no necessity to follow Section 2 by supposing that the Stefan number is large and that there are two time scales, although we require that τ be small compared with thermal and latent heat time scales. As in Section 2 [see (2.21)] we obtain

(3.10) $$u_3 = -\frac{\lambda}{n}\left[a(t)\frac{\cosh n(s_2 - \xi)}{\sinh n\delta}U_2 - b(t)\frac{\cosh n(\xi - s_1)}{\sinh n\delta}U_1\right],$$

from which

(3.11, 12) $$(\mathbf{n}\cdot\nabla u_3)_{S_1} = \lambda a U_2, \quad (\mathbf{n}\cdot\nabla u_3)_{S_2} = \lambda b U_1.$$

Since $U_1 = U_2$ to leading order (= U, say), we see from (2.20) that

$$[\mathbf{n} \cdot \nabla u_3]_{S_2}^{S_1} = \lambda U. \tag{3.13}$$

By (3.5) this is

$$\lambda U = [\mathbf{n} \cdot \nabla u_1]_{S_1} - [\mathbf{n} \cdot \nabla u_2]_{S_2}, \tag{3.14}$$

which is essentially the Stefan jump condition, which holds even though latent heat is released or absorbed volumetrically throughout the blurred front rather than at a sharp surface as when $\tau = 0$, and $s_i < s_c$ or $\tau < 2$.

As in cases (2.27f) or (2.27m) there is an $O(n^{-1})$ discontinuity in u at the trailing edge, whose removal requires a tiny adjustment to the solution behind the trailing edge.

4. A MORE GENERAL RELAXATION LAW

So far we have, as in Hills and Roberts (1987b), concentrated on the simplest of all possible relaxation laws, in which the relaxation time, τ, is constant if ϕ lies within (0,1) but which is infinite if $\phi = 0$ or $\phi = 1$. Reality is perhaps better modelled by a law such as

$$\tau = \tau_0 / \phi(1 - \phi), \tag{4.1}$$

which holds for all ϕ, including 0 and 1. The disadvantage of such a law is, of course, the greater analytic complexity it creates. Its advantage is first that there are no sharp leading or trailing edges to a blurred front. The front can be located by a single surface, e.g. that on which $u = 0$. Such a front resembles more an internal boundary layer which smoothly matches to the surrounding liquid and solid. Second (and because of this), assumptions such as (2.11) and (3.8) become redundant; ambiguities, such as the one arising below (2.11) and at corresponding points in the argument of Section 3, do not arise.

In the limit (3.9) we now have [cf. (2.4) and (2.5)]

$$u_t = u_{\xi\xi} + \lambda \phi_t + r, \tag{4.2}$$

$$\tau \phi_t + \phi(1 - \phi) u = 0, \tag{4.3}$$

where we have omitted the subscript 0 from τ_0. Introduce a new variable v to replace ϕ:

$$\phi = \left(1 + e^{v/\tau}\right)^{-1}. \tag{4.4}$$

By (4.3) we now have

$$u = v_t, \tag{4.5}$$

so that by (4.2)

$$v_{t\xi\xi} = \frac{\lambda}{\tau}\frac{e^{v/\tau}}{\left(1+e^{v/\tau}\right)^2}\, v_t + v_{tt} - r. \tag{4.6}$$

Integration of (4.6) gives

$$v_{\xi\xi} = -\frac{\lambda}{1+e^{v/\tau}} + v_t - rt + \lambda f(\xi), \tag{4.7}$$

where $f(\xi)$ is at this stage arbitrary. In the limit $\tau \to 0$, (4.7) becomes

$$v_{\xi\xi} = v_t - rt + \lambda f(\xi), \quad \text{if} \quad v > 0, \tag{4.8L}$$

$$v_{\xi\xi} = v_t - rt + \lambda f(\xi) - \lambda, \quad \text{if} \quad v < 0, \tag{4.8S}$$

corresponding to pure liquid and solid respectively, as (4.4) shows. If $v = O(\tau)$ as $\tau \to 0$ however, ϕ takes values between 0 and 1, corresponding to states within a blurred front, and (4.7) must be used instead of (4.8).

In the limit $n^2 \equiv \lambda/\tau \gg 1$ of large Stefan number, the left-hand side of (4.2) is small. Then v_{tt} may be omitted from (4.6), and v_t from (4.7) and (4.8). Moreover, if a zero of v exists in the domain [at $\xi = s(t)$ say], i.e. if a front exists, its thickness is of order n^{-1} so that $f(\xi)$ is approximately $f_s(t) \equiv f(s(t))$. We may then integrate (4.7) to obtain

$$\frac{1}{2\lambda\tau}\left(\frac{\mathrm{d}v}{\mathrm{d}\xi}\right)^2 = \ln\left(e^{-v/\tau}+1\right) + \left[f_s - \frac{rt}{\lambda}\right]\frac{v}{\tau} + g, \tag{4.9}$$

where $g(t)$ is an arbitrary function of integration. We now have

$$\xi - s(t) = \frac{1}{2^{1/2}n}\int_0^{v/\tau} \frac{\mathrm{d}w}{\left[\ln\left(e^{-w}+1\right) + \left(f_s - rt/\lambda\right)w + g\right]^{1/2}}, \tag{4.10}$$

where $s(t)$ is (see above) the location of the zero of v [note: since v increases with ξ, we took the positive square root in (4.10)].

By differentiating (4.10) with respect to t we obtain approximately

$$-\dot{s}(t) = \frac{vt}{(2\lambda\tau)^{1/2}}\frac{1}{\left[\ln\left(e^{-v/\tau}+1\right) + \left(f_s - rt/\lambda\right)(v/\tau) + g\right]^{1/2}},$$

or, by (4.9),

$$v_t + \dot{s} v_\xi \approx 0 . \tag{4.11}$$

Since $v = 0$ on $\xi = s$, $D_t v = 0$ thereon, where D_t denotes differentiation with respect to the frame moving with the front at the point concerned. This confirms (4.11).

By (4.5) and the ξ-derivative of (4.11), we may re-interpret (4.8) as

$$u_\xi = -\dot{s}\left[\lambda f_s(t) - rt\right], \quad \text{if} \quad v > 0, \tag{4.12L}$$

$$u_\xi = -\dot{s}\left[\lambda f_s(t) - rt - \lambda\right], \quad \text{if} \quad v < 0. \tag{4.12S}$$

This shows how f_s may be determined from the gradient of u on either side of the blurred front. Stefan's law is recovered by subtraction.

ACKNOWLEDGEMENT

The research of one of the authors (P.H.R.) is sponsored by the U.S. Office of Naval Research under contract N00014-86-K-0691 with the University of California, Los Angeles.

REFERENCES

Atthey, D.R., 1974, *A finite difference scheme for melting problems*, J. Inst. Maths. Applics. **30**, 353.

Binder, K., 1987, *Theory of first-order phase transitions*, Rep. Progr. Phys. **50**, 738.

Caginalp, G., Fife, P.C., 1987, *Elliptic problems involving phase boundaries satisfying a curvature condition*, IMA J. Appl. Math. **38**, 195.

Fasano, A., Primicerio, M., 1985, *Phase-change with volumetric heat sources: Stefan's scheme vs. enthalpy formulation*, Suppl. Boll. UMI (IV) **5**, 131.

Hills, R.N., Roberts, P.H., 1991, *On the life cycle of a mush*, (with appendix by A.M. Soward), Stability Appl. Anal. Cts. Media **1**, 1.

Hills, R.N., Roberts, P.H., 1987a, *Relaxation effects in a mixed phase region: I. General theory*, J. Non Equilib. Thermodyn. **12**, 169.

Hills, R.N., Roberts, P.H., 1987b, *Relaxation effects in a mixed phase region: II. Illustrative examples*, J. Non Equilib. Thermodyn. **12**, 183.

Hohenberg, P.C., Halperin, B.I., 1977, *Theory of dynamic critical phenomena*, Rev. Modern Phys. **49**, 435.

Langer, J.S., 1971, *Theory of spinodal decomposition in alloys*, Ann. Phys. **65**, 53.

Lifshitz, I.M., Slezov, V.V., 1959, *Kinetics of diffusive decomposition of supersaturated solid solutions*, Sov. Phys. JETP **35**, 331.

Meirmanov, A., 1981, *An example of nonexistence of a classical solution of the Stefan problem*, Sov. Math. Dokl. **23**, 564.

Ockendon, J.R., 1975, *Techniques of analysis in moving boundary problems*, in: Heat, Flow and Diffusion (eds. J.R. Ockendon & W.R. Hodgkins), Clarendon Press, Oxford.

Oleinik, O.A., 1960, *A method of solution of the general Stefan problem*, Sov. Math. Dokl. **1**, 1350.

Penrose, O., Fife, P.C., 1990, *Thermodynamically consistent models of phase-field type for the kinetics of phase transitions*, Physica D **43**, 44.

Primicerio, M., 1982, *Mushy regions in phase-change problems*, in: Applied Nonlinear Functional Analysis (eds. Gorenflo & Hoffmann), Lang, Frankfurt.

Rubenstein, L.I., 1971, *The Stefan problem*, Am. Math. Soc. Trans. **27**, American Mathematical Society, Providence.

Manuscript received: July 14, 1992.

PUBLICATION LIST

1. P.H. Roberts, *On the reflection and refraction of hydromagnetic waves.* Astrophys. J., **121**, 1955, 720-730.
2. P.H. Roberts, *The equilibrium of magnetic stars.* Astrophys. J., **122**, 1955, 508-512.
3. P.H. Roberts, *Hydromagnetic disturbances in a fluid of finite conductivity.* Astrophys. J., **122**, 1955, 315-326.
4. J.W. Chamberlain, P.H. Roberts, *Turbulence spectrum in Chandrasekhar's theory.* Phys. Rev., **99**, 1955, 1674-1677.
5. P.H. Roberts, *Twisted magnetic fields.* Astrophys. J., **124**, 1956, 430-442.
6. P.H. Roberts, *How old is the Earth?* Streven, **9**, 1956, 512-520.
7. P.H. Roberts, *The Earth's magnetism.* Streven, **9**, 1956, 898-905.
8. P.H. Roberts, *Hydromagnetic disturbances in a fluid of finite conductivity, II.* Astrophys. J., **126**, 1957, 418-428.
9. P.H. Roberts, *Why a geophysical year?* Streven, **10**, 1957, 850- 855.
10. P.H. Roberts, *On the application of a statistical approximation to the theory of turbulent diffusion.* J. Rat. Mech. & Anal., **6**, 1957, 781-800.
11. F.C. Bertiau, P.H. Roberts, *Distribution of gravitational field at the center of globular and spherical clusters.* Astrophys. J., **128**, 1958, 130-138.
12. P.H. Roberts, *Propagation of induced fields through the core.* Ann. Geophys., **XV**, 1959, 75-86.
13. P.H. Roberts, H.D. Ursell, *Random walk on a sphere and on a Riemannian manifold.* Phil. Trans. R. Soc. Lond., **A252**, 1960, 317-356.
14. P.H. Roberts, *Characteristic value problems posed by differential equations arising in hydrodynamics and hydromagnetics.* J. Math. Anal. and Applic., **1**, 1960, 195-214.
15. R. Hide, P.H. Roberts, *Hydromagnetic flow due to an oscillating plane.* Rev. Mod. Phys., **32**, 1960, 799-806.
16. P.H. Roberts, T. Tatsumi, *The decay of magnetohydrodynamic turbulence.* J. Math. & Mech., **9**, 1960, 697-714.
17. P.H. Roberts, *Analytical theory of turbulent diffusion.* Courant Institute Rep., **HSN-2**, 1960.
18. P.H. Roberts, *Analytical theory of turbulent diffusion.* J. Fluid Mech., **11**, 1960, 257-283.
19. R. Hide, P.H. Roberts, *The origin of the main geomagnetic field.* Phys. and Chem. of the Earth **4**, Pergamon Press, 1961, 25-98.
20. P.H. Roberts, F.J. Lowes, *Earth currents of deep internal origin.* J. Geophys. Res., **66**, 1961, 1243-1254.
21. P.H. Roberts, *On Bourret's hypothesis concerning turbulent diffusion.* Can. J. Phys., **39**, 1961, 1291-1299.
22. P.H. Roberts, A.D. Boardman, *The effect of a vertical magnetic field on the propagation of gravity waves along the plane surface of a semi- infinite viscous, electrically conducting fluid.* Astrophys. J., **135**, 1962, 552-592.
23. T. Chamalaun, P.H. Roberts, *The theory of convection in spherical shells and its application to the problem of thermal convection in the Earth's mantle.* Continental Drift. International Geophysics Series **3**. Academic Press, 1962, 177-194.
24. P.H. Roberts, *On the superpotential and supermatrix of a heterogeneous ellipsoid.* Astrophys. J., **136**, 1962, 1108-1114.
25. R. Hide, P.H. Roberts, *Some elemementary problems in magneto- hydrodynamics.* Adv. in Appl. Mech., **4**, Academic Press, 1962, 215-316.
26. P.H. Roberts, *The effect of a vertical magnetic field on Rayleigh- Taylor instability.* Astrophys. J., **137**, 1963, 679-689.
27. P.H. Roberts, K. Stewartson, *On the stability of a Maclaurin spheroid of small viscosity.* Astrophys. J., **137**, 1963, 777-790.
28. P.H. Roberts, *On highly rotating polytropes I.* Astrophys. J., **137**, 1963, 1129-1141.
29. K. Stewartson, P.H. Roberts, *On the motion of a liquid in a spheroidal cavity of a precessing rigid body.* J. Fluid Mech., **17**, 1963, 1-20.
30. P.H. Roberts, *On highly rotating polytropes II.* Astrophys. J., **138**, 1963, 809-819.
31. S. Chandrasekhar, P.H. Roberts, *The ellipticity of a slowly rotating configuration.* Astrophys. J., **138**, 1963, 801-808.

32. P.H. Roberts, S. Scott, *Truncation errors in the spherical harmonic analysis of the geomagnetic field and the problem of downward extrapolation.* J. Geomag. Geoelec., **XV**, 1964, 148-160.
33. P.H. Roberts, *The stability of hydromagnetic Couette flow.* Proc. Camb. Phil. Soc., **60**, 1964, 635-651.
34. M. Hurley, P.H. Roberts, *On highly rotating polytropes III.* Astrophys. J., **140**, 1964, 583-598.
35. P.H. Roberts, *The figures of the Earth and other rotating bodies.* University of Newcastle upon Tyne, 1964.
36. P.H. Roberts, K. Stewartson, *On the motion of a liquid in a spheroidal cavity of a precessing rigid body, II.* Proc. Camb. Phil. Soc., **61**, 1965, 279-288.
37. R.J. Donnelly, K.W. Schwartz, P.H. Roberts, *Experiments on the stability of viscous flow between rotating cylinders, VI: Finite-amplitude experiments.* Proc. R. Soc. Lond., A**283**, 1965, 531-546.
38. M. Hurley, P.H. Roberts, *On highly rotating polytropes IV.* Astrophys. J. Suppl., **XI**, 1965, 95-119.
39. P.H. Roberts, *On the thermal instability of a highly rotating fluid sphere.* Astrophys. J., **141**, 1965, 240-250.
40. D.N. Limber, P.H. Roberts, *On highly rotating polytropes V.* Astrophys. J., **141**, 1965, 1439-1442.
41. P.H. Roberts, S. Scott, *On analysis of the secular variation, 1. A hydromagnetic constraint: Theory.* J. Geomag. Geoelec., **17**, 1965, 137-151.
42. P.H. Roberts, *Convection in a self-gravitating fluid sphere.* Mathematika, **12**, 1965, 128-136.
43. M. Hurley, P.H. Roberts, K. Wright, *The oscillations of gas spheres.* Astrophys. J., **143**, 1966, 535-551.
44. P.H. Roberts, *On the relationship between Galerkin's method and the local potential method.* Non-Equilibrium Thermodynamics, Variational Techniques and Stability. University of Chicago Press, 1966, 299-302.
45. P.H. Roberts, *On non-linear Bénard convection.* Non-Equilibrium Thermodynamics, Variational Techniques & Stability. University of Chicago Press, 1966, 125-157.
46. P.H. Roberts, *Frozen fields, core motions and the secular variation.* Magnetism and the Cosmos. Oliver and Boyd, 1966, 95-107.
47. R.D. Gibson, P.H. Roberts, *Some comments on the theory of homogeneous dynamos.* Magnetism and the Cosmos. Oliver and Boyd, 1966, 108-120.
48. P.H. Roberts, *Magnetism and the Cosmos.* Edited jointly with W. Hindmarsh, F.J. Lowes and S.K. Runcorn. Oliver and Boyd, 1966.
49. P.H. Roberts, *An Introduction to Magnetohydrodynamics.* Longmans, Green and Co., 1967.
50. P.H. Roberts, *Singularities of Hartmann Layers.* Proc. R. Soc. Lond., A**300**, 1967, 94-107.
51. P.H. Roberts, *Geomagnetic dynamos.* Proceedings of Woods Hole Geophysical Summer School, **1**, 1967.
52. P.H. Roberts, *Convection in horizontal layers with internal heat generation: Theory.* J. Fluid Mech., **30**, 1967, 33-49.
53. P.H. Roberts, *On the thermal instability of a rotating-fluid sphere containing heat sources.* Phil. Trans. R. Soc. Lond., A**263**, 1968, 93-117.
54. J.G. Tough, P.H. Roberts, *Nearly symmetric hydromagnetic dynamos.* Phys. Earth Planet. Int., **1**, 1968, 288-296.
55. W.I. Glaberson, R.J. Donnelly, P.H. Roberts, *Hydromagnetic duct flow: Theory and experiment.* Phys. Fluids, **11**, 1968, 2192-2199.
56. R.D. Gibson, P.H. Roberts, *The Bullard-Gellman dynamo.* Appl. Mod. Phys. to Earth & Planet. Inter., Wiley, 1969, 577-602.
57. P.H. Roberts, *Electrohydrodynamic convection.* Quart. J. Mech. & Appl. Math., **22**, 1969, 211-220.
58. R.J. Donnelly, P.H. Roberts, *Stochastic theory of the interaction of ions and quantized vortices in helium II.* Proc. R. Soc. Lond. A., **312**, 1969, 519-551.
59. R.J. Donnelly, P.H. Roberts, *Roton-phonon relaxation times in helium II.* Phys. Lett. A., **30**, 1969, 468-469.
60. R.J. Donnelly, P.H. Roberts, *Nucleation of quantized vortex rings by ions in helium II.* Phys. Rev. Lett., **23**, 1969, 1491-1495.
61. P.H. Roberts, R.J. Donnelly, *Theory of the onset of superflow.* Phys. Rev. Lett., **24**, 1970, 367-371.
62. P.H. Roberts, R.J. Donnelly, *Dynamics of vortex rings.* Phys. Lett., **31**, 1970, 137-138.
63. P. Baldwin, P.H. Roberts, *The critical layer in stratified shear flow.* Mathematika, **17**, 1970, 102-119.
64. A.E. Cook, P.H. Roberts, *The Rikitake two-disc dynamo system.* Proc. Camb. Phil. Soc., **68**, 1970, 547-569.
65. I.A. Eltayeb, P.H. Roberts, *On the hydromagnetics of rotating fluids.* Astrophys. J., **162**, 1970, 699-701.

66. P.H. Roberts, *Thomas Henry Havelock.* Bull. Lond. Math. Soc., **2**, 1970, 221-232.
67. R.J. Donnelly, P.H. Roberts, *Stochastic theory of the nucleation of quantized vortices in superfluid helium.* Phil. Trans. R. Soc. Lond., A**271**, 1971, 41-100.
68. P.H. Roberts, J. Grant, *Motions in a Bose condensate. I: The structure of a large circular vortex.* J. Phys., **4**, 1971, 55-72.
69. D.M. Strayer, R.J. Donnelly, P.H. Roberts, *Motion of ions at finite velocities in helium II.* Phys. Rev. Lett., **26**, 1971, 165-169.
70. P.H. Roberts, *Dynamo theory of geomagnetism.* World Magnetic Survey 1957-1969, IUGG Publ., 1971, 123-131.
71. P.H. Roberts, M. Stix, *The turbulent dynamo. A translation of a series of papers by F. Krause, K.-H. Rädler and M. Steenbeck.* NCAR Tech. Note TN/1A-60, 1971.
72. P.H. Roberts, *A transformation of the stellar wind equations.* Astrophys. Lett., **9**, 1971, 79-80.
73. A.M. Binnie, P.H. Roberts, *Thomas Henry Havelock 1877-1968.* Biograph. Not. R. Soc. Lond., **17**, 1971, 327-377.
74. P.H. Roberts, *Dynamo theory.* Lectures in Appl. Math., Amer. Math. Soc., **14**, 1971, 129-206.
75. B. Durney, P.H. Roberts, *On the theory of stellar winds.* Astrophys. J., **170**, 1971, 319-323.
76. P.H. Roberts, A.M. Soward, *Stellar winds and breezes.* Proc. R. Soc. Lond., A**328**, 1972, 185-215.
77. P.H. Roberts, M. Stix, *α-effect dynamos by the Bullard-Gellman formalism.* Astron. & Astrophys., **18**, 1972, 453-466.
78. P.H. Roberts, *Kinematic dynamo models.* Phil. Trans. R. Soc. Lond., A**272**, 1972, 663-703.
79. P.H. Roberts, A.M. Soward, *Magnetohydrodynamics of the Earth's core.* Ann. Rev. Fluid Mech., **4**, 1972, 117-154.
80. P. Baldwin, P.H. Roberts, *On resistive instabilities.* Phil. Trans. R. Soc. Lond., A**272**, 1972, 303-330.
81. R.N. Hills, P.H. Roberts, *On Landau's two-fluid model for helium II.* J. Inst. Math. Applics., **9**, 1972, 56-67.
82. P.H. Roberts, *Electromagnetic core-mantle coupling.* J. Geomag. Geoelec., **24**, 1972, 231-259.
83. P.H. Roberts, R.J. Donnelly, *Dynamics of rotons.* Phys. Lett., **43A**, 1973, 1-2.
84. R.J. Donnelly, P.H. Roberts, *An analogy between the theory of superfluidity and the theory of dielectrics.* Phys. Lett., **43A**, 1973, 199-200.
85. P.H. Roberts, *A Hamiltonian theory for weakly interacting vortices.* Mathematika, **19**, 1972, 169-179.
86. F. Krause, P.H. Roberts, *Some problems of mean field electrodynamics.* Astrophys. J., **181**, 1973, 977-992.
87. P.H. Roberts, R.J. Donnelly, *Superfluid mechanics.* Ann. Rev. Fluid Mech., **6**, 1974, 179-225.
88. F. Krause, P.H. Roberts, *Bochner's theorem and mean-field electrodynamics.* Mathematika, **20**, 1973, 24-33.
89. F. Krause, P.H. Roberts, *Comments on the paper: "On the application of Cramer's Theorem to axisymmetric in compressible turbulence", by I. Lerche.* Astrophys. & Space Sci., **22**, 1973, 193-195.
90. P.H. Roberts, *Dynamo effect in mirror-symmetric turbulence in the Kraichnan direct interaction approximation.* Magnetnaya gidrodinamika, **2**, 1975, 3-8.
91. P.H. Roberts, R.J. Donnelly, *Scattering and binding of rotons.* J. Low Temp. Phys., **15**, 1974, 1-27.
92. J. Grant, P.H. Roberts, *Motions in a Bose condensate III. The structure and effective masses of charged and uncharged impurities.* J. Phys. A., **7**, 1974, 260-279.
93. P.H. Roberts, W.J. Pardee, *Bound states of the two-roton Schrödinger equation.* J. Phys. A., **7**, 1974, 1283-1292.
94. P.H. Roberts, *Concepts in hydrodynamic stability theory.* Advan. Chem. Phys., **32**, 1975, 17-35.
95. P.H. Roberts, K. Stewartson, *On finite amplitude convection in a rotating magnetic system.* Phil. Trans. R. Soc. Lond., A**277**, 1974, 287-315.
96. P.H. Roberts, *Magnetic core-mantle coupling by inertial waves.* Phys. Earth Planet. Int., **8**, 1974, 389-390.
97. P.H. Roberts, A.M. Soward, *A unified approach to mean field electrodynamics.* Astron. Nachr., **296**, 1975, 49-64.
98. S. Kumar, P.H. Roberts, *A three-dimensional kinematic dynamo.* Proc. R. Soc. Lond., A**344**, 1975, 235-258.
99. P.H. Roberts, K. Stewartson, *On double-roll convection in a rotating magnetic system.* J. Fluid Mech., **68**, 1975, 447-466.
100. P.H. Roberts, A.M. Soward, *On first order smoothing theory.* J. Math. Phys., **16**, 1975, 609-615.

101. P.H. Roberts, *Geophysical magnetohydrodynamics.* Geodynamics Today; A review of the Earth's dynamic processes. Publ. of R. Soc. Lond., 1975, 133-140.
102. P.H. Roberts, *Stellar winds.* Solar Wind 3, Ed. C.T. Russell. Inst. Geophys. & Planet. Phys. Univ. of California, 1976, 231-242.
103. D.L. Goodstein, J.S. Brooks, R.J. Donnelly, P.H. Roberts, *Quasi- particles and thermal expansion.* Phys. Lett., **54A**, 1975, 281-282.
104. S.I. Braginsky, P.H. Roberts, *Magnetic field generation by baroclinic waves.* Proc. R. Soc. Lond., A**347**, 1975, 125-140.
105. A.M. Soward, P.H. Roberts, *Recent developments in dynamo theory.* Magnitnaya gidrodinamika, #1, 1976, 3-51. Corrigendum #1, 1977, 146-147.
106. P.H. Roberts, R.J. Donnelly, *A simple theory of temperature dependent energy levels: application to He II.* Phys. Lett., **55A**, 1976, 443-445.
107. P.H. Roberts, M. Stix, *On Vainshtein's simplest dynamo instability.* Phys. Earth Planet. Int., **12**, 1976, 19-21.
108. F. Krause, P.H. Roberts, *The high conductivity limit in mean field electrodynamics.* J. Math. Phys., **17**, 1976, 1808-1809.
109. R.J. Donnelly, P.H. Roberts, *A theory of temperature-dependent energy levels: Thermodynamic properties of He II.* J. Low Temp. Phys., **27**, 1977, 687-736.
110. R.N. Hills, P.H. Roberts, *Healing and relaxation in flows of Helium II. I: Generalization of Landau's equations.* Int. J. Engng. Sci., **15**, 1977, 305-316.
111. R.N. Hills, P.H. Roberts, *Superfluid mechanics for a high density of vortex lines.* Archiv. Rat. Mech. Anal., **66**, 1977, 43-71.
112. D.E. Loper, P.H. Roberts, *On the motion of an iron-alloy core containing a slurry. I: General theory.* Geophys. & Astrophys. Fluid Dynam., **9**, 1978, 289-321.
113. P.H. Roberts, K. Stewartson, *The effect of finite electrical and thermal conductivities on magnetic buoyancy in a rotating gas.* Astron. Nachr., **298**, 1977, 311-318.
114. P.H. Roberts, *Magnetohydrodynamics.* Invited Survey Article in Encyclopedia of Physics, Eds. R.G. Lerner & G.L. Trigg. Addison-Wesley Co., 1981, 555-561.
115. P.H. Roberts, *The electron bubble in liquid helium.* Proc. Int. Coll. on Drops and Bubbles, Their Science and the Systems they Model, **I**. JPL publication, 1974.
116. P.H. Roberts, *Diffusive instabilities in magnetohydrodynamic convection.* Instabilities Hydrodynamiques en Convection Libre, Forcée et Mixte, Eds. J. Cl. Legros & J.K. Platten. Springer, 1977, 27-29.
117. R.N. Hills, P.H. Roberts, *Healing and relaxation in flows of Helium II. II: First, second and fourth sound.* J. Low Temp. Phys., **30**, 1978, 709-727.
118. P.H. Roberts, R.W. Walden, R.J. Donnelly, *Bound states of rotons to impurities in He II.* J. Low Temp. Phys., **31**, 1978, 389-408.
119. R.J. Donnelly, R.W. Walden, P.H. Roberts, *Dielectric model of roton interactions in dilute solutions of* ^3He *in* ^4He. J. Low Temp. Phys., **31**, 1978, 375-387.
120. R.J. Donnelly, P.H. Roberts, *Need for more precise thermodynamic & neutron scattering data on liquid helium.* J. Phys. C., **10**, 1977, 683-685.
121. P.H. Roberts, D.E. Loper, *On the diffusive instability of some simple steady magnetohydrodynamic flows.* J. Fluid Mech., **90**, 1979, 641-668.
122. P.H. Roberts, R.N. Hills, R.J. Donnelly, *Calculation of the static healing length in helium II.* Phys. Lett., **70A**, 1979, 437-440.
123. P.H. Roberts, R.N. Hills, *On the interface conditions between a liquid and its vapour with application to helium II.* Non-Equilib. Thermodyn, **4**, 1979, 131-142.
124. P.H. Roberts, *MAC waves.* Geofluiddynamical Wave Mechanics Research, Ed. W. Criminale. Publ. of Appl. Math. Dept. Univ. of Wash., Seattle, 1977, 218-225.
125. R.N. Hills, P.H. Roberts, *Healing and relaxation in flows of helium II. III: Pure superflow.* J. Phys. C., **11**, 1978, 4485-4499.
126. R.J. Donnelly, R.N. Hills, P.H. Roberts, *Superflow in restricted geometries.* Phys. Rev. Lttr., **42**, 1979, 725-728.
127. P.H. Roberts, *Magneto-convection in a rapidly rotating fluid* in Rotating Fluids in Geophysics, Eds. P.H. Roberts, A.M. Soward. Academic Press, 1978, 421-435.

128. C.A. Jones, P.H. Roberts, *The boundary layer method for pulsating stars.* Geophys. & Astrophys. Fluid Dynam., **14**, 1979, 61-101.
129. R. Hide, P.H. Roberts, *How strong is the magnetic field in the Earth's liquid core?* Phys. Earth Planet. Int., **20**, 1979, 124-126.
130. R.N. Hills, P.H. Roberts, *Healing and relaxation in flows of Helium II. IV: Comparison with Khalatnikov-Lebedev theory.* J. Low Temp. Phys., **40**, 1980, 117-134.
131. D.E. Loper, P.H. Roberts, *Are planetary dynamos driven by gravitational settling?* Phys. Earth Planet. Int., **20**, 1979, 192-193.
132. R.W. James, P.H. Roberts, D.E. Winch, *The Cowling anti-dynamo theorem.* Geophys. & Astrophys. Fluid Dynam., **15**, 1980, 149-160.
133. D.E. Loper, P.H. Roberts, *On the motion of an iron-alloy core containing a slurry. II. A simple model.* Geophys. & Astrophys. Fluid Dynam., **16**, 1980, 83-127.
134. D.E. Loper, P.H. Roberts, *A study of conditions at the inner core boundary of the Earth.* Phys. Earth Planet. Int., **24**, 1981, 302-307.
135. P.H. Roberts, *On the Flowers-Ruderman instability mechanism.* Astron. Nachr., **302**, 1981, 65-70.
136. F. Krause, P.H. Roberts, *Strange attractor character of large-scale non-linear dynamos.* Adv. Space Res., **1**, 1981, 231-240.
137. P.H. Roberts, *Equilibria & stability of a fluid type II superconductor.* Quart. J. Mech. & Appl. Math., **34**, 1981, 327-343.
138. D.R. Fearn, D.E. Loper, P.H. Roberts, *Structure of the Earth's inner core.* Nature, **292**, 1981, 232-233.
139. D.E. Loper, P.H. Roberts, *Compositional convection and the gravitationally powered dynamo.* Stellar & Planetary Magnetism, Ed. A.M. Soward. Gordon & Breach, 1983, 297-327.
140. P.H. Roberts, D.E. Loper, *Towards a theory of the structure and evolution of a dendrite layer.* Stellar and Planetary Magnetism, Ed. A. M. Soward. Gordon & Breach, 1983, 329-349.
141. C.A. Jones, P.H. Roberts, *Motions in a Bose condensate. IV. Axisymmetric solitary waves.* J. Phys. A, **15**, 1982, 2599-2619.
142. S.J. Putterman, P.H. Roberts, *Non-linear hydrodynamics and a one fluid theory of superfluid* He^4. Phys. Lett., **89A**, 1982, 444-447.
143. S.J. Putterman, P.H. Roberts, *Classical non-linear waves in dispersive nonlocal media, and the theory of superfluidity.* Physica, **117A**, 1983, 369-397.
144. P.H. Roberts, *Random non-linear electromagnetic waves in vacuo.* Physics, **119A**, 1983, 274-290.
145. P.H. Roberts, A.M. Soward, *The effect of a weak magnetic field on stellar oscillations.* Mon. Not. R. Astr. Soc., **205**, 1983, 1171-1189.
146. R.N. Hills, D.E. Loper, P.H. Roberts, *A thermodynamically consistent model of a mushy zone.* Quart. J. Mech. & Appl. Math., **36**, 1983, 505-539.
147. P.H. Roberts, D.E. Winch, *On random rotations.* Adv. Appl. Prob., **16**, 1984, 638-655.
148. M. Stix, P.H. Roberts, *Time-dependent electromagnetic core-mantle coupling.* Phys. Earth Planet. Int., **36**, 1984, 49-60.
149. P.H. Roberts, *The effect of a weak magnetic field on stellar oscillations.* Astron. Nach., **305**, 1984, 245-249.
150. S.J. Putterman, P.H. Roberts, *Random waves in classical nonlinear Grassmann field. I: Boltzmann equation and scalar field coupling.* Physica, **131A**, 1985, 35-50.
151. S.J. Putterman, P.H. Roberts, *Random waves in a classical nonlinear Grassmann field. II: Scattering by zero point noise and the transition to Fermi statistics.* Physica, **131A**, 1985, 51-63.
152. A. Larraza, S.J. Putterman, P.H. Roberts, *A universal l/f power spectrum as the accumulation point of wave turbulance.* Phys. Rev. Lettr., **55**, 1985, 897-900.
153. C.A. Jones, S.J. Putterman, P.H. Roberts, *Motions in a Bose condensate. V: Stability of solitary wave solutions of the non-linear Schrödinger equations in two and three dimensions.* J. Phys. A., **19**, 1986, 2991-3011.
154. P.H. Roberts, *Dynamo theory.* Proc. NATO Advance Study Institute on Irreversible Phenomena and Dynamical Systems Analysis in Geosciences, Eds. C. Nicolis and G. Nicolis. Reidel, 1987, 73-133.
155. P.H. Roberts, *Convection in spherical systems.* Proc. NATO Advanced Study Institute on Irreversible Phenomena and Dynamical Systems Analysis in Geosciences, Eds., C. Nicolis and G. Nicolis. Reidel, 1987, 53-71.
156. P.H. Roberts, *Rotating Fluids in Geophysics.* Edited jointly with A.M. Soward. Academic Press, 1978.

157. C.A. Jones, P.H. Roberts, *What happens to vortex rings that shrink?*, 75th Jubilee Conference on Helium-4, Ed. J.G.M. Armitage. World Scientific Press, 1983, 113-114.
158. R.N. Hills, P.H. Roberts, *Relaxation effects in a mixed phase region. I: General theory.* J. Nonequilib. Thermodynam., **12**, 1987, 169-181.
159. R.N. Hills, P.H. Roberts, *Relaxation effects in a mixed phase region. II: Illustrative Examples.* J. Nonequilib. Thermodynam., **12**, 1987, 183-195.
160. D. Gubbins, P.H. Roberts, *Magnetohydrodynamics of the Earth's core.* Geomagnetism **2**, Ed. J.A. Jacobs. Academic Press, 1987, 1-183.
161. P.H. Roberts, D. Gubbins, *Origin of the main field: Kinematics.* Geomagnetism **2**, Ed. J.A. Jacobs. Academic Press, 1987, 185-249.
162. P.H. Roberts, *Origin of the main field: Dynamics.* Geomagnetism **2**, Ed. J.A. Jacobs. Academic Press, 1987, 251-306.
163. P.H. Roberts, D.E. Loper, *Dynamical processes in slurries.* Structure and Dynamics of Partially Solidified Systems, Ed. D.E. Loper. Martinus-Nijhoff, 1987, 229-290.
164. D.E. Loper, P.H. Roberts, *A Boussinesq model of a slurry.* Structure and Dynamics of Partially Solidified Systems, Ed. D.E. Loper. Martinus-Nijhoff, 1987, 293-323.
165. R.N. Hills, P.H. Roberts, *Towards a non-local theory of helium II.* Quart. J. Mech. & Appl. Math., **40**, 1987, 279-301.
166. S.I. Braginsky, P.H. Roberts, *A model-Z geodynamo.* Geophys. & Astrophys. Fluid Dynam., **38**, 1987, 327-349.
167. D.E. Loper, P.H. Roberts, *A simple mathematical model of slurry.* Hydrodynamical Behavior and Interacting Particle Systems, Ed. G. Papanicolaou. Springer, 1987, 113-116.
168. R.N. Hills, P.H. Roberts, *On the use of Fick's law in regions of mixed phase.* Int. Comm. Heat Mass Transf., **15**, 1988, 113-119.
169. D.R. Fearn, P.H. Roberts, A.M. Soward, *Convection, stability and the dynamo.* Energy, Stability and Convection, Eds. G.P. Galdi and B. Straughan. Longmans, 1988, pp.60-324.
170. P.C. Pal, P.H. Roberts, *Long term polarity stability and strength of the geomagnetic dipole.* Nature, **331**, 1988, 702-705.
171. S. Putterman, P.H. Roberts, W. Fiszdon, *Theory of 2D plasma waves as a sensitive probe of nonlinear effects due to compressibility of* ^4He. Phys. Lett, A, **128**, 1988, 203-206.
172. R.N. Hills, P.H. Roberts, *On the formulation of diffusive mixture theories for two-phase regions.* J. Engng. Math., **22**, 1988, 93-106.
173. M.A. Berger, P.H. Roberts, *On the winding number problem with finite steps.* Adv. Appl. Prob., **20**, 1988, 261-274.
174. S. Putterman, P.H. Roberts, *Fokker-Planck equation for interacting waves. Dispersion law renormalization and wave turbulence.* Phys. Rpts., **168**, 1988, 209-263.
175. R.N. Hills, P.H. Roberts, *A generalized Scheil-Pfann equation for a dynamical theory of a mushy zone.* Int. J. Non-linear Mech., **23**, 1988, 327-339.
176. P.H. Roberts, *Future of Geodynamo Theory.* Geophys. & Astrophys. Fluid Dynam., **44**, 1988, 3-31.
177. P.H. Roberts, *On topographic core-mantle coupling.* Geophys. & Astrophys. Fluid Dynam., **44**, 1988, 181-187.
178. R.N. Hills, P.H. Roberts, *Healing and relaxation in flows of Helium II. V. Thermodynamic Potential and Applications.* J. Low Temp. Phys., **74**, 1989, 185-203.
179. P.H. Roberts, *From Taylor state to model-Z?* Geophys. & Astrophys. Fluid Dynam., **49**, 1988, 143-160.
180. P.H. Roberts, *Core-mantle coupling.* The Encyclopedia of Solid Earth Geophysics, Ed. D.E. James. Van Nostrand, 1989, 148-160.
181. R.N. Hills, P.H. Roberts, *A macroscopic model of phase coarsening.* Int. J. Non-linear Mech., **25**, 1990, 319-329.
182. C.A. Jones, P.H. Roberts, D.J. Galloway, *Compressible convection in the presence of rotation and a magnetic field.* Geophys. & Astrophys. Fluid Dynam., **53**, 1990, 145-182.
183. W. Kuang, P.H. Roberts, *Resistive instabilities in rapidly rotating fluids: linear theory of the tearing mode.* Geophys. & Astrophys. Fluid Dynam., **55**, 1990, 199-239.
184. C.A. Jones, P.H. Roberts, *Magnetoconvection in rapidly rotating Boussinesq and compressible fluids.* Geophys. & Astrophys. Fluid Dynam., **55**, 1990, 263-308.
185. R.N. Hills, P.H. Roberts, *On the life cycle of a mush.* Stab. & Appl. Anal. Cont. Media, **1**, 1991, 1-26.

186. J. Bloxham, P.H. Roberts, *The geomagnetic main field and the geodynamo.* Rev. Geophys. Supp. (US National Report to International Union of Geodesy and Geophysics 1987-1990), 1991, 428-432.
187. P.H. Roberts, *Magnetoconvection patterns in rotating convection zones.* The Sun and Cool Stars, Eds. I. Tuominen, D. Moss and G. Rüdiger. Springer, 1991, 37-56.
188. M. Kono, P.H. Roberts, *Small amplitude solutions of the dynamo problem, 1. The adjoint system and its solutions.* J. Geomag. Geoelectr., **43**, 1991, 839-862.
189. P.H. Roberts, *Dynamo theory.* Chaotic Processes in the Geological Sciences, Ed. D.A. Yuen. Springer, 1992, 237-280.
190. R.N. Hills, P.H. Roberts, *On the motion of a fluid that is incompressible in a generalized sense, and its relationship to the Boussinesq approximation.* Stab. & Appl. Anal. Cont. Media, **1**, 1991, 205-212.
191. S.J. Putterman, P.H. Roberts, *Elementary excitations of nonlinear classical fields and the theory of superfluid He^4.* Microscopic Theories of Superfluids, Ed. G. Grioli. Cambridge University Press, 1991, 4-21.
192. W. Kuang, P.H. Roberts, *Resistive instabilities in rapidly rotating fluids: linear theory of the g-mode.* Geophys. & Astrophys. Fluid Dynam., **60**, 1991, 295-333.
193. D.E. Loper, P.H. Roberts, *Core-mantle system.* Encyclopedia of Earth System Science **1**, Ed. W.A. Nierenberg. Academic Press, 1992, 655-663.
194. P.H. Roberts, *Geomagnetism.* Encyclopedia of Earth System Science **2**, Ed. W.A. Nierenberg. Academic Press, 1992, 277-294.
195. P.H. Roberts, *Geomagnetism; origin.* Encyclopedia of Earth Science System **2**, Ed. W.A. Nierenberg. Academic Press, 1992, 295-309.
196. R.N. Hills, D.E. Loper, P.H. Roberts, *On continuum models for momentum, heat and species transport in solid-liquid phase change systems.* Int. Comm. Heat Mass Transfer, **19**, 1992, 585-594.
197. B. Denado, A. Larraza, S. Putterman, P. Roberts, *Nonlinear theory of localized standing waves.* Phys. Rev. Lett., **69**, 1992, 597-600.
198. P.H. Roberts, A.M. Soward, *Dynamo theory.* Ann. Rev. Fluid Mech., **24**, 1992, 459-512.
199. M. Kono, P.H. Roberts, *Small amplitude solutions of the dynamo problem. 2. The case of α^2-dynamos.* Geophys. & Astrophys. Fluid Dynam., **67**, 1992, 65-85.
200. W. Kuang, P.H. Roberts, *Resistive instabilities in rapidly rotating fluids: linear theory of the convective modes.* Geophys. & Astrophys. Fluid Dynam., **67**, 1992, 129-162.
201. R.N. Hills, P.H. Roberts, *Relaxation effects and the evolution of a mush.* Stab. & Appl. Anal. Cont. Mech., **2**, 1992, 305-321.
202. M.B. Haeri, S.J. Putterman, A. Garcia, P.H. Roberts, *Quantum wave turbulence.* Phys. Rev. E, **47**, 1993, 739-742.
203. S. Putterman, P.H. Roberts, *Nonlinear theory of modulated standing waves; domain walls, kinks and breathers.* Proc. R. Soc. Lond., A**440**, 1993, 135-148.
204. S. Lan, W. Kuang, P.H. Roberts, *Ideal instabilities in rapidly rotating MHD systems that have critical layers.* Geophys. & Astrophys. Fluid Dynam., **69**, 1993, 133-160.
205. P.H. Roberts, T.H. Jensen, *Homogeneous dynamos: Theory and practice.* Phys. Fluids B, **5**, 1993, 2657-2662.
206. C.C. Wu, P.H. Roberts, *Shock wave propagation in a sonoluminescing gas bubble.* Phys. Rev. Lett., **70**, 1993, 3424-3427.
207. R.N. Hills, P.H. Roberts, *A note on the kinetic conditions at a supercooled interface.* Int. Comm. Heat Mass Transf., **20**, 1993, 407-416.
208. G.A. Glatzmaier, P.H. Roberts, *Intermediate dynamo models.* J. Geomag. & Geoelectr., **45**, 1993, 1605-1616.
209. P.H. Roberts, *The geodynamo.* The Cosmic Dynamo, Eds. F. Krause, G. Rüdiger and K.-H. Rädler. Reidel, 1993, 431-440.
210. P.H. Roberts, *Dynamo theory.* Astrophysical Fluid Dynamics, Eds. J.-P. Zahn and J. Zinn-Justin. Elsevier, 1993, 229-323.
211. C.C. Wu, P.H. Roberts, *A model of sonoluminescence.* Proc. R. Soc. Lond., A**445**, 1994, 323-349.
212. P.H. Roberts, *Magnetohydrodynamics*, Encyclopedia of Appl. Physics, **9**, Ed. G. Trigg. Amer. Inst. Phys., 1994, 111-156.
213. B.P. Barber, C.C. Wu, R. Löfstedt, S.J. Putterman, P.H. Roberts, *Sensitivity of sonoluminescence to experimental parameters.* Phys. Rev. Ltrs., **72**, 1994, 1380-1383.

214. P.H. Roberts, *Fundamentals of dynamo theory*, Lectures on Solar & Planetary Dynamos, Eds. M.R.E Proctor and A.D. Gilbert. Cambridge University Press, 1994, 1-58.
215. P.H. Roberts, *The solar dynamo*, The Solar Engine and its Influence on Terrestrial Atmosphere and Climate, Ed. E. Nesme–Ribes. Springer, 1994, 1-26.
216. S.I. Braginsky, P.H. Roberts, *From Taylor state to model–Z*, Geophys. & Astrophys. Fluid Dynam., **77**, 1994, 3-13.
217. M. Kono, P.H. Roberts, *Kinematic dynamos with anisotropic $\alpha-$effect*, Geophys. & Astrophys. Fluid Dynam., **77**, 1994, 27-53.
218. R.D. Gibson, P.H. Roberts, *Solution to a class of integral equations with an application to magnetohydrodynamics*, Meth. & Appl. Anal., **1**, 1994, 421-433.
219. P.H. Roberts, E. Schatzman, *The solar dynamo mechanism.* Cours de Structure Interne; Production et rôle du champ magnétique. Structure Interne des Étoiles et des Planètes Géantes, Eds. A.M. Hubert, E. Schatzman. Observatoire de Paris, 1994, 61-117.
220. T. Nakajima, P.H. Roberts, *A mapping method for solving dynamo equations*, Proc. R. Soc. Lond. A**448**, 1995, 1-28.
221. S.I. Braginsky, P.H. Roberts, *Equations governing convection in Earth's core and the geodynamo*, Geophys. & Astrophys. Fluid Dynam., **79**, 1995, 1-97.
222. P.H. Roberts, *On the ratio of two double integrals*, SIAM Review, **37**, 1995, 244-245.
223. P.H. Roberts, *Dynamics of the core, geodynamo*, (US National Report to the International Union of Geodesy and Geophysics), Reviews of Geophys., Supp., 1995, 443–450.
224. T. Nakajima, P.H. Roberts, *An application of a mapping method to the asymmetric dynamo equations.*, Phys. Earth Planet. Int., **91**, 1995, 53-61.
225. G.A. Glatzmaier, P.H. Roberts, *A three–dimensional convective dynamo solution with rotating and finitely conducting inner core and mantle*, Phys. Earth Planet. Int., **91**, 1995, 63-75.
226. G.A. Glatzmaier, P.H. Roberts, *A three–dimensional self–consitent computer simulation of the geomagnetic field*, Nature, **377**, 1995, 203-209 (and cover).
227. R.N. Hills, P.H. Roberts, *On the influence of solid-phase volume fraction on the adiabatic coarsening of a mushy zone.* Int. Comm. in Heat Mass Transfer, **22**, 1995, 771-777.
228. D.E. Winch, P.H. Roberts, *Derivatives of addition theorems for Legendre functions*, J. Austr. Math. Soc. B, **37**, 1995, 212-234.
229. R.F. Sinclair, S. Putterman, P.H. Roberts, *Quantum Fokker–Planck equation for interacting waves*, Phil. Trans. R. Soc. Lond. **354**, 1996, 951-978.
230. N.B. Morley, P.H. Roberts, *Solutions of uniform open–channel, liquid metal flow in a strong, oblique magnetic field.* Phys. Fluids, **8**, 1996, 923-935.
231. P.H. Roberts, C.C. Wu, *Structure and stability of a spherical implosion.* Phys. Letrs. A, **213**, 1996, 59-64.
232. G.A. Glatzmaier, P.H. Roberts, *An anelastic evolutionary geodynamo simulation driven by compositional and thermal convection.* Physica D, **97**, 1996, 81-94.
233. C.C. Wu, P.H. Roberts, *Structure and stability of a spherical shock wave in a van der Waals gas.* Quart. J. Mech. appl. Math., **49**, 1996, 501-543.
234. G.A. Glatzmaier, P.H. Roberts, *Rotation and magnetism of Earth's inner core*, Science, **274**, 1996, 1887-1892 (and cover).
235. G.A. Glatzmaier, P.H. Roberts, *Magnetic sounding of planetary interiors*, Phys. Earth Planet. Inter., **98**, 1996, 207-220.
236. P.H. Roberts, *Equations governing convection in Earth's core and the geodynamo*, Mathematical Aspects of Thermal Convection. Institute of Mathematical Analysis, Kyoto University, Japan, 1996, 92-104.
237. G.A. Glatzmaier, P.H. Roberts, *Numerical simulations of the geodynamo*, Acta Astronom. Geophys. Univ. Com., **19**, 1997, 125-143.
238. G.A. Glatzmaier, P.H. Roberts, *Simulating the geodynamo*, Contemp. Phys., **38**, 1997, 269-288 (and cover).
239. K. Zhang, P.H. Roberts, *Thermal inertial waves in a rotating fluid layer: Exact and asymptotic solutions*, Phys. Fluids, **9**, 1997, 1980-1987.
240. M.J. Shearer, P.H. Roberts, *The hidden ocean at the top of Earth's core.* Dynam. Atmos. & Oceans, **27**, 631-647 (1997).

241. G.A. Glatzmaier, P.H. Roberts, *Numerical simulations of Earth's magnetic field* Geowissenschaften, **15**, 1997, 95-99.
242. R.N. Hills, P.H. Roberts, *Microstructural coarsening kinetics for a mushy zone of a pure material*, Int. J. Non-lin. Mech., **32**, 1997, 1003-1013.
243. C.C. Wu, P.H. Roberts, *On rectified diffusion and sononluminescence*, Theor. & Comp. Fluid Dynam., **10** (1998), 357-372.
244. S.J. Putterman, P.H. Roberts, *Comment on "Bubble shape oscillations and the onset of sonoluminescence"*, Phys. Rev. Letrs., **80**, 1998, 3666-3667.
245. K. Zhang, P.H. Roberts, *A note on the stabilising & destabilising effects of Ekman boundary layers*, Geophys. & Astrophys. Fluid Dynam., **88**, 1998, 215-223.
246. G.A. Glatzmaier, P.H. Roberts, *Dynamo theory then and now*, Int. J. Engng. Sci., **36**, 1998, 1325-1338.
247. P.H. Roberts, C.C.Wu, *The decay of bubble oscillations.* Phys. Fluids, **10**, 1998, 3227-3229.
248. C.C. Wu, P.H. Roberts, *Bubble shape instabilities and sonoluminescence*, Physics Letts. A., **250**, 1998, 131-136.
249. C.C. Wu, P.H. Roberts, *Richtmyer-Meshkov instability and the dynamics of the magnetosphere*, Geophys. Res. Lettrs., **26**, 1999, 655-658.
250. M. Matsushima, T. Nakajima, P.H. Roberts, *The anisotropy of local turbulence in Earth's core*, Earth Planets Space, **51**, 1999, 277-286.
251. N.G. Berloff, P.H. Roberts, *Motions in a bose condensate VI. Vortices in a nonlocal model*, J. Phys. A: Math. Gen., **32**, 1999, 5611-5625.
252. R.D. Gibson, P.H. Roberts, *Solution to a class of integral equations with an application to magneto-hydrodynamics*, In *Recent Advances in Fluid Mechanics*, Eds. P.L. Sanchev and M. Venkatachalappa, Gordon & Breach, London, 1999, pp. 35-48.
253. G.A. Glatzmaier, R.S. Coe, L. Hongre, P.H. Roberts, *The role of the Earth's mantle in controlling the frequency of geomagnetic reversals*, *Nature*, **401**, 1999, 885-890.
254. C.A. Jones, P.H. Roberts, *Convection driven dynamos in a rotating plane layer*, *J. Fluid Mech.*, **404**, 2000, 311-343.
255. N.G. Berloff, P.H. Roberts, *Motions in a bose condensate VII. Boundary layer separation*, J. Phys. A: Math. Gen., **33**, 2000, 4025-4038.
256. G.A. Glatzmaier, P.H. Roberts, *A test of the frozen flux approximation using geodynamo simulations*, Proc. R. Soc. Lond., A**358**, 2000, 1109-1121.
257. P.H. Roberts, K. Zhang, *Thermal generation of Alfvén waves in oscillatory magnetoconvection.* J. Fluid Mech., **420**, 2000, 201-223.
258. G.A. Glatzmaier, P.H. Roberts, *Geodynamo theory and simulations.* Rev. Mod. Phys., **72**, 2000, 1081-1124.
259. P.H. Roberts, C.A. Jones, *The onset of magnetoconvection at large Prandtl number in a rotating layer I. Finite magnetic diffusion.* Geophys. & Astrophys. Fluid Dynam., **92**, 2000, 289-325.
260. C.A. Jones, P.H. Roberts, *The onset of magnetoconvection at large Prandtl number in a rotating layer II. Small magnetic diffusion.* Geophys. & Astrophys. Fluid Dynam., **93**, 2000, 173-226.
261. N.G. Berloff, P.H. Roberts, *Roton creation and vortex nucleation in superfluids.* Phys. Rev., **A247**, 2000, 69-74.
262. N.G. Berloff, P.H. Roberts, *Motions in a Bose condensate VIII. The electron bubble.* J. Phys. A. Math. Gen., **34**, 2001, 81-91.
263. G.A. Glatzmaier, P.H. Roberts, *The geodynamo past, present and future.* Geophys. & Astrophys. Fluid Dynam., **94**, 2001, 47-84.
264. D.E. Loper, P.H. Roberts, *Mush-chimney convection.* Stud. Appl. Math., **106**, 2001, 187-227.
265. N.G. Berloff, P.H. Roberts, *Capture of an impurity by a vortex line in a Bose condensate.* Phys. Rev. B, **63**, 2001, 81-91.
266. M. Kono, P.H. Roberts, *Definition of the Rayleigh number for geodynamo simulations.* Phys. Earth Planet. Interiors, **128**, 2001, 13-24.
267. N.G. Berloff, P.H. Roberts, *Motion in a Bose condensate: IX. Crow instability of antiparallel vortex pairs.* J. Phys. A. Math. Gen., **34**, 2001, 10057–10066.
268. P.H. Roberts, N.G. Berloff, *The nonlinear Schrödinger equation as a model of superfluidity.* Quantum Vortex Dynamics and Superfluid Turbulence, eds. C.F. Barenghi, R.J. Donnelly, W.F. Vinen. Springer, Berlin, pp. 2001, 235–257.

269. N.G. Berloff, P.H. Roberts, *Vortices in nonlocal condensate models of superfluid helium.* Quantum Vortex Dynamics and Superfluid Turbulence, eds. C.F. Barenghi, R.J. Donnelly, W.F. Vinen. Springer, Berlin, 2001, pp. 268–275.
270. S.I. Braginsky, P.H. Roberts, *On the theory of convection in the Earth's core.* Advances in Nonlinear Dynamos (The Fluid Mechanics of Astrophysics and Geophysics, **9**). Eds. A. Ferriz-Mas & M. Nuñez, Taylor & Francis, London, 2003, pp. 60–82.