LECTURE NOTES IN PHYSICS

Edith Falgarone Thierry Passot (Eds.)

Turbulence and Magnetic Fields in Astrophysics



Springer

Lecture Notes in Physics

Editorial Board

R. Beig, Wien, Austria
B.-G. Englert, Ismaning, Germany
U. Frisch, Nice, France
P. Hänggi, Augsburg, Germany
K. Hepp, Zürich, Switzerland
W. Hillebrandt, Garching, Germany
D. Imboden, Zürich, Switzerland
R. L. Jaffe, Cambridge, MA, USA
R. Lipowsky, Golm, Germany
H. v. Löhneysen, Karlsruhe, Germany
I. Ojima, Kyoto, Japan
D. Sornette, Nice, France, and Los Angeles, CA, USA
S. Theisen, Golm, Germany
W. Weise, Trento, Italy, and Garching, Germany
J. Wess, München, Germany
J. Zittartz, Köln, Germany

Springer
Berlin
Heidelberg
New York
Barcelona
Hong Kong
London
Milan
Paris
Tokyo

Physics and Astronomy ONLINE LIBRARY
http://www.springer.de/phys/

Editorial Policy

The series *Lecture Notes in Physics* (LNP), founded in 1969, reports new developments in physics research and teaching -- quickly, informally but with a high quality. Manuscripts to be considered for publication are topical volumes consisting of a limited number of contributions, carefully edited and closely related to each other. Each contribution should contain at least partly original and previously unpublished material, be written in a clear, pedagogical style and aimed at a broader readership, especially graduate students and nonspecialist researchers wishing to familiarize themselves with the topic concerned. For this reason, traditional proceedings cannot be considered for this series though volumes to appear in this series are often based on material presented at conferences, workshops and schools (in exceptional cases the original papers and/or those not included in the printed book may be added on an accompanying CD ROM, together with the abstracts of posters and other material suitable for publication, e.g. large tables, colour pictures, program codes, etc.).

Acceptance

A project can only be accepted tentatively for publication, by both the editorial board and the publisher, following thorough examination of the material submitted. The book proposal sent to the publisher should consist at least of a preliminary table of contents outlining the structure of the book together with abstracts of all contributions to be included.
Final acceptance is issued by the series editor in charge, in consultation with the publisher, only after receiving the complete manuscript. Final acceptance, possibly requiring minor corrections, usually follows the tentative acceptance unless the final manuscript differs significantly from expectations (project outline). In particular, the series editors are entitled to reject individual contributions if they do not meet the high quality standards of this series. The final manuscript must be camera-ready, and should include both an informative introduction and a sufficiently detailed subject index.

Contractual Aspects

Publication in LNP is free of charge. There is no formal contract, no royalties are paid, and no bulk orders are required, although special discounts are offered in this case. The volume editors receive jointly 30 free copies for their personal use and are entitled, as are the contributing authors, to purchase Springer books at a reduced rate. The publisher secures the copyright for each volume. As a rule, no reprints of individual contributions can be supplied.

Manuscript Submission

The manuscript in its final and approved version must be submitted in camera-ready form. The corresponding electronic source files are also required for the production process, in particular the online version. Technical assistance in compiling the final manuscript can be provided by the publisher's production editor(s), especially with regard to the publisher's own Latex macro package which has been specially designed for this series.

Online Version/ LNP Homepage

LNP homepage (list of available titles, aims and scope, editorial contacts etc.):
http://www.springer.de/phys/books/lnpp/

LNP online (abstracts, full-texts, subscriptions etc.):
http://link.springer.de/series/lnpp/

Edith Falgarone Thierry Passot (Eds.)

Turbulence and Magnetic Fields in Astrophysics



Springer

Editors

Edith Falgarone
Laboratoire de Radioastronomie,
Ecole Normale Supérieure,
24, Rue Lhomond,
75231 Paris Cedex 05, France

Thierry Passot
CNRS,
Observatoire de la Côte d'azur,
B.P. 4229,
06304 Nice Cédex 4, France

Cover Picture: (see contribution "MHD Turbulence in Star-Forming Regions and the Interstellar Medium" by M. Mac Low in this volume)

Library of Congress Cataloging-in-Publication Data.

Turbulence and magnetic fields in astrophysics / Edith Falgarone, Thierry Passot (eds.).
p. cm. -- (Lecture notes in physics, ISSN 0075-8450; 614)
Includes bibliographical references.
ISBN 3-540-00274-X
1. Astrophysics. 2. Magnetohydrodynamics. 3. Turbulence. I. Falgarone, E. (Edith) II. Passot, Thierry, 1960- III. Series.

QB462.8.T87 2003
523.01'886--dc21

2002042925

ISSN 0075-8450
ISBN 3-540-00274-X Springer-Verlag Berlin Heidelberg New York

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilm or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer-Verlag. Violations are liable for prosecution under the German Copyright Law.

Springer-Verlag Berlin Heidelberg New York
a member of BertelsmannSpringer Science+Business Media GmbH

http://www.springer.de

© Springer-Verlag Berlin Heidelberg 2003
Printed in Germany

The use of general descriptive names, registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

Typesetting: Camera-ready by the authors/editor
Camera-data conversion by Steingraeber Satztechnik GmbH Heidelberg
Cover design: *design & production*, Heidelberg

Printed on acid-free paper
57/3141/du - 5 4 3 2 1 0

Preface

This book contains review articles of most of the topics addressed at the conference on *Simulations of Magnetohydrodynamic turbulence in astrophysics: recent achievements and perspectives* which took place from July 2 to 6, 2001 at the Institut Henri Poincaré in Paris. We made the choice to publish these lectures in a tutorial form so that they can be read by a broad audience. As a result, this book does not give an exhaustive view of all the subjects addressed during the conference.

The main objective of this workshop which gathered about 90 scientists from different fields, was to present and confront recent results on the topic of turbulence in magnetized astrophysical environments. A second objective was to discuss the latest generation of numerical codes, such as those using adaptive mesh refinement (AMR) techniques.

During a plenary discussion at the end of the workshop discussions were held on several topics, often at the heart of vivid controversies. Topics included the timescale for the dissipation of magneto-hydrodynamical (MHD) turbulence, the role of boundary conditions, the characteristics of imbalanced turbulence, the validity of the polytropic approach to Alfvén waves support within interstellar clouds, the source of turbulence inside clouds devoid of stellar activity, the timescale for star formation, the Alfvén Mach number of interstellar gas motions, the formation process for helical fields in the interstellar medium. The impact of small upon large scales was also discussed. Does an incorrect treatment of small scales (e.g. neglect of the Hall effect) affect large scales? What is the impact on reconnection? Is there a relation between the coherent structures of magnetic fields (sheets, tubes ...) and the plasma parameter β at the scale considered? What is the effect of magnetic helicity ejection on the dynamo mechanism? The articles presented in this book address many of these issues, and illustrate the fact that many of them remain open questions.

The book has been divided into four chapters. In the first chapter, general topics in incompressible and compressible MHD turbulence are discussed. These topics include scaling laws, intermittency, energy decay, reduced MHD (with a discussion on coronal heating models), a new regime of turbulence at very large Prandtl number and weak turbulence. In particular, the issue of anisotropy and its relation with the Goldreich-Sridhar and the Iroshnikov-Kraichnan theory is fully discussed. The second chapter contains a general discussion of numerical issues in astrophysical MHD turbulence and a presentation of the NIRVANA

code. The third chapter is devoted to the physics of the interstellar medium and star formation, with a special emphasis on the role of turbulence. First, recent observations of magnetic fields (intensity and direction) are reviewed. The interplay between turbulence, magnetic fields, self-gravity, thermal instabilities and energy injection is studied and results are interpreted to account for the formation, lifetime and star formation activity of clouds. Diagnostic tools are presented that allow a quantitative comparison between numerical modeling and the observable universe. A scenario of star formation which reproduces the stellar initial mass function (IMF) is proposed. The stability of filaments of matter threaded by helical fields is discussed and compared to observations. Turbulence in accretion disks is also reviewed. The last chapter is devoted to new scenarios of magnetic reconnection, in particular in collisionless or turbulent plasmas, and their relation with dynamo theory. The role of boundaries and helicity in astrophysical dynamos is particularly stressed.

We wish to express our special thanks to the lecturers for their stimulating presentations and to all the participants who contributed to the success of this meeting. We greatly appreciated the involvement of anonymous referees whose comments were very valuable in making the papers readable by a large audience.

We are especially indebted to Michel Pérault, Annie Touchant and Marie-Françoise Ducos, without whom the workshop would not have run so smoothly. We also gratefully acknowledge the support of the staff of the Institut Henri Poincaré for help with the practical details in the organization of the meeting. The workshop benefitted from the support of CNRS through the National Program "Physique et Chimie du Milieu Interstellaire", the research group "Turbulence" and the advisory committee of Section 14, and from the Observatoire de Paris and the Laboratoire de Radioastronomie.

Paris, Nice
August 2002

Edith Falgarone
Thierry Passot

Table of Contents

Part I MHD Turbulence

The Evolving Phenomenological View on Magnetohydrodynamic Turbulence
Wolf-Christian Müller, Dieter Biskamp 3

Coronal Heating and Reduced MHD
Sean Oughton, Pablo Dmitruk, William H. Matthaeus 28

MHD Turbulence: Scaling Laws and Astrophysical Implications
Jungyeon Cho, Alex Lazarian, Ethan T. Vishniac 56

Part II Numerical Approches

Numerical Simulations of Magnetic Fields in Astrophysical Turbulence
Ellen G. Zweibel, Fabian Heitsch, Yuhong Fan 101

Adaptive Mesh Refinement in MHD Modeling. Realization, Tests and Application
Udo Ziegler 127

Part III Turbulence in the ISM

Observations of Interstellar Magnetic Fields
Richard Crutcher, Carl Heiles, Thomas Troland 155

MHD Turbulence in Star-Forming Regions and the Interstellar Medium
Mordecai-Mark Mac Low 182

Thermal Instability and Magnetic Pressure in the Turbulent Interstellar Medium
Enrique Vázquez-Semadeni, Adriana Gazol, Thierry Passot, Javier Sánchez-Salcedo 213

Developing Diagnostics of Molecular Clouds Using Numerical MHD Simulations
Eve C. Ostriker 252

Star Formation and the Initial Mass Function
Åke Nordlund, Paolo Padoan 271

The Structure and Formation of Filamentary Molecular Clouds
Jason D. Fiege 299

Numerical Simulations of MHD Turbulence in Accretion Disks
Steven A. Balbus, John F. Hawley 329

Part IV Current Issues in Reconnection and Astrophysical Dynamos

Recent Developments in Collisionless Reconnection Theory: Applications to Laboratory and Astrophysical Plasmas
Amitava Bhattacharjee, Zhiwei W. Ma, Xiaogang Wang 351

Problems and Progress in Astrophysical Dynamos
Ethan T. Vishniac, Alex Lazarian, Jungyeon Cho 376

The Helicity Issue in Large Scale Dynamos
Axel Brandenburg 402

Recent Developments in Magnetic Dynamo Theory
Eric G. Blackman 432

List of Contributors

Steven A. Balbus
Dept. of Astronomy, University of Virginia, Charlottesville, VA 22901
sb@virginia.edu

Amitava Bhattacharjee
Center for Magnetic Reconnection Studies, Department of Physics and Astronomy, University of Iowa, Iowa City, IA 52242, USA
amitava-bhattacharjee@uiowa.edu

Dieter Biskamp
Centre for Interdisciplinary Plasma Science, Max-Planck-Institut für Plasmaphysik, 85748 Garching, Germany
dfb@ipp.mpg.de

Eric G. Blackman
Department of Physics and Astronomy, and Laboratory for Laser Energetics, University of Rochester, Rochester NY 14627, USA
blackman@pas.rochester.edu

Axel Brandenburg
NORDITA, Blegdamsvej 17, DK-2100 Copenhagen Ø, Denmark
brandenb@nordita.dk

Jungyeon Cho
University of Wisconsin, Madison WI53706, USA
cho@astro.wisc.edu

Richard Crutcher
University of Illinois, Urbana IL 61801, USA
crutcher@astro.uiuc.edu

Pablo Dmitruk
Bartol Research Institute, University of Delaware, Newark, DE 19716, USA
PABLO@bartol.udel.edu

Yuhong Fan
High Altitude Observatory, NCAR, Boulder, CO 80307, USA
yfan@ucar.edu

Jason D. Fiege
Canadian Institute for Theoretical Astrophysics, McLennan Labs, University of Toronto, 60 St. George Street, Toronto, Ontario, Canada M5S 3H8
fiege@cita.utoronto.ca

Adriana Gazol
Instituto de Astronomía, UNAM, Campus Morelia, Apdo. Postal 3-72, Morelia, Michoacán, MEXICO
a.gazol@astrosmo.unam.mx

John F. Hawley
Dept. of Astronomy, University of Virginia, Charlottesville, VA 22901
jh8h@virginia.edu

Carl Heiles
University of California, Berkeley CA 91820, USA
heiles@vermi.Berkeley.edu

Fabian Heitsch
JILA, University of Colorado, Boulder CO 80309, USA
heitsch@colorado.edu

Alex Lazarian
University of Wisconsin, Madison WI53706, USA
lazarian@astro.wisc.edu

Zhiwei W. Ma
Center for Magnetic Reconnection Studies, Department of Physics and Astronomy, University of Iowa, Iowa City, IA 52242, USA
weima@plasma5.physics.uiowa.edu

Mordecai-Mark Mac Low
Dept. of Astrophysics, American Museum of Natural History, 79th Street and Central Park W., New York, NY, 10024-5192, USA
mordecai@amnh.org

William H. Matthaeus
Bartol Research Institute, University of Delaware, Newark, DE 19716, USA
yswhm@bxclu.bartol.udel.edu

Wolf-Christian Müller
Centre for Interdisciplinary Plasma Science, Max-Planck-Institut für Plasmaphysik, 85748 Garching, Germany
Wolf.Mueller@cips.mpg.de

Åke Nordlund
Astronomical Observatory / NBIfAFG and Theoretical Astrophysics Center, Juliane Maries Vej 30, DK-2100 Copenhagen Ø, Denmark
aake@astro.ku.dk

Eve C. Ostriker
Department of Astronomy, University of Maryland, College Park, MD 20742, USA
ostriker@astro.umd.edu

Sean Oughton
Department of Mathematics, University of Waikato, Hamilton, New Zealand
seano@waikato.ac.nz

Paolo Padoan
Jet Propulsion Laboratory, 4800 Oak Grove Drive, MS 169-506 California Institute of Technology, Pasadena, CA 91109-8099, USA
padoan@jpl.nasa.gov

Thierry Passot
CNRS, Observatoire de la Côte d'azur, B.P. 4229, 06304, Nice, Cédex 4, France
passot@obs-nice.fr

Javier Sánchez-Salcedo
Instituto de Astronomía, UNAM, Apdo. Postal 70-264, México, D.F., 04510, México
jsanchez@astroscu.unam.mx

Thomas Troland
University of Kentucky, Lexington KY 40502, USA
troland@pa.uky.edu

Enrique Vázquez-Semadeni
Instituto de Astronomía, UNAM, Campus Morelia, Apdo. Postal 3-72, Morelia, Michoacán, MEXICO
e.vazquez@astrosmo.unam.mx

Ethan T. Vishniac
Johns Hopkins University, Baltimore MD21218, USA
ethan@skysrv.pha.jhu.edu

Xiaogang Wang
Department of Physics, Dalian University of Technology, Dalian, China 116024
xiaogang@plasma4.physics.uiowa.edu

Udo Ziegler
Astrophysikalisches Institut Potsdam, An der Sternwarte 16,D-14482 Potsdam, Germany
uziegler@aip.de

Ellen G. Zweibel
JILA, University of Colorado, Boulder CO 80309, USA
zweibel@solarz.colorado.edu

The Evolving Phenomenological View on Magnetohydrodynamic Turbulence

Wolf-Christian Müller and Dieter Biskamp

Centre for Interdisciplinary Plasma Science,
Max-Planck-Institut für Plasmaphysik, 85748 Garching, Germany

Abstract. The current level of phenomenological MHD turbulence modelling is outlined together with recent results of experimental measurements and direct numerical simulations. After reviewing the spatially isotropic Kolmogorov and Iroshnikov-Kraichnan phenomenologies, the model of Goldreich and Sridhar is discussed, which introduces the concept of local anisotropy caused by the magnetic field. The She-Lévêque picture of intermittency dealing with the anomalous scaling of the two-point structure functions is also considered, followed by a summary of the phenomenological theory of turbulent energy-decay. It is shown how the combination of phenomenology and numerical simulation leads to the physical model of homogeneous MHD turbulence.

1 Introduction

Nature confronts the astronomical observer with a large variety of systems which – by experimental measurement or theoretical prediction – are perceived as plasmas in turbulent motion. The ubiquity of turbulence in the universe is a direct consequence of the fact that the dynamics of almost every non-stationary astrophysical plasma flow is characterized by a wide range of length scales and velocities. Therefore, the theoretical description of these flows seems to be one of the major keys for a better understanding of as diverse observations as the generation of large-scale magnetic fields accompanying many celestial objects by the turbulent dynamo effect [1,2], the dynamics of stellar winds and their interaction with planetary magnetospheres [3], the discrepancy between observed and predicted life-times of star-forming molecular clouds in the insterstellar medium [4], and the angular momentum transport within accretion disks prone to magnetorotational instability [5], to name a few examples.

In spite of many efforts turbulence has so far withstood all attempts to tame its complexity by first-principle-based theoretical treatment. The development of a rigorous statistical formalism for hydrodynamic turbulence, which has been initiated in the 1930s by Taylor, von Kármán, Howarth, Keller, and Friedman [6], culminated during the 1970s in a variety of stochastic models and statistical approximations (see, e.g., [7,8,9,10,11,12,13] and the reviews in [14,15]). Kraichnan's direct-interaction approximation (DIA) [7], for example, represents a self-consistent expansion of the Navier-Stokes and magnetohydrodynamic (MHD) equations. However, the DIA formalism is not invariant under random Galilean transformations, leading to, e.g., an incorrect energy spectrum scaling in the hydrodynamic case. The necessary invariance property has been restored by

E. Falgarone and T. Passot (Eds.): LNP 614, pp. 3–27, 2003.
© Springer-Verlag Berlin Heidelberg 2003

heuristic modifications to the theory, known as the Lagrangian History DIA [8] (see [16] for a detailed presentation).

The algebraic structure of the mentioned theories is quite involved and, consequently, ties the derivation of experimentally verifiable statements to severely simplifying assumptions on the turbulent system. Therefore, the only statistical approximation having been applied fruitfully to MHD turbulence is the eddy-damped quasi-normal Markovian approximation (EDQNM) [17] in the field of dynamo theory. The EDQNM approach yields evolution equations for the second-order statistical moments of the turbulent fields by modeling the fourth-order cumulants as linear damping terms where the associated eddy damping rate acts as a free parameter.

A different line-of-thought began in the 1940s with Kolmogorov's model of turbulence, tracing back to the ostensive picture put forward by Richardson in the 1920s (for an introduction see, for example, [18,19]). Instead of formally expanding or modifying the underlying differential equations, the Kolmogorov theory is built on phenomenological hypotheses based on observation and physical intuition. Though in general incapable of giving exact numerical values of constants or prefactors, phenomenologies built in this way can predict certain important scaling properties of turbulence, e.g., the exponents of power-laws observed in the Fourier energy spectra (see below).

In this article we outline the current state of MHD turbulence phenomenology and put recent theoretical developments in context with results of experimental measurements and direct numerical simulations (DNS) which are crucial for the verification of phenomenological hypotheses. Section 2 briefly summarizes the MHD equations and their basic features in the context of turbulent flows. In Sect. 3 the main concepts of turbulence phenomenology are presented, followed by a view on models of intermittency and turbulent energy-decay in Sects. 4 and 5, respectively. Section 6 gives a brief summary of experimental and recent numerical results, comparing them to theory.

2 Properties of Magnetohydrodynamics

To keep the theoretical treatment of plasma turbulence tractable, a number of simplifying assumptions have to be adopted. To this end, the resistive MHD approximation is applied, regarding the plasma as a quasi-neutral, electrically conducting, single fluid featuring viscous and resistive dissipation.

This implies that the characteristic spatial scales of the system are much larger than the mean free path of its microscopic constituents and that the shortest resolved time scales are those of the fast magnetosonic wave. Concerning inner friction, the MHD fluid is assumed to be Newtonian.

The MHD equations can be derived in a formally consistent way from the conservation laws of mass, momentum and electric charge in combination with Maxwell's equations [20]. A more heuristic approach can, for example, be found in [21].

For additional simplicity the turbulent flow is considered as being incompressible, presupposing the fluid velocity to be significantly smaller than the slowest speed of compressible wave propagation and effectively eliminating the magnetosonic modes. The mass density ϱ_0 is assumed to be spatially uniform.

In this framework the equations of incompressible MHD read

$$\partial_t \boldsymbol{v} = -\boldsymbol{v}\cdot\nabla\boldsymbol{v} - \nabla p - \boldsymbol{b}\times(\nabla\times\boldsymbol{b}) + \mu\Delta\boldsymbol{v}\ , \tag{1}$$

$$\partial_t \boldsymbol{b} = \nabla\times(\boldsymbol{v}\times\boldsymbol{b}) + \eta\Delta\boldsymbol{b}\ , \tag{2}$$

$$\nabla\cdot\boldsymbol{v} = \nabla\cdot\boldsymbol{b} = 0\ , \tag{3}$$

with the dimensionless fluid velocity $\boldsymbol{v}$, magnetic field $\boldsymbol{b}$, and thermodynamic pressure p. All quantities can be dressed with units using a characteristic length L_0, velocity v_0 and mass density ϱ_0 as reference quantities.

Because of incompressibility the pressure p is a merely passive variable, being determined by a Poisson equation which is the divergence of the momentum balance (1). It is, therefore, possible to eliminate p from the equations by solely considering the rotational part of (1) which gives an evolution equation for the vorticity $\boldsymbol{\omega} = \nabla\times\boldsymbol{v}$,

$$\partial_t\boldsymbol{\omega} = \nabla\times(\boldsymbol{v}\times\boldsymbol{\omega} - \boldsymbol{b}\times\boldsymbol{j}) + \mu\Delta\boldsymbol{\omega}\,,$$

introducing the electric current density $\boldsymbol{j} = \nabla\times\boldsymbol{b}$.

The dynamic state of a MHD flow is fully characterized by the non-dimensional parameters

$$\mu = \mathsf{Re}^{-1} = \frac{\mu_*}{L_0 v_0}, \quad \eta = \mathsf{Rm}^{-1} = \frac{\eta_*}{L_0 v_0},$$

where Re is the kinetic Reynolds number defined with the kinematic viscosity μ_* and Rm the magnetic Reynolds number involving the magnetic diffusivity η_* of the MHD fluid. The two parameters are rough estimates of the strength of the nonlinearities in the respective equation compared to dissipation introduced by the last terms on the right-hand-sides of (1) and (2). If both Re and Rm become sufficiently larger than unity, a MHD flow undergoes a transition from the laminar state, characterized by a stationary stream-line topology, to turbulence, where the fluid motion seems to be erratic and unpredictable.

MHD turbulence can be coarsely classified by considering the kinetic energy of the flow field $E^{\mathrm{K}} = \frac{1}{2}\int_V \mathrm{d}V v^2$ and the magnetic energy $E^{\mathrm{M}} = \frac{1}{2}\int_V \mathrm{d}V b^2$ with V denoting the spatial volume of the system. The two extremes, which will not be treated here and represent fields of research on their own right, are characterized by $E^{\mathrm{K}} \gg E^{\mathrm{M}}$, where the magnetic field is passively advected by the fluid, and $E^{\mathrm{K}} \ll E^{\mathrm{M}}$ with a strong magnetic field forcing the fluid motion into quasi-two-dimensionality. While the former condition is typical for the dynamo problem, i.e., the amplification of a magnetic field by plasma turbulence, the latter is, for example, encountered in the solar corona or terrestrial laboratory experiments with magnetic plasma confinement. For studying inherent properties of the nonlinear interaction between the turbulent fields $\boldsymbol{v}$ and $\boldsymbol{b}$, we will further on concentrate on turbulence with $E^{\mathrm{K}} \sim E^{\mathrm{M}}$ and a magnetic Prandtl number $\mathsf{Pr_m} = \mu/\eta$ equal to unity.

2.1 The Ideal Invariants

If the turbulent system is closed with appropriate, e.g., periodic, boundary conditions, there exist a number of invariants in the ideal limit $\mu = \eta = 0$ [22]. The quadratic ones are particularly important since they remain invariant when truncating the spatial Fourier representation of the turbulent system at a finite wavenumber. For incompressible, three-dimensional MHD these 'rugged' ideal invariants are given together with their decay rates at finite viscosity and magnetic diffusivity:
total energy

$$E = E^{\mathrm{K}} + E^{\mathrm{M}} = \frac{1}{2}\int_V \mathrm{d}V\,(v^2 + b^2), \quad \dot{E} = -\int_V dV\ (\mu\omega^2 + \eta j^2),$$

magnetic helicity

$$H^{\mathrm{M}} = \frac{1}{2}\int_V \mathrm{d}V\,(\boldsymbol{a}\cdot\boldsymbol{b}), \quad \dot{H}^{\mathrm{M}} = -\eta\int_V dV\ \boldsymbol{j}\cdot\boldsymbol{B},$$

and cross helicity

$$H^{\mathrm{C}} = \frac{1}{2}\int_V \mathrm{d}V\,(\boldsymbol{v}\cdot\boldsymbol{b}), \quad \dot{H}^{\mathrm{C}} = -\frac{\mu+\eta}{2}\int_V dV\ \boldsymbol{\omega}\cdot\boldsymbol{j}$$

with the magnetic vector potential $\boldsymbol{a}$, $\boldsymbol{b} = \nabla\times\boldsymbol{a}$.

The three ideal invariants play a prominent role in the characterization of the macroscopic properties of MHD turbulence. The magnetic helicity H^{M} is a measure for linkage and twist of the magnetic field lines [23] while the cross helicity H^{C} gives the overall correlation of magnetic and velocity field.

2.2 Elsässer Variables

An equivalent formulation of (1)–(3), introducing the Elsässer fields $\boldsymbol{z}^{\pm}=\boldsymbol{v}\pm\boldsymbol{b}$ [24], reads

$$\nabla\cdot\boldsymbol{z}^{\pm} = 0\ , \tag{4}$$

$$\partial_t\boldsymbol{z}^{\pm} = -\boldsymbol{z}^{\mp}\cdot\nabla\boldsymbol{z}^{\pm} - \nabla(p + \frac{b^2}{2}) + \frac{\mu+\eta}{2}\Delta\boldsymbol{z}^{\pm} + \frac{\mu-\eta}{2}\Delta\boldsymbol{z}^{\mp}\,. \tag{5}$$

The Elsässer variables are more fundamental in incompressible MHD turbulence than $\boldsymbol{v}$ and $\boldsymbol{b}$ alone as the symmetry of (4) and (5) suggests. Moreover, in the presence of a background magnetic field $|\boldsymbol{b}_0| \gg |\boldsymbol{b}|$ it is possible to interprete the $\boldsymbol{z}^{\pm}$ fields as shear Alfvén-wave packets traveling along $\boldsymbol{b}_0$, where $\boldsymbol{z}^{+}$ and $\boldsymbol{z}^{-}$ propagate in opposite directions.

In the following text $\{\boldsymbol{v},\boldsymbol{b}\}$ or $\boldsymbol{z}^{\pm}$ are used interchangeably, depending on which variable set yields the clearest presentation.

3 Phenomenologies of MHD Turbulence

The most prominent characteristic of turbulence is the seemingly unpredictable variation of the flow observables. Their random values follow statistical distributions which allow to define smoothly evolving quantities, e.g., ensemble averages denoted by angular brackets, $\langle \bullet \rangle$.

Hereinafter, the focus will be on the inherent properties of 'fully developed' turbulence where the turbulent system has reached dynamical equilibrium between the energy inflow, supplied by driving mechanisms, e.g., a large-scale shear instability, and the energy outflow caused by small-scale dissipation. The equilibrium does not have to be stationary in time.

The turbulent fields $\boldsymbol{v}$ and $\boldsymbol{b}$ are assumed to be spatially homogeneous in their statistical properties. This assumption is only approximately valid in small spatial regions of real turbulence but strongly facilitates analytical calculations.

Statistical homogeneity together with the quasi-ergodicity hypotheses allow to replace the ensemble averages introduced above by spatial or (in the stationary case) time averages. Ignoring the influence of (far-away) boundaries, periodic boundary conditions are applied at the outer surfaces of the volume V containing the turbulent flow.

Besides its apparent randomness, turbulence has remarkable similarity properties. Certain functions derived from the turbulent fields, e.g., the moments of the two-point probability distribution function or the total energy evolution in time, exhibit invariance under scaling transformations[1] if the influence of initial and boundary conditions are negligible. Such a self-similar function asymptotically follows a power-law, $f(x) = Cx^{\alpha}$, where the scaling exponent α can often be derived phenomenologically without exact knowledge about either the functional form or the numerical value of the coefficient C.

3.1 Kolmogorov Picture (K41)

The K41 phenomenology [25,26], which does not include a magnetic field but represents the foundation of nearly all phenomenological turbulence models, regards a turbulent velocity field as the superposition of structures ('eddies') characterized by a spatial scale $\boldsymbol{\ell}$ and the associated velocity field increment $\delta v_\ell = [\boldsymbol{v}(\boldsymbol{r}+\boldsymbol{\ell}) - \boldsymbol{v}(\boldsymbol{r})] \cdot \boldsymbol{\ell}/\ell$. Due to additionally assumed statistical isotropy the field increments depend solely on ℓ, allowing to define the characteristic eddy velocity

$$v_\ell = \langle \delta v_\ell^2 \rangle^{1/2} \ .$$

The K41 picture differentiates three spatial scale ranges: (i) the energy-containing scales, driving the flow, (ii) the dissipation range at smallest scales, where dissipative effects dominate, removing energy from the system, and (iii) the 'inertial' range where nonlinear interactions govern the dynamics and the

[1] That is to say, $f(\lambda x) = \lambda^{\alpha} f(x)$.

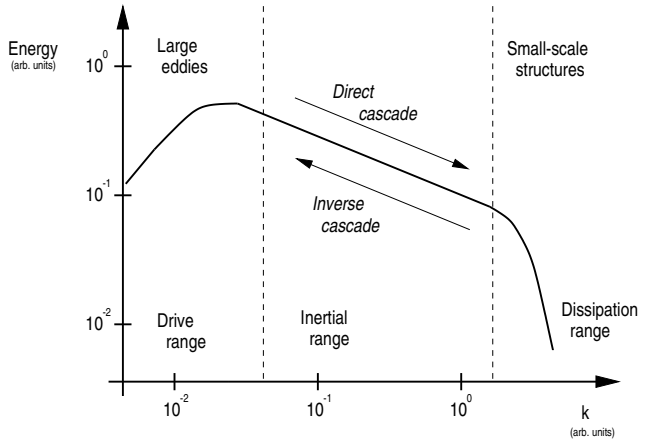


Fig. 1. Schematic representation of the K41 picture of turbulence showing the spatial energy spectrum as an example.

influence of driving and dissipation are negligible (see Fig. 1). Within the latter region it is experimentally observed that the structure functions of order p

$$S_p^v(\ell) = \langle \delta v_\ell^p \rangle$$

exhibit spatial self-similarity, i.e., they follow power-laws in ℓ with constant p-dependent exponents ζ_p.

The K41 theory predicts values for these scaling exponents in the limiting case of infinite Reynolds number. Since the structure functions are the statistical moments of the two-point probability distribution of the respective turbulent field, their scaling exponents provide fundamental information about the statistical properties of turbulence, e.g., $S_2^v(\ell)$ is directly linked to the one-dimensional energy spectrum while the p-dependence of the ζ_p characterizes the intermittency of flow structures. The relation of the K41 picture to the asymptotic statistical invariance properties of fully-developed turbulence is treated in detail in [18].

The dynamics of the spectral energy flow in the inertial range is idealized as a *Richardson cascade* meaning that in the spatial hierarchy of eddies the larger ones constantly transfer a fraction of their energy towards smaller counterparts [27]. The cascade is termed 'direct' because the resulting energy flux is headed towards smaller spatial scales with the size of interacting eddies differing only slightly, i.e., the transfer proceeds in small local steps in spatial Fourier space. Nonlinear turbulent cascades also work towards larger scales, e.g. the inverse cascade of magnetic helicity in three-dimensional MHD turbulence.[2]

The cascades are caused by the nonlinearities in the Navier-Stokes and MHD equations which redistribute the cascading quantities dissipationless among different scales of motion. In particular, the energy flux can be estimated by order of magnitude as v_ℓ^2/τ_ℓ with the nonlinear eddy-turnover time $\tau_\ell = \ell/v_\ell$.

[2] The inverse cascade of magnetic helicity leads to the build-up of large-scale magnetic structures and is intimately linked to the turbulent dynamo effect [17].

K41 Results

In fully developed turbulence including all assumptions described above the nonlinear energy-flux in the inertial range equals the energy dissipation, ε. This yields by order of magnitude

$$\begin{aligned} \varepsilon &\overset{!}{\sim} \frac{v_\ell^2}{\tau_\ell} = \frac{v_\ell^3}{\ell} \qquad \Rightarrow v_\ell \sim (\varepsilon\ell)^{1/3} \\ &\Rightarrow S_p^v(\ell) \sim (\varepsilon\ell)^{p/3} \,. \end{aligned} \tag{6}$$

For the case $p = 3$ an exact relation exists, the four-fifth law (see, e.g., [18]):

$$S_3^v(\ell) = -\frac{4}{5}\varepsilon\ell \,.$$

The Fourier transform of $S_2^v(\boldsymbol{\ell})$ scales like the one-dimensional energy spectrum defined as $E_1(k_1) = \frac{1}{2}\int \mathrm{d}k_2 \int \mathrm{d}k_3 \boldsymbol{v}(\boldsymbol{k}) \cdot \boldsymbol{v}^*(\boldsymbol{k})$ with $k_i = (2\pi)/\ell_i$. In isotropic turbulence one usually considers the angle-integrated energy spectrum

$$E(k) = \int_0^{4\pi} \mathrm{d}\Omega \, E(\boldsymbol{k})|_{|\boldsymbol{k}|=k} \,, \tag{7}$$

defining $E(\boldsymbol{k}) = \frac{1}{2}|\boldsymbol{v}(\boldsymbol{k})|^2$. The real-space exponent ζ_2 of $S_2^v(\boldsymbol{\ell})$ corresponds to an inertial range Fourier-space scaling $\sim k^{-(1+\zeta_2)}$ which yields the experimentally well-supported Kolmogorov spectrum

$$E(k) = C_\mathrm{K} \varepsilon^{2/3} k^{-5/3} \,,$$

with the Kolmogorov constant $C_\mathrm{K} \approx 1.6$.

The Kolmogorov dissipation length $\ell_{\mathrm{K}41}$ gives an order of magnitude estimate of the spatial scales where energy dissipation $\sim v_\ell^2 \mu/\ell^2$ begins to dominate the nonlinear energy-dynamics $\sim v_\ell^2/\tau_\ell$, which marks the beginning of the dissipation range. With the approximations introduced above,

$$\begin{aligned} \frac{\mu}{\ell^2} &\overset{!}{\sim} \tau_\ell^{-1} = v_\ell/\ell \sim (\varepsilon\ell)^{1/3}/\ell \\ &\Rightarrow \ell_{K41} = \left(\frac{\mu^3}{\varepsilon}\right)^{1/4} . \end{aligned}$$

3.2 Iroshnikov-Kraichnan Picture (IK)

The IK phenomenology [28,29] for MHD turbulence is a modification of the K41 approach introducing a different model of the nonlinear energy flux. While the K41 energy cascade is caused by mutual eddy scrambling, in the IK picture energy is transported mainly by Alfvén waves propagating along the local magnetic field. The spectral energy redistribution takes place during the nonlinear scattering of colliding Alfvén-wave packets.

In the following the Elsässer quantity z_ℓ is used analogously to v_ℓ in the hydrodynamic case. Restricting consideration to MHD turbulence with small mean correlation $\rho = |H^{\rm C}|/(E^{\rm K}E^{\rm M})^{1/2}$ it is not necessary to distinguish z_ℓ^+ and z_ℓ^-. For finite ρ the respective energy spectra scalings may differ considerably [30], demanding a different theoretical approach [31].

The coupling of magnetic and velocity field through the generation and attenuation of Alfvén waves, termed *Alfvén effect*, leads to local equipartition of magnetic and kinetic energy. The physical process can be illustrated by regarding the plasma motions at a fixed scale ℓ significantly smaller than the integral scale of the turbulent system. Eddies of size ℓ are interacting with the magnetic field $\boldsymbol{b}_0$, carried by the largest energy-containing swirls, in the way that velocity perturbations v_ℓ are triggering Alfvén waves via locally deforming the quasi-constant guide field $\boldsymbol{b}_0$. The incompressible transverse waves then travel along $\boldsymbol{b}_0$. If the involved perturbations δv, δB are small compared to b_0, one has approximately $\delta v = \pm\delta B$.

For two colliding Alfvén waves of extent ℓ the interaction time to nonlinearly exchange energy is given by $\tau_{\rm A} = \ell/b_0$.[3] Due to $b_0 \gg b_\ell$ this time is much shorter than the corresponding K41 transfer time $\tau_\ell = \ell/z_\ell$ and thus the time τ_* entering the IK energy flux z_ℓ^2/τ_* is enlarged by the factor $\tau_\ell/\tau_{\rm A}$ compared to the K41 case.

IK Results

With the approximations and transformations presented above, in particular

$$v_\ell \longrightarrow z_\ell \ , \quad \tau_\ell \longrightarrow \frac{\tau_\ell}{\tau_{\rm A}}\tau_\ell \ , \tag{8}$$

all K41 results can be readily modified for the IK picture. The basic inertial range scaling of the Elsässer fields in the IK case is given by

$$z_\ell \sim (\varepsilon b_0 \ell)^{1/4} \quad \Rightarrow \quad S_p^z(\ell) \sim (\varepsilon b_0 \ell)^{p/4} \ .$$

Analogously the energy spectrum is obtained as

$$E(k) = C_{\rm IK}(\varepsilon b_0)^{1/2} k^{-3/2}$$

while the IK dissipation length is found to be

$$\ell_{\rm IK} = \left(\tfrac{b_0\eta^2}{\varepsilon}\right)^{1/3}. \tag{9}$$

There also exists an exact relation corresponding to the four-fifth law for incompressible, three-dimensional MHD ([32], [33]),

$$\sum_{i=1}^{3} \langle \delta z_\ell^{\mp} (\delta_i z_\ell^{\pm})^2 \rangle = -\frac{4}{3}\varepsilon^{\pm}\ell \, ,$$

[3] In the chosen dimensionless system the magnetic field is given as the Alfvén speed at reference density, $\boldsymbol{v}_{\rm A} = \boldsymbol{b}/(4\pi\varrho_0)^{1/2}$, in multiples of the reference velocity v_0.

with $\varepsilon^{\pm} = \frac{1}{2}\int_V \mathrm{d}V[\mu\omega^2 + \eta j^2 \pm (\mu+\eta)\boldsymbol{\omega}\cdot\boldsymbol{j}]$, $\delta z_\ell^{\mp}$ denoting the longitudinal field increments introduced in Sect. 3.1 and $\delta_i z_\ell^{\pm} = (z^{\pm}(\boldsymbol{r}+\boldsymbol{\ell}) - z^{\pm}(\boldsymbol{r}))\cdot\boldsymbol{e}_i$ introducing the unit vector $\boldsymbol{e}_i$ from the orthogonal base of an arbitrary coordinate system.

3.3 Goldreich-Sridhar Picture (GS)

The K41 and IK phenomenologies are spatially isotropic in the way that turbulent structures are characterized by a single length-scale ℓ. However, in the presence of a magnetic field the characteristic interaction time of Alfvén-wave like distortions of extent λ along the field lines, $\tau_\lambda \sim \lambda/b_0$, is typically much shorter than the nonlinear turnover time $\tau_l \sim l/z_l$ of their components of extent l and amplitude z_l perpendicular to the local magnetic field. In addition, the nonlinear energy flux exhibits a pronounced anisotropy, being hampered along the direction of the magnetic field [34,35].

Against this background Goldreich and Sridhar developed a semi-phenomenological model [36,37,38] which takes the magnetic-field-induced spatial anisotropy into account. Though the local mean field $\boldsymbol{b}_0$ is usually assumed to stem from an external source, it might as well be generated by eddies on larger spatial scales analogous to the IK picture. In accordance with the weakness of field-parallel energy transfer, the GS model predicts the scaling of turbulent fluctuations perpendicular to $\boldsymbol{b}_0$ (for an important exception see the discussion of GS2 below).

The GS model consists of three parts. The first (GS1) being a weak turbulence formalism in combination with several assumptions [36] which are still subject to controversial discussion [39,40,41,42]. Since GS1 is basically a formal perturbative expansion for weak turbulence (see also [41]), we abstain from considering it here. However, a result in this framework, which led to the development of phenomenologies for what Goldreich and Sridhar call 'intermediate' (GS2) and 'strong' (GS3) MHD turbulence, is the strengthening of nonlinear interactions with increasing spectral wavenumber. This behaviour invalidates the technique underlying GS1 if the considered spatial scales of motion are sufficiently small.

Intermediate Turbulence

In intermediate MHD turbulence [38] the contribution of higher order interaction terms in the expansion GS1 can not be neglected anymore, demanding a different approach. To this end the IK phenomenology is extended to include anisotropy effects. One considers an Alfvén-wave like eddie of size λ and amplitude z_λ parallel to the local magnetic field and, respectively, l and z_l perpendicular to it. The nonlinear time-scale transformation given in (8) can be generalized by associating the nonlinear turnover time τ_l with swirling motion perpendicular to $\boldsymbol{b}_0$, while the parallel interaction time is given by τ_λ. Analogously to (8) the transformation of the K41 nonlinear turnover time, τ_ℓ, into the frame of the

anisotropic IK phenomenology GS2 reads

$$\tau_\ell \longrightarrow \frac{\tau_l}{\tau_\lambda}\tau_l \,, \tag{10}$$

leading to

$$z_l \sim (\varepsilon b_0/\lambda)^{1/4} l^{1/2} \quad \Rightarrow \quad S_p^z(l) \sim (\varepsilon b_0/\lambda)^{p/4} l^{p/2} \,.$$

The space increment l is perpendicular to the local magnetic field, contrary to the quantity ℓ in the K41 and IK phenomenologies which denotes an increment in an arbitrary direction. By $z_l^2 \sim k_l E(k_l)$ with $k_l \sim l^{-1}$ and $k_\lambda \sim \lambda^{-1}$ the perpendicular energy spectrum is obtained as

$$E(k_l) \sim (\varepsilon b_0 k_\lambda)^{1/2} k_l^{-2} \,.$$

Since a field-parallel energy cascade is precluded (see above), k_λ is constant by order of magnitude under variation of k_l.

Strong Turbulence

The anisotropic IK phenomenology GS2 still assumes the distortion due to a single nonlinear interaction to be small compared to the amplitude of the interacting eddies. The strengthening of nonlinear effects with increasing spectral wavenumber, therefore, renders the GS2 model invalid if the considered spatial scales of motion are small enough. This gives rise to the GS3 picture of strong turbulence [37].

The GS3 phenomenology can be derived from the GS2 model by presuming a hypothetical equality[4] of the characteristic time scales parallel and perpendicular to the local magnetic field, in the language of Goldreich and Sridhar a 'critical balance'. This is, in fact, nothing else than setting the prefactor τ_l/τ_λ on the right-hand-side of (10) equal to unity, i.e.,

$$\tau_\ell \longrightarrow \tau_l \sim l/z_l \sim \lambda/b_0 \,.$$

Carrying out the already presented steps, one arrives at

$$z_l \sim (\varepsilon l)^{1/3} \quad \Rightarrow \quad S_p^z(l) \sim (\varepsilon l)^{p/3} \,, \tag{11}$$

which gives a Kolmogorov-like perpendicular energy spectrum,

$$E(k_l) \sim \varepsilon^{2/3} k_l^{-5/3} \,.$$

The GS3 model, which is in principle an anisotropic K41 phenomenology due to the assumed equality of parallel and perpendicular time-scales, yields a relation between the corresponding spatial scales of the turbulent eddies. By $\lambda/b_0 \sim l/z_l$ in combination with (11) one obtains

$$\lambda \sim \varepsilon^{-1/3} b_0 l^{2/3} \,, \tag{12}$$

[4] by order of magnitude

which implies that eddies become elongated along $\boldsymbol{b}_0$ with decreasing spatial scale. Another consequence of (12) is an energy cascade parallel to $\boldsymbol{b}_0$ with the associated one-dimensional energy spectrum

$$E(k_\lambda) \sim \left(\varepsilon^3 b_0^{-5}\right)^{1/2} k_\lambda^{-5/2} .$$

4 Intermittency

The K41 and IK models, which implicitly assume spatial uniformity of energy dissipation, ε, give the scaling exponents ζ_p of structure functions of arbitrary order p as $\zeta_p^{\mathrm{K41}} = p/3$ and $\zeta_p^{\mathrm{IK}} = p/4$. There is, however, experimental evidence for significant deviations from the predicted linear behaviour in hydrodynamic turbulence (cf. [43] for a review), in the turbulent solar wind [44] as well as in DNS of MHD turbulence [45,46,47].

This *anomalous scaling* is attributed to a departure from self-similarity of the turbulent fields in the inertial range. The effect can be linked to the spatial distribution of the turbulent structures responsible for energy dissipation by Kolmogorov's refined similarity hypothesis [48] which introduces the local energy dissipation in a sphere of radius ℓ, ε_ℓ. This approach is motivated by the observation that energy dissipation is not homogeneously distributed in space. Instead, the turbulent system is interspersed with small regions of intense dissipation, resulting in a pronounced spatial intermittency.

4.1 She-Lévêque Model (SL)

Among the phenomenological theories dealing with anomalous structure-function scaling (cf. [18,43]), the model of She and Lévêque [49] is exceptional as it (i) only contains parameters which can be estimated by physical reasoning and (ii) achieves very good agreement with experimental measurements.

The fundamental hypotheses of the SL phenomenology postulates a hierarchical relation between the statistical moments of ε_ℓ. In their original work She and Lévêque attribute this property to a "hidden symmetry" of the Navier-Stokes equations. It could be shown that the hypotheses is equivalent to assuming a logarithmic Poisson distribution of the ε_ℓ over ℓ [50], the symmetry being a generalized scale-covariance of the flow field [50,51].

In generic form, neglecting time-scale intermittency, the scaling exponents of the structure functions in the SL model depend on two parameters, g and C_0 (see, e.g., [52]),

$$\zeta_p^{\mathrm{GSL}} = \frac{p}{g}\left(1 - \frac{2}{g}\right) + C_0\left[1 - \left(1 - \frac{2}{gC_0}\right)^{p/g}\right] .$$

While g is linked to the model of nonlinear energy transfer via the basic scaling relation $v_\ell \sim \ell^{1/g}$, C_0 can be interpreted as the co-dimension of the smallest turbulent structures. In three-dimensional hydrodynamic turbulence these

structures are quasi-one-dimensional vorticity filaments ($C_0 = 2$), contrary to the MHD case where micro-current and vorticity sheets ($C_0 = 1$) are forming [47].

Hydrodynamic Intermittency

In the three-dimensional hydrodynamic case, one assumes K41 scaling $v_\ell \sim l^{1/3}$, i.e., $g = 3$, and vortex filaments with $C_0 = 2$. This yields the SL-model which fits most experimental observations very well:

$$\zeta_p^{\mathrm{SL}} = \frac{p}{9} + 2 - 2\left(\frac{2}{3}\right)^{p/3} \quad .$$

MHD Intermittency

Assuming the IK picture, $z_\ell \sim \ell^{1/4}$, i.e., $g = 4$, in three-dimensional MHD turbulence together with $C_0 = 1$ [52,53] results in

$$\zeta_p^{\mathrm{IK}} = \frac{p}{8} + 1 - \left(\frac{1}{2}\right)^{p/4} \quad .$$

High-resolution DNS [47], however, give scalings that are clearly different to this prediction. Instead, convincing agreement is achieved when referring to the K41 phenomenology with $C_0 = 1$ which is also consistent with other observations from three-dimensional MHD turbulence (see below):

$$\zeta_p^{\mathrm{MHD}} = \frac{p}{9} + 1 - \left(\frac{1}{3}\right)^{p/3} \quad . \tag{13}$$

5 Energy Decay

The energy evolution in freely decaying MHD turbulence represents a classical problem which can be treated phenomenologically. While the energy of a laminar flow decays exponentially $E(t) \sim \exp(-2\eta k^2 t)$, the decay of turbulence exhibits a power-law behaviour $\sim t^{-\beta}$. In the hydrodynamic case, Kolmogorov derived the decay exponent $\beta = 10/7$ [54] by assuming the, meanwhile questioned, invariance of the Loitsianksii integral $\sim \int \mathrm{d}r r^4 \langle v(\boldsymbol{x}+\boldsymbol{r}) v(\boldsymbol{x})\rangle$ (see, e.g., [19]). Laboratory experiments, on the other hand, do not give a clear picture with β ranging between 1 and 2 (cf., for example, [55]).

Recently, the invariance of an analogue to the Loitsianskii integral in MHD turbulence has been postulated, using Elsässer variables in the corresponding correlation function [56]. In combination with the IK concept of an Alfvén-wave dominated nonlinear energy-transfer $\beta = 5/6$ is obtained. However, the derivation contains an inconsistency [47] which puts the result in question.

Selective Decay

It appears that finite magnetic helicity is characteristic of most astrophysical plasmas, since magnetic turbulence usually occurs in rotating systems, where the combined action of Coriolis and buoyancy forces naturally lead to twisted field lines. For finite H^{M}, selective decay, i.e., the much slower decay of H^{M} compared to that of the energy, will dominate the dynamics. The basic idea has been developed for the decay of enstrophy, $\Omega = \frac{1}{2}\int \boldsymbol{\omega}^2 \mathrm{d}2S$, in 2D hydrodynamic turbulence [57] and was later applied to MHD turbulence [58,59] (and references therein). For high Reynolds number, H^{M} can be considered invariant during energy decay, a property which is more robust than the questionable invariance of the Loitsianskii integral and allows the construction of a phenomenological model for the energy decay.

Defining the characteristic length scale ℓ_0 by $E^{\mathrm{M}}\ell_0 = H^{\mathrm{M}}$ and applying the relations H^{M} =const, $\ell_0 \sim E^{3/2}/\varepsilon$, and $E \sim E^{\mathrm{M}}$ one finds

$$(\mathrm{d}E/\mathrm{d}t)H^{\mathrm{M}}E^{-5/2} = \mathrm{const}, \tag{14}$$

which has the asymptotic solution $E \sim t^{-2/3}$ [21].

Departure from Self-similarity

DNS reveal significant deviations from the phenomenological decay law (14) (see below), which are attributed to a departure from self-similarity in the energy decay. In particular the ratio $\Gamma = E^{\mathrm{K}}/E^{\mathrm{M}}$ is not constant, as implicitly assumed in the derivation of (14), but decreases at a rate comparable to that of the energy, $\Gamma \simeq 0.1 \times E/H^{\mathrm{M}}$ [60].

To account explicitly for the variation of Γ, it is assumed that the most important nonlinearities in the MHD equations arise from the advective terms, giving

$$-\frac{\mathrm{d}E}{\mathrm{d}t} \sim \boldsymbol{v}\cdot\nabla E \sim (E^{\mathrm{K}})^{1/2}\frac{E}{\ell_0}.$$

Substitution of the integral scale ℓ_0 introduced above gives

$$\frac{E^{5/2}}{\varepsilon H^{\mathrm{M}}}\,\frac{\Gamma^{1/2}}{(1+\Gamma)^{3/2}} = \mathrm{const}. \tag{15}$$

which together with $\Gamma \sim E/H^{\mathrm{M}}$ describes the energy decay.

In the limit $\Gamma \ll 1$, which is the asymptotic state decaying MHD turbulence is observed to strive to, one finds the similarity solution $E \sim t^{-1/2}$. For finite Γ, the decay is somewhat steeper, flattening to $t^{-1/2}$ as Γ becomes small [47].

6 Experimental and Numerical Evidence

Since all phenomenologies presented in previous sections are based on hypotheses, their experimental verification is indispensable, creating the link to nature's first principles that pure mathematics cannot provide, yet.

In the laboratory, experimentally accessible large-Reynolds number MHD turbulence is only found in the large devices for magnetically confined nuclear fusion. There, however, it is the result of an unwanted disruption of the discharge and, consequently, has to be avoided [21]. The only natural system showing fully developed MHD turbulence and being accessible to in-situ measurements is the solar wind, in particular the regions of comparably slow bulk speed. This medium has meanwhile been explored by space probes within a radial distance between 0.3 and 80 astronomical units from the sun and is observed to show fluctuations of flow velocity and magnetic field over a wide range of spatial scales together with a spectral energy distribution often following a power-law $\sim k^{-1.7}$ [61]. Though this spectral index seems to support the K41 picture there are still uncertainties about the nature of the solar wind whose explication might lie beyond the framework of incompressible homogeneous MHD turbulence.

Therefore, the main source of experimental knowledge stems from DNS of idealized systems. The numerical approach allows for 'perfect' diagnostics under well-defined conditions but is restricted by the limitation of computational resources which today allow to obtain Reynolds numbers of the order 10^3 [47] compared to real astrophysical values of at least 10^8 [1].[5] Nevertheless, high-resolution DNS are a useful instrument for studying the similarity properties of MHD turbulence when being aware of the existence of finite-Reynolds-number effects.

6.1 Numerics

Pseudospectral methods (see, e.g., [62]), discretizing the turbulent fields in Fourier space, have become the standard numerical tool for DNS of homogeneous MHD turbulence.

While, compared to real-space finite-difference algorithms, spatial differentiation is strongly simplified, i.e., reduced to wave-vector multiplication, nonlinear terms become convolution sums in Fourier space which are numerically expensive to evaluate. The computational cost can, however, be reduced to the level of finite-difference schemes, i.e., on a three-dimensional grid by three orders of magnitude, by computing the convolutions in real space and applying Fast-Fourier-Transforms (FFTs) for the shuttling between real and Fourier space.

Apart from the restriction to periodic boundary conditions, pseudospectral Fourier-space algorithms involve an unwanted effect known as 'aliasing' [63]. This numerical error is introduced by mapping the result of a real-space product (the Fourier-space convolution) on finite Fourier space. It can be eliminated by surrounding the Fourier-space grid of linear extension N with a buffer zone of size $N/2$, taking up the erroneous aliasing modes. In practice, the aliasing error can be reduced to the order of the discretization error by restricting the Fourier-space volume of interest to a sphere of radius N [64]. It should also be

[5] As the Reynolds numbers are a measure for the spatial range between the largest energy-containing and the smallest dissipative scales, increasing them demands a finer spatial grid to numerically resolve all motions of the turbulent flow.

noted that pseudospectral schemes are not very well-suited for massively-parallel computing architectures since FFTs involve a high level of processor-to-processor communication. As a consequence, parallel pseudospectral codes usually do not reach the speed of their finite-difference correspondents.

These drawbacks are outweighed by higher precision at same spatial resolution of pseudospectral schemes compared to finite-difference algorithms. Moreover, pseudospectral codes signal insufficient spatial resolution for given Reynolds numbers by an energy pile-up at the largest wavenumbers. This is not the case for finite-difference methods (for a comparative study of finite-difference and pseudospectral schemes in turbulence DNS, see [65]). Finally, the theoretical models of turbulence rest on the reduction of the turbulent system into a set of structures at different spatial scales – an idea inherent to the pseudospectral approach.

Figure 2 displays, as an example, the real-space vorticity in a typical periodic-cube configuration of a high-resolution pseudospectral DNS with 512^3 spatial Fourier modes.

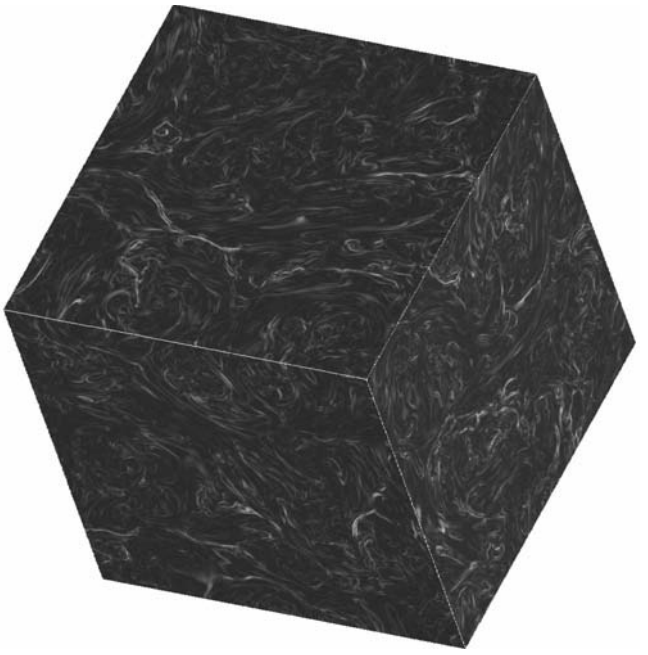


Fig. 2. Cuts, in (x, y), (x, z), (y, z) planes, of vorticity $|\boldsymbol{\omega}|$ in 512^3-DNS of decaying MHD turbulence.

6.2 Energy Decay

The temporal evolution of energy, E, in three-dimensional decaying MHD turbulence has been subject to various numerical investigations (see, e.g., [66,67,60] for

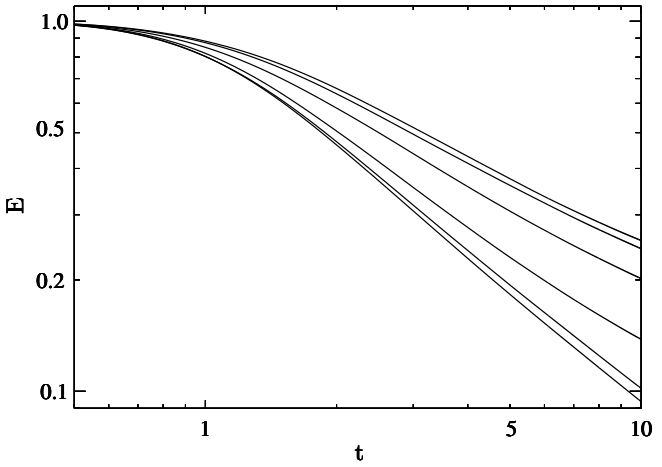


Fig. 3. Energy evolution for varying magnetic helicity $H^M = 0, 0.15, 0.38, 0.7, 0.92, 1[H^{\rm M}_{\rm max}]$ (bottom to top) in DNS of decaying MHD turbulence.

DNS with the currently highest spatial resolution). The theoretical predictions for the exponent of the power-law $E \sim t^{-\beta}$ range between 2/3 and 10/7 (see Sect. 5). As can be inferred from Fig. 3, the decay rate strongly depends on the magnetic helicity $H^{\rm M}$ in the system, being the lower the larger $|H^{\rm M}|$. The only phenomenologies of turbulent energy decay which take finite magnetic helicity into consideration are (14) and (15). DNS show that the simple phenomenology (14), built on the full self-similarity of energy decay, does not apply. In contrast, the approximate constancy of expression (15) over the period of fully developed turbulence ($t > 3$) shown in Fig. 4, which depicts data gained with DNS of varying resolution (512^3, 256^3) and at different values of $H^{\rm M}$ [60], validates the associated simple phenomenology.

In the case of vanishing $H^{\rm M}$, one finds a different decay law [47],

$$({\rm d}E/{\rm d}t)E^{-2} = {\rm const}, \tag{16}$$

which corresponds to the solution $E \sim t^{-1}$ in agreement with DNS of three-dimensional compressible MHD turbulence [66,67]. The decay is similar as in two-dimensional MHD turbulence [68], though, contrary to the latter, there is no obvious selective decay process for $H^{\rm M} = 0$ in the 3D case.

As for $H^{\rm M} = 0$ there is no clear conserved quantity, it is difficult to interpret the decay law (16), a situation familiar from hydrodynamic turbulence. If the decay depends sensitively on the small-k spectrum, the t^{-1}-law may not be universal. The fact that the energy decay is more rapid for $H^{\rm M} \simeq 0$ than for finite $H^{\rm M}$, is plausible, since a nonhelical field is less constrained. The transition between the t^{-1} and the slower $t^{-1/2}$ behaviour is not abrupt, but occurs over a certain range $H^{\rm M}/H^{\rm M}_{\rm max} \simeq 0.2$-$0.3$.

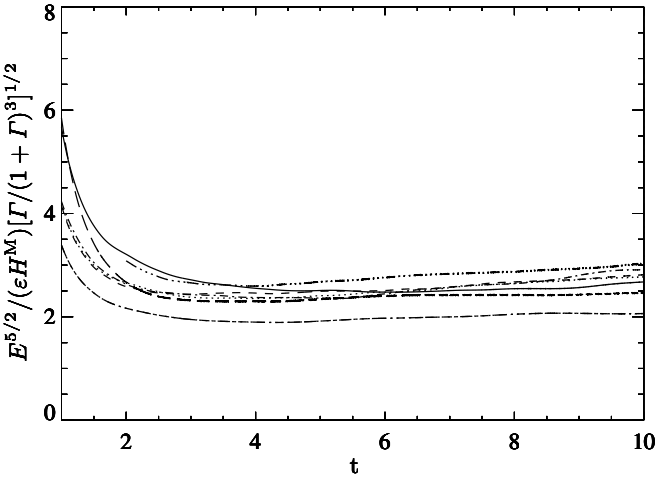


Fig. 4. Relation (15) for various DNS runs of decaying MHD turbulence with 512^3 and 256^3 collocation points and different values of $H^{\rm M}$ (cf. [60]).

6.3 Energy Spectra

The angle-integrated total energy spectrum (7) represents a specific aspect of the statistical properties of turbulence expressed by the two-point correlation tensors of the respective turbulent fields $\langle \boldsymbol{v}(\boldsymbol{r})\boldsymbol{v}(\boldsymbol{r}')\rangle$ and $\langle \boldsymbol{b}(\boldsymbol{r})\boldsymbol{b}(\boldsymbol{r}')\rangle$.

The scaling exponent α of the observed inertial-range power-law $\sim k^{-\alpha}$ permits to draw conclusions about the validity of the different turbulence phenomenologies which yield explicit values of this number, i.e., $\alpha = 5/3$ for K41 and $\alpha = 3/2$ for IK in isotropic turbulence. The difference between the two values is, however, small, making it difficult to distinguish them experimentally. To verify the values of α for the one-dimensional perpendicular and parallel energy spectra predicted by the GS3 phenomenolgy, 5/3 and 5/2 respectively, DNS with a significant mean magnetic field have to be performed, thereby giving up global isotropy.

Figure 5 depicts a time-scatter plot of the total energy spectrum found in DNS of fully-developed decaying MHD turbulence. To bring out the universal shape of the spectrum, abscissa and ordinate are normalized with a generalized dissipation length and by assuming a Kolmogorov-type spectral function (see [47] for details). The curves are compensated with a factor $k^{5/3}$, yielding a horizontal behaviour in the inertial-range where K41 scaling applies.

The spectra seem to support the K41 model and invalidate the IK phenomenology in isotropic MHD turbulence, in agreement with the solar wind observations mentioned before. The Alfvén-effect is nonetheless present in the system, as Fig. 6 shows. With increasing wavenumber this mechanism becomes more efficient, leading to approximate equipartition between kinetic and magnetic energy where the specific value of $E^{\rm M}(k)/E^{\rm K}(k)$ in the dissipation range depends on $\mathsf{Pr_m}$. With decreasing wavenumber the dynamo effect, striving to-

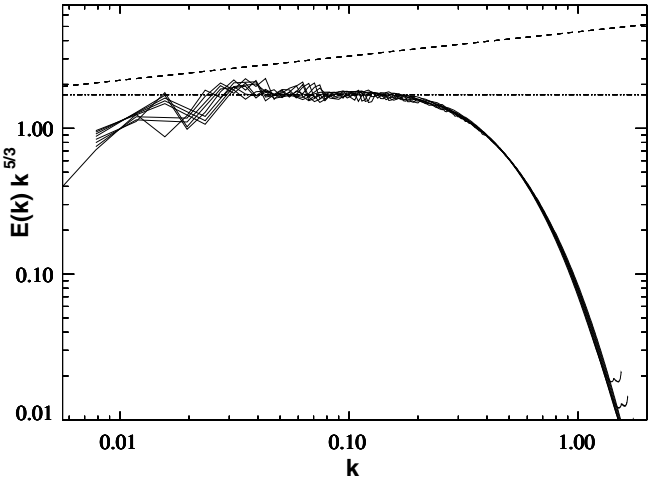


Fig. 5. Time scatter of angle-integrated, normalized and compensated energy spectra in 512^3-DNS of decaying MHD turbulence. Dash-dotted line: K41 scaling, dashed line: IK scaling, (after [47]).

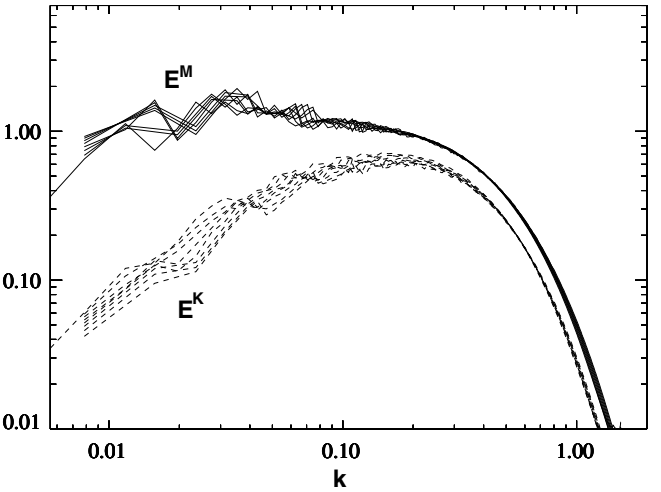


Fig. 6. Kinetic (dashed line) and magnetic (solid line) fractions of the total energy spectra in 512^3-DNS of decaying MHD turbulence, presented in Fig. 5 (after [47]).

wards a magnetically dominated state, takes over which results in the observed decline of the spectral kinetic energy compared to the magnetic one.

6.4 Intermittency

The non-intermittent phenomenologies K41 and IK predict the isotropic structure-function exponents as $\zeta_p^{\mathrm{K41}} = p/3$ and $\zeta_p^{\mathrm{IK}} = p/4$, respectively.

Experimentally observed deviations from these linear relations are attributed to the intermittency of dissipative small-scale structures. To quantify intermittency effects it is necessary to consider the scaling exponents of structure functions of various orders.

Figure 7 shows, as an example, a third order structure function from a 512^3 DNS of a homogeneously turbulent MHD system. The scaling exponent of the approximate power-law observed in the increment interval $\ell \in [20, 200]$ is estimated via the logarithmic derivative depicted in the inset.

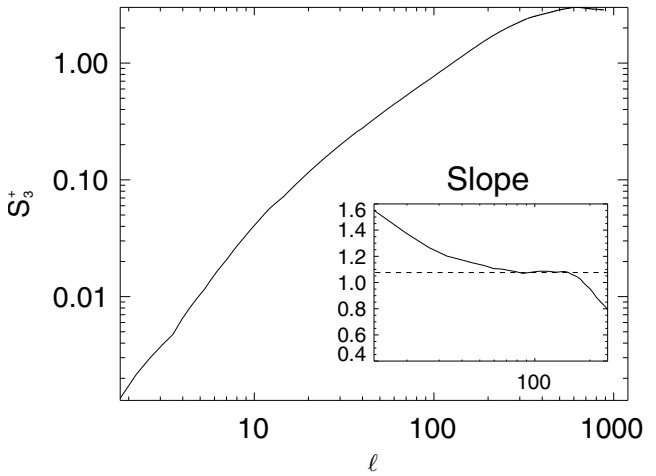


Fig. 7. Energy-normalized third-order z^+-structure function from 512^3-DNS of homogeneous MHD turbulence. The abscissa is normalized with the K41 dissipation length. The inset gives the logarithmic derivative, used to estimate the approximate power-law exponent.

The limited system size of DNS in space and time gives rise to statistical noise spoiling the higher-order structure function scalings. Fortunately, structure functions of different order deviate qualitatively in the same manner from their 'ideal' shape, a property termed 'extended self-similarity' (ESS) [69]. Hence, the error can be reduced and the scaling range substantially enlarged, by considering the relative scaling exponents $\xi_{p,q} = \zeta_p/\zeta_q$ obtained when plotting structure functions of various orders p as functions of a reference structure function of fixed order q whose scaling exponent can be measured with sufficient reliability.

Figure 8 displays examples of ESS curves together with their estimated slopes in units of the S_3^+ scaling exponent. By making use of ESS, DNS scaling exponents up to order eight can be determined with sufficient precision. This allows to test intermittency phenomenologies (see, for example, [18]) of which the Log-Poisson model (see Sect. 4.1) seems to be the most promising candidate.

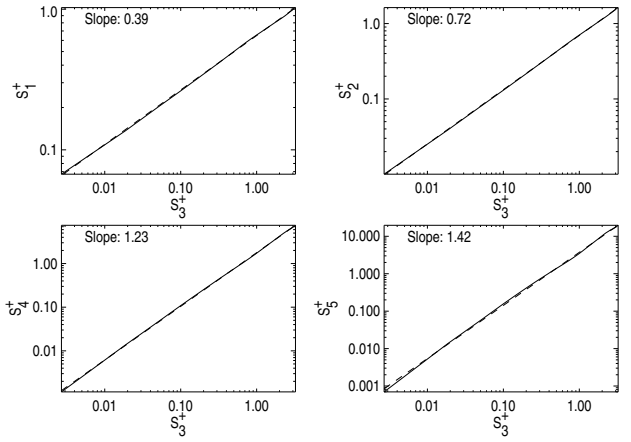


Fig. 8. Extended self-similarity plots of various z^+-structure functions (solid lines) against the corresponding third-order function. The dashed lines show linear least-square fits yielding the logarithmic ESS slopes $\xi_{p,3}$.

A modification of the phenomenology, (13), fits the scaling exponents of the isotropic[6] structure functions up to $p = 8$ in isotropic three-dimensional MHD turbulence with good precision and is moreover consistent with the finding of a K41-like energy spectrum [46]. The MHD model (13) is indicated in Fig. 9 by the dashed line, while the stars denote the corresponding exponents obtained from DNS.

The success of (13) in reproducing the scaling exponents of isotropic structure functions in three-dimensional MHD turbulence without a mean magnetic field corroborates the K41 picture of advectively dominated energy flux. The IK phenomenology, on the other hand, is obviously not valid while for the verification of the GS theory a different diagnostic approach should be taken which will be presented in the following section.

6.5 Local Anisotropy

The GS picture has questioned the local isotropy assumption of K41 and IK in the MHD case. It is therefore reasonable to ask to which degree the local anisotropy introduced by the magnetic field can be detected in globally isotropic MHD turbulence. Since the angle-integrated energy spectrum alone does not provide such information, the scaling properties of structure functions of different order with space increments parallel, $S_p^z(\lambda)$, and perpendicular, $S_p^z(l)$, to the local magnetic field have to be considered.

As a first step, the mean component of the local magnetic field $\overline{\boldsymbol{b}}$ for a fixed spatial scale, l, and at a fixed reference point in space, $\boldsymbol{r}$, has to be determined.

[6] The space increment being taken without reference to a certain direction.

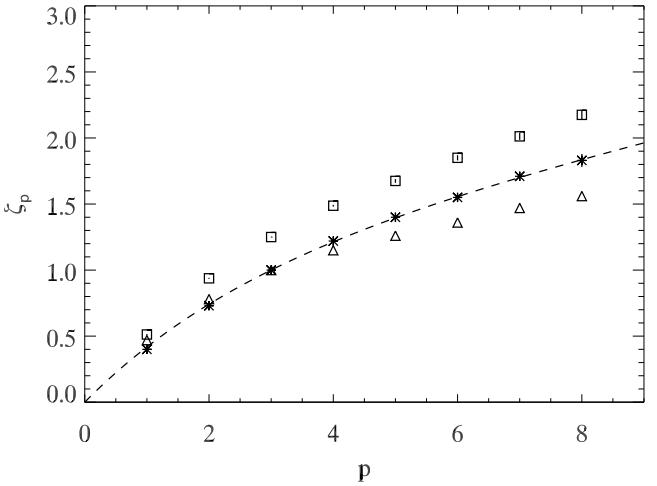


Fig. 9. Scaling exponents of the z^+-structure functions in 512^3-DNS of decaying isotropic MHD turbulence. Triangles: Two-dimensional MHD turbulence [45,70]. Stars: Isotropic *and* perpendicular structure functions. Squares: Parallel structure functions. Dashed line: MHD intermittency model, (13).

This is done by applying a top-hat filter operator $\mathsf{F}(\boldsymbol{r}, \boldsymbol{r}')$ of width l to the turbulent magnetic field $\boldsymbol{b}$,

$$\overline{\boldsymbol{b}}(\boldsymbol{r}) = \mathsf{F} \circ \boldsymbol{b} = \int_V \mathrm{d}^3\boldsymbol{r}' F(|\boldsymbol{r} - \boldsymbol{r}'|)\boldsymbol{b}(\boldsymbol{r}')\,,$$

with the top-hat filter kernel

$$F(|\boldsymbol{r} - \boldsymbol{r}'|) = \begin{cases} 3/(4\pi R^3), & |\boldsymbol{r} - \boldsymbol{r}'| \le R \\ 0, & |\boldsymbol{r} - \boldsymbol{r}'| > R\,. \end{cases}$$

Then, the perpendicular and parallel structure functions are calculated with the space increment towards the second test point being taken perpendicular and parallel to the local $\overline{\boldsymbol{b}}$ vector, respectively. The filter operation is repeated for every increment with the filter width set to the increment length.

The perpendicular and parallel structure functions obtained by this procedure in 512^3-DNS of decaying isotropic MHD turbulence are shown in Fig. 9. One observes a pronounced anisotropy in the statistics parallel and perpendicular to the field, as presumed in the GS picture. Some evidence has also been found in DNS at moderate spatial resolution and with a mean magnetic field component [71,72,73]. In the isotropic turbulence case without a mean magnetic field the perpendicular structure function scalings coincide with the isotropic ones. This is not surprising as the larger value of the parallel structure-function exponent at $p = 2$ versus the perpendicular scaling implies a sharper decline of the parallel energy spectrum compared to the perpendicular one which also explains the less intermittent behaviour of the parallel field exponents.

To examine the dependence of the observed local statistical anisotropy on the magnetic field strength, DNS with a mean magnetic field $\overline{\boldsymbol{b}}$ of varying modulus are considered. Figure 10 displays the obtained scaling exponents of MHD turbulence with a mean field strength $\overline{b} \in \{1, 2, 5\}$ together with the results of isotropic and two-dimensional systems (same symbols as in Fig. 9).

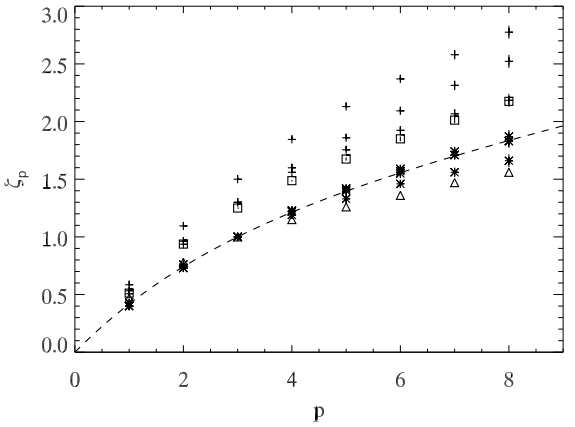


Fig. 10. Scaling exponents of the z^+-structure functions in DNS of decaying isotropic (512^3) and forced mean-field ($512^2 \times 128$, 256^3) MHD turbulence. Same symbols as in Fig. 9. Additional stars: Perpendicular structure function scalings for $\overline{b}$ = 1,2,5 (top to bottom). Crosses: Parallel scaling exponents (increasing $\overline{b}$ bottom to top). Dashed line: MHD intermittency model, (13).

The crosses denoting the parallel scaling data (mean field increase from bottom to top) indicate that the small-scale structures, as expected, become less and less intermittent in the field-parallel direction as the mean magnetic field gets stronger. The flow loses its turbulent character in the direction of $\boldsymbol{b}_0$. This can formally be seen in expression (9) where a growth of b_0 leads to an increase of the dissipation length, effectively reducing the Reynolds numbers of the corresponding flow. The associated physical mechanism is the alignment of the dissipative structures, micro-current and vorticity sheets, along the mean field.

Perpendicular to $\overline{\boldsymbol{b}}$ (stars in Fig. 10) the mean field effect is quite weak. Nevertheless, for the largest value $\overline{b} = 5$ the expected transition towards two-dimensional incompressible MHD turbulence is observed to occur. This state corresponds to the limiting case of a reduced-MHD description of a compressible three-dimensional MHD flow permeated by a strong mean magnetic field (see, e.g., [21]).

7 Summary

A brief overview of the current phenomenological framework for three-dimensional homogeneous MHD turbulence has been given, combined with recent numerical and experimental results.

Within the idealized statistical approach, the K41 phenomenology can be considered as a solid foundation for further theoretical developments. The presented numerical results show that the IK picture does not seem to hold, being at variance with the findings of a K41 energy spectrum although the Alfvén effect is present in the turbulent system. This is in agreement with the observation of approximate K41 energy-spectra in the solar wind.

The Log-Poisson intermittency phenomenology has proven to be a valuable tool for linking observed isotropic structure-function scaling exponents with a physical model of the turbulent dynamics. As the spatial dimension of the smallest dissipative structures can easily be determined by DNS, the second free parameter of the model allows to test different hypotheses on the energy cascade mechanism. This diagnostic is more general than the energy spectra alone and furthermore allows to quantify intermittency effects.

It is strong evidence in favour of the K41 phenomenology that the MHD intermittency model which assumes K41-like nonlinear energy dynamics leads to very good agreement with the obtained numerical data. By contrast, assuming Alfvén-wave cascade dynamics, the quintessence of the IK picture, yields a Log-Poisson model that does not agree with the DNS results. This leads to the conclusion that the isotropic statistical properties of homogeneous three-dimensional MHD turbulence without a mean magnetic field are well described within the K41 framework.

Local anisotropy in the nonlinear dynamics introduced by the presence of the magnetic field is addressed by the GS phenomenology in its three variants. This picture of MHD turbulence predicts three different consecutive inertial ranges in the perpendicular dynamics due to a strengthening of nonlinear interactions with spectral wavenumber. Though there is some support for the 'strong turbulence' part, stemming from DNS with a mean magnetic field, the validity of the GS hypotheses have still to be scrutinized with high-resolution numerical data.

References

1. Y.B. Zeldovich, A.A. Ruzmaikin, D.D. Sokoloff: *Magnetic Fields In Astrophysics* (Gordon and Breach Science Publishers, New York, 1983)
2. D. Biskamp: *Magnetic Reconnection in Plasmas* (Cambridge University Press, Cambridge, 2000)
3. M.L. Goldstein, D.A. Roberts: Physics of Plasmas **6**(11), 4154–4160 (1999)
4. E.G. Zweibel: Physics of Plasmas **6**(5), 1725–1731 (1999)
5. J.F. Hawley, S.A. Balbus: Physics of Plasmas **6**(12), 4444–4449 (1999)
6. G.K. Batchelor: *The Theory of Homogeneous Turbulence* (Cambridge University Press, Cambridge, 1993)
7. R.H. Kraichnan: Journal of Fluid Mechanics **5**, 497–543 (1959)

8. R.H. Kraichnan: Physics of Fluids **8**(4), 575–598 (1965)
9. S.F. Edwards: Journal of Fluid Mechanics **18**, 239–273 (1964)
10. J.R. Herring: Physics of Fluids **8**(12), 2219–2225 (1965)
11. S.A. Orszag: Journal of Fluid Mechanics **41**(2), 363–386 (1970)
12. R.H. Kraichnan: Journal of Fluid Mechanics **47**(3), 513–524 (1971)
13. U. Frisch, M. Lesieur, A. Brissaud: Journal of Fluid Mechanics **65**(1), 145–152 (1974)
14. H.A. Rose, P.L. Sulem: Journal de Physique **39**(5), 441–483 (1978)
15. M. Lesieur: *Turbulence in Fluids* (Kluwer Academic Publishers, Dordrecht, 1997)
16. D.C. Leslie: *Developments in the Theory of Turbulence* (Clarendon Press, Oxford, 1983)
17. A. Pouquet, U. Frisch, J. Léorat: Journal of Fluid Mechanics **77**(2), 321–354 (1976)
18. U. Frisch: *Turbulence* (Cambridge University Press, Cambridge, 1996)
19. J.O. Hinze: *Turbulence* (McGraw-Hill, New York, 1987)
20. D.I. Braginskii: Reviews of Plasma Physics **1**, 205–311 (1965)
21. D. Biskamp: *Nonlinear Magnetohydrodynamics* (Cambridge University Press, Cambridge, 1993)
22. L. Woltjer: Proceedings of the National Academy of Sciences **44**(9), 833–841 (1958)
23. H.K. Moffatt: Journal of Fluid Mechanics **35**(1), 117–129 (1969)
24. W.M. Elsässer: Physical Review **79**, 183 (1950)
25. A.N. Kolmogorov: Proceedings of the Royal Society A **434**, 9–13 (1991). [Dokl. Akad. Nauk SSSR, 30(4), 1941]
26. A.N. Kolmogorov: Proceedings of the Royal Society A **434**, 15–17 (1991). [Dokl. Akad. Nauk SSSR, 32(1), 1941]
27. L.F. Richardson: *Weather Prediction by Numerical Process* (Cambridge University Press, Cambridge, 1922)
28. P.S. Iroshnikov: Soviet Astronomy **7**, 566–571 (1964). [Astron. Zh., 40:742, 1963]
29. R.H. Kraichnan: Physics of Fluids **8**(7), 1385–1387 (1965)
30. H. Politano, A. Pouquet, P.L. Sulem: Physics of Fluids B **1**(12), 2330–2339 (1989)
31. R. Grappin, A. Pouquet, J. Léorat: Astronomy and Astrophysics **126**, 51–58 (1983)
32. H. Politano, A. Pouquet: Physical Review E **57**(1), R21–R24 (1998)
33. H. Politano, A. Pouquet: Geophysical Research Letters **25**(3), 273–276 (1998)
34. J.V. Shebalin, W.H. Matthaeus, D. Montgomery: Journal of Plasma Physics **29**(3), 525–547 (1983)
35. R. Grappin: Physics of Fluids **29**(8), 2433–2443 (1986)
36. S. Sridhar, P. Goldreich: Astrophysical Journal **432**, 612–621 (1994)
37. P. Goldreich, S. Sridhar: Astrophysical Journal **438**, 763–775 (1995)
38. P. Goldreich, S. Sridhar: Astrophysical Journal **485**, 680–688 (1997)
39. D. Montgomery, W.H. Matthaeus: Astrophysical Journal **447**, 706–707 (1995)
40. C.S. Ng, A. Bhattacharjee: Astrophysical Journal **465**, 845–854 (1996)
41. S. Galtier, S.V. Nazarenko, A.C. Newell, A. Pouquet: Journal of Plasma Physics **63**(5), 447–488 (2000)
42. S.V. Nazarenko, A.C. Newell, S. Galtier: Physica D **152–153**, 646–652 (2001)
43. K.R. Sreenivasan, R.A. Antonia: Annual Review of Fluid Mechanics **29**, 435–472 (1997)
44. L.F. Burlaga: Journal of Geophysical Research **96**(A4), 5847–5851 (1991)
45. H. Politano, A. Pouquet, V. Carbone: Europhysics Letters **43**(5), 516–521 (1998)
46. W.C. Müller, D. Biskamp: Physical Review Letters **84**(3), 475–478 (2000)
47. D. Biskamp, W.C. Müller: Physics of Plasmas **7**(12), 4889–4900 (2000)

48. A.N. Kolmogorov: Journal of Fluid Mechanics **13**, 82–85 (1962)
49. Z.S. She, E. Lévêque: Physical Review Letters **72**(3), 336–339 (1994)
50. B. Dubrulle: Physical Review Letters **73**(7), 959–962 (1994)
51. Z.S. She, E.C. Waymire: Physical Review Letters **74**(2), 262–265 (1995)
52. H. Politano, A. Pouquet: Physical Review E **52**(1), 636–641 (1995)
53. R. Grauer, J. Krug, C. Marliani: Physics Letters A **195**, 335–338 (1994)
54. A.N. Kolmogorov: Doklady Akademiia Nauk SSSR **31**, 538–540 (1941)
55. M.R. Smith, R. J. Donnelly, N. Goldenfeld und W. F. Vinen: Physical Review Letters **71**(16), 2583–2586 (1993)
56. S. Galtier, H. Politano, A. Pouquet: Physical Review Letters **79**(15), 2807–2810 (1997)
57. G.K. Batchelor: Physics of Fluids Supplement II **12**(3), 233–239 (1969)
58. T. Hatori: Journal of the Physical Society of Japan **53**(8), 2539–2545 (1984)
59. A.C. Ting, W.H. Matthaeus, D. Montgomery: Physics of Fluids **29**(10), 3261–3274 (1986)
60. D. Biskamp, W.C. Müller: Physical Review Letters **83**(11), 2195–2198 (1999)
61. R.J. Leamon, C.W. Smith, N.F. Ness, W.H. Matthaeus, H.K. Wong: Journal of Geophysical Research **103**(A3), 4775–4787 (1998)
62. C. Canuto, M.Y. Hussaini, A. Quarteroni, T.A. Zang: *Spectral Methods in Fluid Dynamics* (Springer-Verlag, New York, 1988)
63. S.A. Orszag: Studies in Applied Mathematics **51**(3), 253–259 (1972)
64. A. Vincent, M. Meneguzzi: Journal of Fluid Mechanics **225**, 1–20 (1991)
65. J.R. Herring, S.A. Orszag, R.H. Kraichnan, D.G. Fox: Journal of Fluid Mechanics **66**(3), 417–444 (1974)
66. M.M. Mac Low, R.S. Klessen, A. Burkert, M.D. Smith: Physical Review Letters **80**(13), 2754–2757 (1998)
67. J.M. Stone, E.C. Ostriker, C.F. Gammie: Astrophysical Journal **508**, L99–L102 (1998)
68. D. Biskamp und H. Welter: Physics of Fluids B **1**(10), 1964–1979 (1989)
69. R. Benzi, S. Ciliberto, R. Tripiccione, C. Baudet, F. Massaioli, S. Succi: Physical Review E **48**(1), R29–R32 (1993)
70. D. Biskamp, E. Schwarz: Physics of Plasmas **8**(7), 3282–3292 (2001)
71. J. Cho, E.T. Vishniac: Astrophysical Journal **539**, 273–282 (2000)
72. J. Cho, A. Lazarian, E.T. Vishniac: Astrophysical Journal **564**, 291–301 (2002)
73. L.J. Milano, W.H. Matthaeus, P. Dmitruk: Physics of Plasmas **8**(6), 2673–2681 (2001)

Coronal Heating and Reduced MHD

Sean Oughton[1], Pablo Dmitruk[2], and William H. Matthaeus[2]

[1] Department of Mathematics, University of Waikato, Hamilton, New Zealand
[2] Bartol Research Institute, University of Delaware, Newark, DE 19716, USA

Abstract. We review the use of reduced magnetohydrodynamics (RMHD) in coronal heating models, with particular emphasis on models for magnetically open regions. A brief review of the nature of the coronal heating problem is presented first, followed by detailed discussion regarding the assumptions and features of RMHD and its applicability to the dynamics of the solar corona. We then review a class of heating models based on quasi-2D turbulent cascades driven by low-frequency Alfvén waves.

1 Introduction

The fact that the solar corona is over a hundred times hotter than the visible solar surface was first established in 1939 [1,2]. Observations indicate that coronal temperatures are typically $\sim 10^6\,\mathrm{K}$, while the temperature of the photosphere is $\approx 6000\,\mathrm{K}$. Although much effort has been expended in attempts to explain this "coronal heating problem", the mechanism underlying the heating process is still not known; see, e.g., [3,4,5,6,7]. In this paper we review first the role of reduced magnetohydrodynamics (RMHD) in coronal dynamics and second a class of heating models which employ RMHD processes in a central way.

Note that it is the *mechanism* of coronal heating which has remained unidentified, not the energy source for the heating. An ample reservoir of energy is available to heat the corona, in the form of the (convectively) turbulent photospheric and subphotospheric motions. The associated energy flux is essentially equivalent to the Poynting flux $\boldsymbol{S} = \boldsymbol{E} \times \boldsymbol{B}$, where the electric field is that induced by the convective motions [8]. Thus, the "problem" in the coronal heating problem is primarily concerned with how to transport this energy up into the corona and then convert it into heat within a few solar radii ($R_{\mathrm{sun}} = 700\,\mathrm{Mm}$). Recent observations [9,10,11,12,13] emphasize that the bulk of the heating needs to occur within this distance, which is a rather strong constraint.

Table 1 lists some of the heating models which have been proposed over the last 60 years. Although no one heating theory has been entirely successful, substantial progress in understanding the situation has nonetheless occurred. A key recognition is that the corona is magnetically dominated, in the sense that there is a large (often approximately vertical) mean magnetic field ($\boldsymbol{B}_0$) throughout it, with the fluctuations in the velocity and magnetic fields being much smaller in magnitude [6,14,15]. As discussed in the subsequent sections, this physical situation of a strong $\boldsymbol{B}_0$ with small (but finite) amplitude fluctuations can be exploited to simplify the full 3D MHD equations to those of reduced MHD. Figure 1 is a sketch depicting the magnetic structure of the corona.

E. Falgarone and T. Passot (Eds.): LNP 614, pp. 28–55, 2003.
© Springer-Verlag Berlin Heidelberg 2003

Table 1. Some candidate heating mechanisms for the solar corona (after Ulmschneider 1996 [5]; see also [4])

Energy Carrier	Dissipation Mechanism
Acoustic waves	
Slow mode MHD waves	Shocks
Longitudinal MHD waves	
Fast mode MHD waves	Landau damping
	Phase mixing
Magnetoacoustic surface waves	Mode coupling
	Resonant absorption
	Mode coupling
	Phase mixing
Alfvén waves	Resonance heating
	Viscous heating
	Turbulent heating
	Landau damping
Current sheets	Reconnection (turbulent/wave heating)
Magnetic field	Nanoflares

Most of the approaches listed in Table 1 can be characterized as *direct*, in the sense that the energy of the carrier is converted directly into heat without passing through any other channel. On the whole, direct methods have so far proved to be either insufficient or inadequate (e.g., [4]), thereby prompting consideration of indirect methods. In an *indirect* heating method the free energy source in the corona is first converted into some other form of excitation, which is itself subsequently converted into heat energy.

Various other divisions of the heating models have been employed, for example wave versus turbulence and acoustic versus magnetic. In the latter case, the magnetic mechanisms are often subdivided into either AC (aka wave) mechanisms, wherein the field guides the (dissipative) waves, and DC (aka current) mechanisms where it is the dissipation of the energy stored in the magnetic field which provides the heating [3,4,8,16,17]. In this review we pay particular attention to an indirect turbulence heating model, where the (quasi-2D) turbulence is driven by low-frequency Alfvén waves.

Note that the traditional schism between wave models and turbulence models can be misleading and/or over-restrictive. A similarly misleading schism has also featured in models for the transport and evolution of solar wind fluctuations [18,19,20,21]. However, in both the corona and the solar wind, observations indicate that the systems are strongly dynamic with wave and turbulence processes

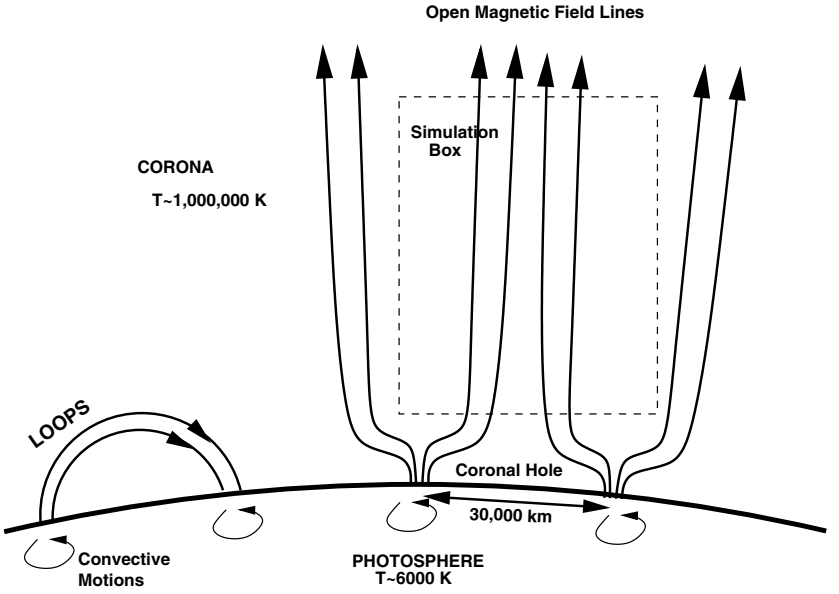


Fig. 1. Cartoon sketch of the (magnetic) structure of the corona. Shown are a few magnetic field-lines delineating coronal loops and coronal holes. Also indicated are the convective motions which power the atmospheric heating. Plasma heating in both the magnetically open and the magnetically closed regions remains unexplained. Note that at photospheric heights the magnetic flux bundles are typically spatially well separated. In contrast, the coronal field is essentially space filling, with adjacent flux bundles subject to shear due to the independent motion of the photospheric "footpoints"

each playing important roles, and, moreover, that activity occurs over a wide range of spatial and temporal scales amenable to MHD-based descriptions. For example, in the corona measurements from various satellite and ground-based instruments show the presence of waves and substantial transverse (to the local vertical) structuring (see, e.g., [7,22]). Observations of counter-propagating waves [11] and reconnection [23] have also been reported, although in such cases a unique interpretation of the data is often not possible. Taken together, the observations imply that coronal dynamics involves both wave effects and nonlinear MHD effects, such as turbulent reconnection and cascades.

The remainder of the paper is organised as follows. Section 2 contains a detailed discussion of the assumptions and characteristics of reduced MHD, along with a critique of their relevance to the corona. In Sect. 3 we review several coronal heating models which employ RMHD, focusing primarily on models for magnetically open regions. The final section summarises our conclusions. Two appendices, one a brief history of RMHD derivations and the other on the subtle relationship between RMHD fluctuations and solutions of the RMHD equations, close the paper.

2 Reduced MHD

The equations of reduced MHD, first derived by Strauss [24], can be obtained by considering the strong $\boldsymbol{B}_0$, low-frequency limit of the full 3D MHD equations. Note that they are *not a linearization* of the MHD equations—the nonlinear terms are always important, by construction. The RMHD equations have also been derived from various other perspectives, some of which are summarised in Appendix A.

In this section, we begin by stating the RMHD equations and then derive the timescale condition which determines whether or not a fluctuation is RMHD in character. Section 2.2 is a discussion of the nature of cascades in RMHD turbulence and their relation to cascades in full 3D turbulence. Section 2.3 reviews the relevance of RMHD to coronal dynamics.

2.1 Equations and Timescale Conditions

The usual primary assumption of RMHD is that there is a strong uniform magnetic field, $\boldsymbol{B}_0 = B_0\hat{\boldsymbol{z}}$, present [24,25,26]. The field is strong in the sense that the amplitudes of the velocity and magnetic field fluctuations are small. Since the full 3D MHD equations have terms like $\boldsymbol{B}_0 \cdot \nabla \boldsymbol{b}$, these terms will be much larger than the nonlinear ones (e.g., $\boldsymbol{v} \cdot \nabla \boldsymbol{v}$), unless gradients along $\boldsymbol{B}_0$ are restricted to be small, as is the case for RMHD. When this situation is enforced, the natural coordinates to use involve a rescaling of the parallel (z) coordinate to be purely large-scale and of B_0 to be O(1).

The RMHD equations are then conveniently written—using non-dimensionalized (Alfvén speed) units—as equations for the evolution of the fluctuations in the fluid vorticity $\omega(x, y, z, t)$ and the magnetic vector potential $a(x, y, z, t)$:

$$\left(\frac{\partial}{\partial t} + \boldsymbol{v}\cdot\nabla_\perp\right)\omega = \boldsymbol{b}\cdot\nabla_\perp j + \nu\nabla_\perp^2\omega + B_0\frac{\partial j}{\partial z}, \tag{1a}$$

$$\left(\frac{\partial}{\partial t} + \boldsymbol{v}\cdot\nabla_\perp\right)a = \eta\nabla_\perp^2 a + B_0\frac{\partial \psi}{\partial z}, \tag{1b}$$

where $\nabla_\perp = (\partial_x, \partial_y, 0)$, $\boldsymbol{v} = \nabla \times \psi\hat{\boldsymbol{z}}$, $\boldsymbol{b} = \nabla \times a\hat{\boldsymbol{z}}$, $\omega = -\nabla_\perp^2\psi$, $j = -\nabla_\perp^2 a$, $\nabla_\perp \cdot \boldsymbol{v} = 0$, and the other symbols have their usual meanings (see, e.g., [25]). Since the equations are written using Alfvén speed units, B_0 is actually the (large-scale) Alfvén speed. However, it is often convenient to still refer to it as the mean field. We emphasize that the z coordinate has been rescaled such that only large-scale variations with z are permissible. For example, the final term in (1a) can be rewritten as

$$\left(\frac{B_0}{\epsilon}\right)\left(\epsilon\frac{\partial}{\partial z}\right)j, \tag{2}$$

where a small parameter ϵ appears explicitly (see below). The two terms in brackets then correspond directly to the actual physical mean field and the gradient operator along it [25].

It is evident that in RMHD the velocity is incompressible and that setting $B_0 = 0$ would yield precisely the equations of incompressible 2D MHD, on a set of independent z-planes. On the other hand, if the only terms retained are the time derivatives and the terms involving B_0, the solutions are parallel-propagating (transverse) Alfvén waves. Thus it is clear that the B_0 terms provide the explicit coupling between the otherwise independent planes of 2D excitation.

A related point is that while $\boldsymbol{v}$ and $\boldsymbol{b}$ depend upon all three spatial coordinates (although as noted the z dependence is "slow") the amplitudes of these excitations are confined to the planes perpendicular to $\boldsymbol{B}_0$.[1] In this sense, the RMHD equations are complementary to systems employing $2\frac{1}{2}$D geometry, wherein vector fields have three components but depend on only two spatial coordinates.

A derivation of the equations is now sketched. Consider a fluctuating magnetofluid[2] threaded by a uniform magnetic field $\boldsymbol{B}_0$. The dynamics of the fluctuations is given by the usual 3D MHD equations in the presence of a mean field. In order to compare the characteristic timescales associated with these fluctuations, it is useful to employ a Fourier decomposition. Accordingly the mean square magnetic fluctuation $\langle b^2 \rangle$ may be decomposed into a spectral distribution. The magnetic energy spectrum, i.e., the distribution of energy over wavenumber magnitude k, may be designated as $E_b(k)$, and satisfies

$$\frac{\langle b^2 \rangle}{2} = \int_0^\infty E_b(k)\,\mathrm{d}k. \tag{3}$$

The contribution to the magnetic energy density due to fluctuations near spatial scale $1/k$ may be estimated in the usual way as $b_k^2 = kE_b(k)$.

In incompressible MHD each Fourier mode has two[3] timescales associated with it:

$$\tau_{\mathrm{A}}(\boldsymbol{k}) = \frac{1}{|\boldsymbol{k}\cdot\boldsymbol{B}_0|} \equiv \frac{1}{|k_\parallel||\boldsymbol{B}_0|}\ , \tag{4a}$$

$$\tau_{\mathrm{NL}}(\boldsymbol{k}) = \frac{1}{kb_k}\ , \tag{4b}$$

where these are respectively referred to as the Alfvén (or wave) timescale and the (direction-averaged) modal nonlinear timescale; $k_\parallel$ is the component of the Fourier wavevector parallel to $\boldsymbol{B}_0$. Two points regarding the Alfvén timescale are

[1] This is not quite the full story. For a derivation beginning from the incompressible equations [25], it can be shown that, to the same order as (1a) and (1b), RMHD also involves fluctuations whose amplitudes are purely parallel to $\boldsymbol{B}_0$. For a derivation beginning with compressible MHD [26] the parallel fluctuations are down by a factor equal to the turbulent Mach number (interestingly, this relative ordering of the parallel amplitude also emerges in incompressible and compressible simulations of the *full* MHD equations [27]). However, in both cases these parallel fluctuations are dynamically passive and thus of reduced interest [25,26].

[2] Density fluctuations are likely to be of interest along with the velocity and magnetic fluctuations.

[3] Where the viscous and resistive timescales are not considered.

noteworthy. First, it is the time it takes an Alfvén wave to propagate a parallel wavelength for the mode (modulo a factor of 2π), and is thus quite distinct from the box-crossing timescale of a wave, which for a box of length $L_{\rm box}$ would be $\tau_{\rm box} = L_{\rm box}/B_0$, independent of $\boldsymbol{k}$. Second, it depends strongly on the orientation of its wavevector but is *in*dependent of the energy of the fluctuation.

The nonlinear timescale as defined in (4b) is local and direction-averaged (over all $\boldsymbol{k}$ with the same $|\boldsymbol{k}|$). Consequently, it cannot be viewed as anything other than a reasonable approximation in the present context, which is expected to be highly anisotropic. Its analogue for hydrodynamic turbulence is, however, familiar and reliable in the context of incompressible homogeneous hydrodynamic turbulence [28], and therefore it is believed that this type of estimate is a good starting point for MHD as well. It is also useful to define the *global* nonlinear timescale $\tau_{\rm NL} = \lambda_\perp/\delta b$, where $\lambda_\perp$ is a lengthscale characteristic of the turbulent (nonlinear) dynamics[4] and $\delta b = \sqrt{\langle b^2 \rangle}$ is the root mean square magnetic fluctuation (assumed $\approx \delta v$).

Forming the ratio of the *modal* timescales yields the central parameter of RMHD:

$$\epsilon_{\rm RMHD}(\boldsymbol{k}) = \frac{\tau_{\rm NL}(\boldsymbol{k})}{\tau_{\rm A}(\boldsymbol{k})} = \frac{k_\parallel}{k}\frac{B_0}{b_k}. \tag{5}$$

This is the ratio of the timescale associated with wave-like activity, which depends upon parallel structure, to the timescale of nonlinear activity, which does not depend upon direction at the current level of approximation. Physically, it appears likely that the set of (Fourier) modes which satisfy $\tau_{\rm NL}(\boldsymbol{k}) \lesssim \tau_{\rm A}(\boldsymbol{k})$—that is those modes for which the wave timescale is *not* dominant—will behave quite differently from those for which it is. Thus, we partition the modes on the basis of the inequality $\epsilon_{\rm RMHD}(\boldsymbol{k}) \lesssim 1$, and investigate the governing equations for the two sets of modes.[5] As it turns out, the modes which satisfy this inequality are the RMHD ones.

Note that the *critical balance* condition employed by Goldreich and Sridhar [29] in a paper on strong MHD turbulence is essentially the condition $\epsilon_{\rm RMHD}(\boldsymbol{k}) = 1$, and is thus related to the boundary between RMHD and non-RMHD fluctuations (cf. Appendix B and Fig. 7a).

Up to this point no major assumptions about the nature of the fluctuations, or the geometry of the system, have been made. As noted above, the usual primary assumption of RMHD is that $\boldsymbol{B}_0$ is a strong field, meaning that its energy density is much greater than that of the fluctuations. Consequently, for each Fourier mode $B_0/b_k \gg 1$, and satisfaction of $\epsilon_{\rm RMHD}(\boldsymbol{k}) \lesssim 1$ requires that $k_\parallel \ll k$. In other words, the RMHD modes have wavevectors which are approximately perpendicular to $\boldsymbol{B}_0$. This leads to two approximations employed frequently in RMHD, namely $k \approx k_\perp$ and $\tau_{\rm NL}(\boldsymbol{k}) \approx 1/(k_\perp b_k)$.

[4] The "$\perp$" subscript is included since we have in mind specialisation to the RMHD case.

[5] For systems which are not in steady-state the partitioning is likely to be time-dependent.

The Strauss [24] and Montgomery [25] derivations of the RMHD equations are perturbation expansions in distinct but related small parameters. Respectively, $\epsilon_{\rm Strauss} = \ell_\perp/\ell_\parallel$ and $\epsilon_{\rm Mont} = \delta b/B_0$, where the ℓs are characteristic lengthscales in the directions indicated by their subscripts (relative to $\boldsymbol{B}_0$). Note that these small parameters are based on global, not modal, quantities. RMHD provides the leading-order dynamics for each of these derivations (cf. Appendix A).

Subsequently, Zank and Matthaeus [26] noted that a more rigorous derivation is also possible. They showed that the 3D *compressible* MHD equations have leading-order solutions which obey the RMHD equations, provided that (i) the sonic Mach number, $M_{\rm s} = \delta v/c_{\rm s}$, is small, (ii) the plasma beta,[6] $\beta_{\rm p}$, is at most of order unity, (iii) characteristic lengths along B_0 are much longer than those perpendicular to it, and (iv) all high-frequency[7] modes have been eliminated. It is condition (iv) which ties their derivation of the RMHD equations to the timescale-based discussion of the RMHD modes presented above.

Note that there is a subtle but important distinction to be made between the set of RMHD *fluctuations* and solutions of the RMHD *equations*. It is quite possible to solve the RMHD equations using initial conditions which include non-RMHD fluctuations, e.g., in simulations with rectangular domains in k-space (see Fig. 7a). The solutions to the equations at later times would, presumably, also include contributions from non-RMHD fluctuations. Several questions then arise as to the consistency of the RMHD equations. For example: Do they adequately capture the physics of non-RMHD fluctuations along with that of the RMHD fluctuations? If an (initial) solution to the RMHD equations contains only RMHD modes, how rapidly will non-RMHD modes be generated? Do RMHD solutions "escape" from the range of validity of the equations themselves? Concerns such as these have prompted some implementations of RMHD to include *ad hoc*—and perhaps unnecessary—parallel dissipation terms which damp energy production at high $k_\parallel$. Further discussion regarding this point is presented in Appendix B.

2.2 Anisotropy and Cascades

The RMHD equations are manifestly anisotropic, containing as they do a strong mean magnetic field with the (dynamically important) fluctuations strictly perpendicular to it. Associated with this is the lengthscale anisotropy between the x and y coordinates and the slowly-varying z coordinate These anisotropies have important consequences for the development and interpretation of RMHD energy cascades.

In RMHD the cascade is such that excitations are principally driven towards high *perpendicular* wavenumbers [25,30,31]. That is, the evolution in physical space involves the development of structure with finer and finer scale perpendicular to $\boldsymbol{B}_0$ (with the smallest scales attained set by the Reynolds numbers).

[6] The plasma beta factor, $\beta_{\rm p}$, is the usual ratio of thermal and magnetic pressures.
[7] Where high-frequency is defined to be on the acoustic timescale or faster.

This inherent feature of RMHD is often identified by referring to a *strong perpendicular cascade.* It should be kept in mind, however, that it is the parallel cascade which is substantially weakened in RMHD, rather than the perpendicular cascade being strengthened [30,31,32,33,34].

To understand the weakening of the parallel cascade it is helpful to consider cascades in fully 3D (i.e., not reduced) MHD turbulence when $\delta b/B_0$ is small. Considering the turbulent fluctuations as Alfvén wavepackets, the strength of B_0 means that the packets are rapidly propagating. Hence, the (nonlinear) interaction time for two counter-propagating wavepackets is much reduced, leading to a weakened cascade. There is, however, an escape clause, since this argument ignores the anisotropy of the wave dispersion relation $\omega_{\rm A} = \boldsymbol{k} \cdot \boldsymbol{B}_0$. If one or more of the wavepackets has large spatial extent (i.e., has a characteristic $k_\parallel$ which is small)—as is the case for RMHD fluctuations—then even though they are moving rapidly through each other, the interaction time for the packets can still be long, and in particular of order $\tau_{\rm NL}$ rather than $1/(kB_0)$. Consequently the cascade associated with these interactions is only weakly affected by the strong B_0. Shebalin et al. [33] presented a "weak turbulence" analysis of such interactions which revealed that *strict* resonant driving occurs only when a propagating mode interacts with a non-propagating (i.e., 2D) mode of opposite cross helicity (see also [35,36]). In such cases the driven mode has exactly the same $k_\parallel$ as that driving mode which is propagating. Thus, in the weak turbulence limit there is strong perpendicular spectral transfer but *no parallel spectral transfer.* The arguments can be generalised to allow for broadened resonant interactions [30,31], and one finds that the interaction of an arbitrary propagating mode with an RMHD mode also leads to strong perpendicular cascade.

Additional support for suppressed parallel cascade with still strong perpendicular cascade comes from analytic work, and compressible and incompressible simulations with a strong mean field [25,27,30,31,32,33,37,38,39,40,41,42]. Moreover, these studies indicate that turbulence which is initially isotropic dynamically reorganises to favor RMHD-type fluctuations. For example, one can examine the evolution of mean wavenumbers computed parallel and perpendicular to B_0 using simulations. Figure 2, reproduced from [31], displays results from two (unforced) incompressible 3D MHD simulations with different values of B_0. When $B_0 = 0$, the mean $k_\perp$ and $k_\parallel$ are essentially the same at all times, showing a monotonic increase as the nonlinear interactions cause transfer of energy to smaller scales, followed by a gradual decline as the turbulence decays. Such isotropy is of course expected when there is no preferred direction. For the $B_0 = 4$ case, in contrast, the mean $k_\parallel$ remains almost constant, indicating that the cascade in this direction has been strongly suppressed. The mean $k_\perp$, however, behaves in the same way as it did in the $B_0 = 0$ simulation.[8]

There is a technical point about the full-MHD perpendicular cascade that warrants mention at this point, namely that it consists of two distinct parts

[8] The maximum value attained, however, is somewhat less than in the isotropic case since the initial conditions contain both RMHD and non-RMHD fluctuations and the latter are influenced by B_0 while the anisotropy is developing.

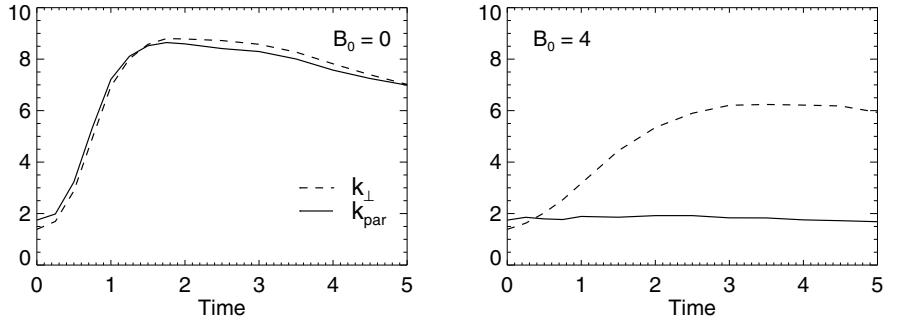


Fig. 2. Mean parallel and perpendicular wavenumbers as a function of time for two fully 3D MHD simulations. The simulations are unforced and have the same initial conditions and Reynolds numbers but different values of B_0. A value of $B_0 = 1$ corresponds to equipartition of the energy in the initial turbulent fluctuations and that in the mean field. Note the lack of evolution in $k_\parallel$ for the $B_0 = 4$ case (figure courtesy of [31])

[30,31]. The first part involves the self-interactions of only the RMHD modes. As these are in effect "zero-frequency" modes, the interactions are (trivially) always resonant. Perpendicular cascade proceeds unimpeded for RMHD modes. Moreover, because the resonance condition is trivial, it does not impose any additional restrictions on parallel spectral transfer, provided that excitations remain within the RMHD region $\epsilon_{\mathrm{RMHD}} \lesssim 1$. Thus, RMHD spectral transfer is essentially isotropic although its region of applicability is emphatically not (cf. Fig. 7). The second type of perpendicular cascade in full-MHD comes from modes that are outside the RMHD region. Sufficiently far outside the $\epsilon_{\mathrm{RMHD}} = 1$ boundary the resonance conditions described by Shebalin et al. [33,35,38,42,43] prevail, and the only surviving incompressible couplings are those that maintain constant $k_\parallel$. This resonant cascade of high-frequency modes to higher $k_\perp$ is nonlocal, since it relies upon driving by (inherently) low-frequency RMHD modes.

2.3 Relevance in the Corona

Here we briefly summarise some of the theoretical and observational support for using RMHD in studies of coronal dynamics.

As noted above, an important characteristic of both coronal holes and coronal loops is that they are permeated by a strong large-scale magnetic field. Moreover, the large-scale field can be approximated as uniform over significant distances (e.g., fractions of a loop length).

Observations of coronal quantities [11,44,45,46] in magnetically open regions suggest that $\delta v \approx 25$–$35\,\mathrm{km\,s^{-1}}$, $c_\mathrm{s} \approx 115\,\mathrm{km\,s^{-1}}$, and that Alfvén speeds are in excess of $1000\,\mathrm{km\,s^{-1}}$. Thus, $M_\mathrm{s} = \delta v / c_\mathrm{s} \approx 0.2$–$0.3$ and $\beta_\mathrm{p} \approx 0.01$, consistent with the assumptions underlying RMHD derivations.

The timescales of the processes involved in heating models can also be examined for consistency with the assumptions of RMHD. In estimating a coronal value for $\tau_\mathrm{NL} = \lambda_\perp / \delta v$ we may take $\lambda_\perp$ to be in the range 3–30 Mm, where the

upper value is given by the size of super-granules (equivalent to the magnetic network separation distance). This yields $\tau_{\rm NL} \approx 10^2$–$10^3$ s. Now let us consider, somewhat arbitrarily, the parallel system size to be one solar radius.[9] Using a nominal Alfvén speed of $1000\,{\rm km\,s^{-1}}$, the associated $\tau_{\rm box}$ is 700 s. The longest period waves that fit within the system thus have $\epsilon_{\rm RMHD} \approx 1$. Hence, for RMHD to be sensibly applied, the driving fluctuations should have periods which are of order a few hundred seconds or longer. Fluctuations associated with $\tau_{\rm A} > \tau_{\rm box}$ do not "fit" within the system, and the nomenclature "low-frequency waves" gives way to "quasi-static fluctuations" as this inequality becomes progressively better satisfied. However, both low-frequency waves and quasi-static fluctuations can be dynamically included in the RMHD formulation.

Continual buffeting from the granular and super-granular motions associated with the convective dynamics of the interior is believed to be the source of the driving fluctuations and can excite Alfvén waves which propagate upwards along the (mean) field in magnetically open regions.[10] Observations indicate that the timescale associated with the coronal base driving fluctuations is $\tau_{\rm ph} \sim 10^3$–10^4 s [7,8,47], which is clearly much longer than the estimates for $\tau_{\rm NL}$.

For magnetically closed regions, like coronal loops, both ends of the "coronal box" are anchored in the photosphere, where the granular motions induce continual displacements of the field-lines. Provided again that those motions proceed on a slow scale, compared to the typical nonlinear time, the RMHD condition will be satisfied. (See the end of Sect. 3.2 for an important distinction between the wave box-crossing time and the modal Alfvénic time in connection with consideration of the RMHD condition.)

In summary, there are good theoretical and observational grounds for using RMHD to study the dynamics of coronal plasma. Parcels of the corona typically have: low $\beta_{\rm p}$, a strong approximately uniform magnetic field with low fluctuation levels ($\delta b/B_0 \ll 1$), and $\epsilon_{\rm RMHD} \lesssim 1$, all of which are in accord with the assumptions underlying RMHD. Thus, we expect that RMHD-based models will capture the essential physics of the low-frequency waves, quasi-static fluctuations, and turbulence that we wish to investigate. From a practical perspective, RMHD also affords considerable advantages in numerical simplicity and efficiency. Naturally, results from RMHD models will only provide an approximation to the true dynamics since, among other effects, it is highly likely that compressibility and kinetic processes will play important roles.

---

[9] This choice is not entirely without justification since observations indicate that heating in open field-line regions operates within about a solar radius above the photosphere.

[10] Similarly, buffeting and other processes (e.g., reconnection events) in the chromosphere and transition region may also excite Alfvén waves or related propagating fluctuations. For the heating model(s) considered below the actual source of the fluctuations is not particularly important. What is important is that such fluctuations are present at some (perhaps only notional) coronal base.

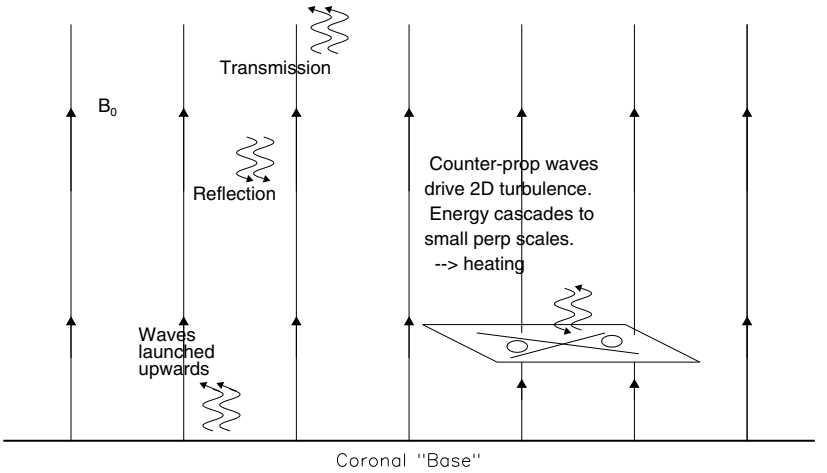


Fig. 3. Cartoon sketch of the heating model for magnetically open regions. Low-frequency waves are reflected off inhomogeneities in the large-scale density and magnetic field. The ensuing population of counter-propagating modes interacts to drive an RMHD (quasi-perpendicular) turbulent cascade, with the energy at the small-scale end of the cascade being converted into heat by kinetic processes

3 Coronal Heating Models

3.1 Coronal Holes

The class of models to be described was first presented by Matthaeus et al. [48] and has subsequently been investigated using phenomenology based models and nonlinear simulations [49,50,51,52,53]. As we shall see, an important distinguishing factor between the various models is the degree of realism and self-consistency associated with the reflection and driving terms.

Figure 3 is a schematic representation of the physics associated with the proposed heating scenario. *Low*-frequency Alfvénic fluctuations, generated somewhere below the coronal base, propagate up into the corona. Whilst progressing through the corona they undergo (*non*-WKB) reflection off the gradients in the background density and magnetic fields. This produces a population of counter-propagating fluctuations, which interact. The interaction is nonlinear and, because the modes are low-frequency, predominantly acts to drive *quasi-2D* fluctuations; these then self-interact to produce a cascade of energy towards small *perpendicular* lengthscales, with the formation of current sheets and the reconnection of transverse magnetic islands featuring prominently in the dynamics. Kinetic effects at these scales convert the turbulent energy into heat.

Models of this kind exhibit several features [51] which void some of the criticisms which have been leveled at direct wave heating models. For example, the usual criticism of (direct) wave heating models in magnetically open regions is that, because of the large propagation speeds ($\sim 10^3\,\mathrm{km\,s^{-1}}$), Alfvén waves propagate out of the region where the heating needs to occur (within $\approx 2\,R_{\mathrm{sun}}$

of the surface) before they can be significantly damped [54,55]. However, in the above indirect model, the waves lose energy primarily due to (nonlinear) transfer of excitation to the quasi-2D turbulent cascade, rather than (direct) viscous and resistive wave damping. For the low-frequency waves considered in the model, this is a relatively fast process and associated with the extraction of significant amounts of wave energy. Provided there is a region containing a population of low-frequency counter-propagating waves these will interact nonlinearly to drive a sustained quasi-2D cascade. Naturally, the *strength* of the cascade will depend on factors such as the abundance of upward and downward propagating modes (and hence the reflection rate), and the wave amplitudes, but the cascade's *existence* is not unduly constrained by the large Alfvén speed.

Note that direct wave damping models usually rely on damping of high-frequency (e.g., ion cyclotron) waves, whereas the class of models described above extracts energy from low-frequency waves. There are several advantages to using low-frequency waves as the input energy source. First, photospheric observations indicate that there is substantial energy in this range of the frequency power spectrum, whereas the spectrum at high frequencies ($\sim$ kHz), although unobserved, is expected to be of much lower amplitude [55,56,57]. Second, non-WKB reflection due to inhomogeneities in the solar atmosphere is most efficient for low-frequency fluctuations [58,59,60,61].

Another feature of the model depicted in Fig. 3 is that since the cascade is quasi-2D, both its dynamics and the associated heat output at small perpendicular scales are largely insensitive to the strength of B_0 [27,33,34,37,38].[11] This aspect may be of relevance in explaining why coronal heating appears to depend only weakly on solar cycle phase. Moreover, since the nascent solar wind is relatively sluggish over the height range where the counter-propagating fluctuations interact [12], one concludes that the heating occurs "in place." This follows since when the low-frequency waves supply energy to the cascade at a height r_1, say, the quasi-2D cascade transfers the energy to small scales at about this same height. Thus, a quasi-2D cascade, coupled with low wind speed, helps circumvent problems associated with the rapid removal of energy from the desired heating region by fast propagating waves.

We turn now to several studies of specific models within the above class of heating model. Note that at least two points make the RMHD equations a natural choice for use in investigating such models. First, the driving in the model is via low-frequency (and hence long wavelength) Alfvén waves propagating along a strong $\boldsymbol{B}_0$. Second, the primary couplings of such counter-propagating waves are to (Fourier) modes with wavevectors which are quasi-perpendicular to $\boldsymbol{B}_0$.

[11] Although the existence of a sufficiently strong B_0 may well be required in order to produce a substantial population of quasi-2D fluctuations in the first place. Moreover, even though the existence and strength of a quasi-2D cascade may be insensitive to the strength of B_0, the wave transmission out of the system and reflection within the system may retain a dependence upon B_0. Thus, the relative strength of turbulence versus wave effects at a fixed frequency of driving is expected to remain sensitive to the size of the Alfvén speed. It is only for the idealized case of "zero-frequency driving" that complete insensitivity to B_0 would be expected.

The initial investigation was via a one-point homogeneous turbulence closure for the energy in the upwards (Z_+^2) and downwards (Z_-^2) quasi-2D/RMHD fluctuations. Terms representing the forcing (F), energy-conserving reflection ($R^{\pm}$), and transmission (T) (see Fig. 3) were included using *ad hoc* rates [49],

$$\frac{dZ_-^2}{dt} = -\frac{Z_-^2 Z_+}{\lambda_\perp} + F - R^- Z_-^2 + R^+ Z_+^2 - T Z_-^2, \tag{6a}$$

$$\frac{dZ_+^2}{dt} = -\frac{Z_+^2 Z_-}{\lambda_\perp} + R^- Z_-^2 - R^+ Z_+^2. \tag{6b}$$

The model was closed by including an equation for the evolution of the characteristic transverse lengthscale of the quasi-2D fluctuations, $\lambda_\perp$, giving a (numerically) tractable system of three nonlinear ODEs [49].

In such a one-point closure the simulation "box" (Fig. 1) shrinks to a point and only volume-averaged information is available. Despite this simplicity, the model has proved to be remarkably robust, giving heating efficiencies in excellent agreement with models where the nonlinearities are treated with considerably more sophistication. In particular, in the limit that the reflection rate is much greater than the transmission rate, one can show analytically that the dissipation rate approaches $F/2$ asymptotically (from below). Defining the *heating efficiency* as the dissipation rate divided by the rate of energy injection, one finds that the maximum heating efficiency is 50% (see Fig. 5).

More sophisticated investigations of the heating scenario can be achieved using spectral method simulations of the nonlinear dynamics [51,53]. Such an approach is superior to the turbulence phenomenology in the sense that it does not assume that the turbulence is self-sustaining, allowing for the possibility that the heating mechanism may not be viable. When performing such simulations, the basic equations of RMHD, (1a,b), require augmenting with terms which account for the forcing, reflection, and transmission. In the simplest case [51] this is achieved using terms which are essentially the same as their analogs in the phenomenology, i.e., F is a body force, and $R^{\pm}$ and T are imposed rates.

Figure 4 shows time histories of several interesting quantities from a (periodic) RMHD simulation of this kind [51], with Reynolds numbers of 800, $F = 1$, $R = R^{\pm} = 0.5$, and $T = 0.3$. The initial conditions involve a seed level of turbulence, with the excited fluctuations having their perpendicular wavenumbers restricted to the band $2 \leq k_\perp \leq 6$ and an approximately flat spectrum in $k_\parallel$. It is evident that a statistically steady state is reached after a few tens of box crossing times. These states are characterised by mixed (normalised) cross helicity[12] $\sigma_c = \langle \boldsymbol{v} \cdot \boldsymbol{b} \rangle / (\langle \boldsymbol{v}^2 \rangle + \langle \boldsymbol{b}^2 \rangle)$, with around 10% of the energy contained in the *strictly* non-propagating (i.e., 2D) fluctuations. Note that the mean-square current is, visually, a much more intermittent quantity than the energy. Examination of the probability density functions of these quantities confirms this impression, with strongly non-gaussian tails characterising the current fluctuations [51]. Since in

[12] As well as being the average correlation between $\boldsymbol{v}$ and $\boldsymbol{b}$, cross helicity is also interpretable as the difference between the energy in upward and downward type fluctuations, e.g., [35,62,63]. Angle brackets denote spatial averaging.

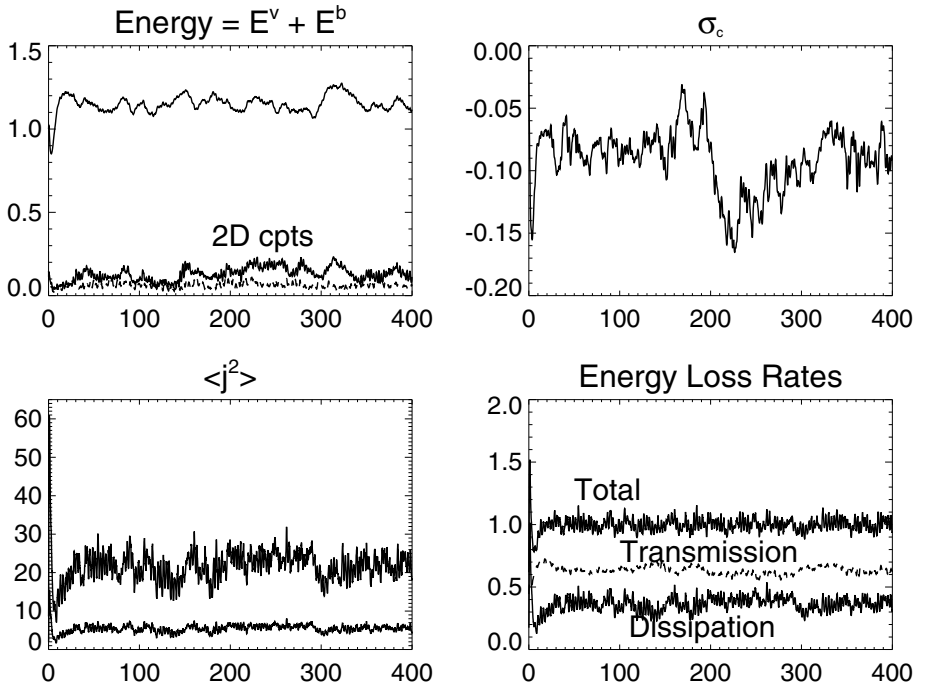


Fig. 4. Time histories from a $256^2 \times 4$ periodic RMHD simulation of the heating model discussed in the text. Shown are the energy, normalised cross helicity (which is bounded by ± 1), mean-square current, and energy loss rates. In the lefthand panels, the lower traces are the contributions to the totals due to the 2D (non-propagating) components. The unit of time is a box-crossing time for an Alfvén wave (after [51])

this incompressible model the heating is due to viscous and resistive dissipation, it is no surprise that the dissipation is also quite intermittent [64,65].

Examination of energy spectra and transverse cross sections in coordinate space (not shown; see [51]) emphasizes that the dynamics is turbulent with broadband spectra, transient current sheets, and reconnection events clearly evident. It is also shown that the long-time statistical properties of the system are insensitive to the total energy of the initial fluctuations, indicating that the heating process is robust.

A central question which all coronal heating models must address concerns their efficiency. Given an input source of energy, how much of it can be sustainably converted into heat within a few solar radii? In the case of the phenomenology and the simulations, it is straightforward to calculate the steady-state rates of dissipation and transmission. Figure 5 plots these two quantities (normalised to the energy injection rate) for a range of reflection rates. The data points are from the RMHD simulations [51] while the curves are calculated using the phenomenology solutions [49]. Remarkably, the results from the simple one-point phenomenology are in excellent accord with those from the full simulations. A key feature which the plot reveals is that even when the reflection rate is very

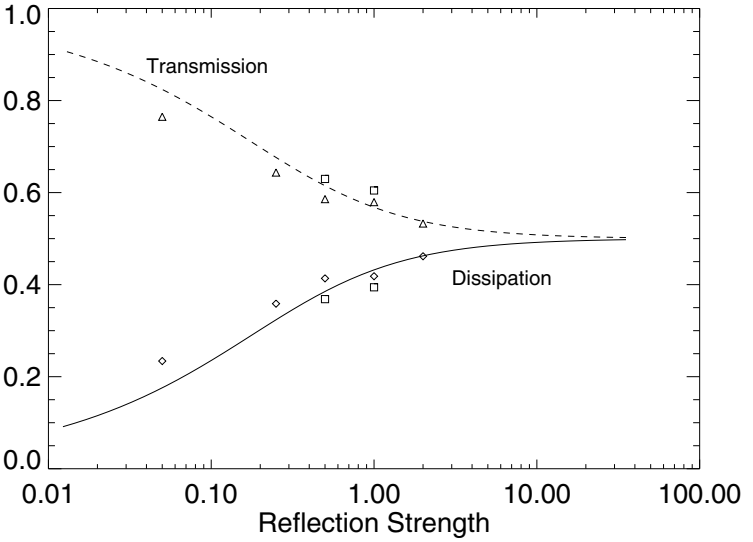


Fig. 5. Rates of dissipation and transmission for the one-point phenomenology (*curves*) and forced periodic RMHD simulations with resolutions of $256^2 \times 4$ (*squares*) and $64^2 \times 4$ (*other symbols*), respectively corresponding to Reynolds numbers of 800 and 200. Quantities are normalised to the energy injection rate, so that the dissipation rate is equivalent to the heating efficiency. Note the excellent agreement between the (simulation) data points and the solutions to the phenomenology, even when the reflection rate is small

weak (e.g., ~ 0.01), the models still yield heating efficiencies of $\sim 10\%$. Note that if the reflection is switched off in either model, the dissipation in the unforced modes decays to zero, so that sustained turbulent heating is not achieved [51,52]. This point is elaborated on shortly.

In summary, results obtained from the phenomenology and the periodic simulations indicate that heating efficiencies of 10%–50% are easily attainable—provided that *some* reflection of the upward propagating low-frequency fluctuations takes place. Given even rather small reflection rates, counter-propagating waves will drive—and sustain—a quasi-2D cascade for a wide range of initial conditions [51].

While these results are encouraging it is evident that they are only a first step. Some of the limitations concern the periodic nature of the boundary conditions and that forcing, reflection, and transmission are included in an *ad hoc* fashion. Models which include these effects in a more self-consistent way are clearly needed.

Recently, Dmitruk et al. [52] have extended the above models by including modeled (mean) coronal density and magnetic field profiles, with the reflection and transmission calculated from these in a consistent manner. Moreover, their heating model also better addresses both the transport of the fluctuations and the driving at the lower boundary. The basic physics is a more realistic imple-

mentation of the scenario sketched in Fig. 3, with upward propagating fluctuations injected at the base, initial conditions involving a seed level of broadband (RMHD) turbulence, and (height dependent) reflection coefficients proportional to the derivative of the large-scale Alfvén velocity. A key feature is that the boundary conditions are no longer periodic.

Using this model Dmitruk et al. investigated the following question: *Under what circumstances can one sustain incompressible (R)MHD turbulence in open geometries when it is driven using unidirectionally propagating Alfvén waves?*[13] In these circumstances it is convenient to use the Elsässer potentials as the basic variables, rather than the vorticity/vector potential formulation employed in (1a,b). Specifically, one introduces the potential for the upwards-type fluctuations, $f = \psi - a$, and that for the downwards-type ones, $g = \psi + a$. We use the word "type" to indicate that not all the fluctuations are propagating, since in general the averages of f and g over the parallel coordinate (denoted s rather than z in [52]) are non-zero. These averages correspond precisely to the existence of non-propagating "structures,"[14] which we refer to as 2D modes, e.g., $\boldsymbol{v}_{2\mathrm{D}}$ where $\partial_s \boldsymbol{v}_{2\mathrm{D}} = 0$.

To address the above question, Dmitruk et al. employed a simulation domain which is periodic in the transverse directions but bounded in the parallel direction by the planes $s = 0$ and $s = 1$. The equations solved are a modification of the basic RMHD equations which allow for a mean field which depends on s, and thus they include reflection in a self-consistent way [61,66,67]. This would introduce a term $\sim j\,\mathrm{d}B_0/\mathrm{d}z$ in (1a), for example. The numerical algorithm employs Fourier-Fourier-Chebyshev spectral collocation with second-order Runge–Kutta timestepping.

Two types of "open" boundary condition were employed on the end s-planes. They correspond to setting either $f = F_0(x, y, t)$ on $s = 0$ and $g = 0$ on $s = 1$, or to setting $\partial_s f = \hat{F}_0(x, y, t)$ on $s = 0$ and $\partial_s g = 0$ on $s = 1$ (i.e., Dirichlet versus Neumann boundary conditions). Since waves are permitted to propagate freely out of the box, there is no need to control the "other" potential on the top and bottom planes as it corresponds to the outward propagating characteristic there. In both cases, the flux of downward waves at the top boundary is enforced to be zero. Physically, one may think of the two types of boundary condition as allowing or disallowing the presence of non-propagating 2D fluctuations [52].

This leads to four fundamentally distinct situations to investigate, made up of the combinations of the two types of boundary condition paired with the presence or absence of reflection (which depends on whether B_0 is a function of s or uniform). Dmitruk et al. examined these situations by driving the $s = 0$ boundary with an F_0 or $\hat{F}_0$ which was monochromatic, meaning that it was sinusoidal in time, with fixed frequency ν, and that only a single transverse

[13] It is well known that in incompressible MHD there are *no* nonlinear interactions when *only* unidirectionally propagating fluctuations are present, i.e., either $\boldsymbol{v} = \boldsymbol{b}$ everywhere, or $\boldsymbol{v} = -\boldsymbol{b}$ everywhere.

[14] Note that these structures are not assumed to be static or even steady. In general they should be treated as 2D or $2\frac{1}{2}$D turbulence.

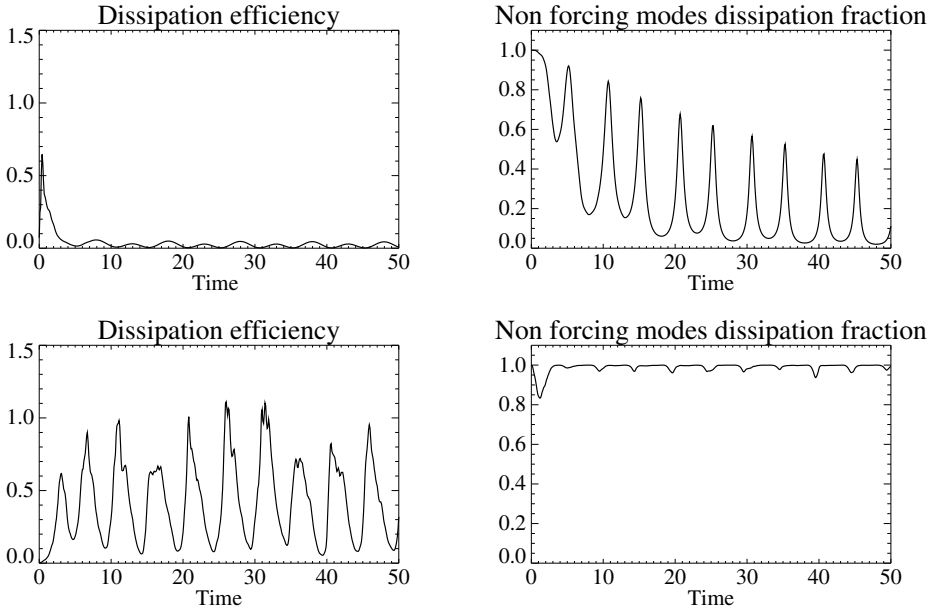


Fig. 6. Heating efficiencies for two runs with boundary conditions which support the presence in the system of non-propagating structures. **Left**: ratio of the dissipation rate to the (period-averaged) input wave energy flux. **Right**: fraction of the dissipation rate which is due to the undriven modes. **Top row**: no reflection. **Bottom row**: reflection. Note that the unit of time is $\tau_{\rm box}$

wavevector (Fourier mode) was so driven. In order to make connection with the coronal context, they employed low-frequency driving with $\nu = 0.1/\tau_{\rm box}$. Detailed discussion and explanation of the results is given in their paper [52]. Below we summarise their main result.

Figure 6 displays the heating efficiencies, as a function of time, for two runs from [52]. The plots on the top row are for the case where the boundary conditions support structures but there is no reflection (i.e., $\partial_s g = 0$, B_0 uniform). Since there is then no mechanism for generating downwards fluctuations, the cross helicity eventually becomes unidirectional. This can be interpreted as the turbulence being turned off due to *dynamic alignment*[15] [62,63]. Thus, although there is significant turbulent dissipation for tens of $\tau_{\rm box}$, this is not sustained over longer timescales. By contrast, the structures-plus-reflection situation (bottom row) does yield sustained turbulent heating, and, moreover, does so with considerable efficiency.

In the other two cases (not shown; see [52]), where there are no structures present, the turbulent dissipation is negligible after a few $\tau_{\rm box}$. In the absence of reflection this occurs because although the initial conditions contained both up-

[15] Maron and Goldreich [68] have recently rechristened dynamic alignment as "growth of imbalance".

wards and downwards propagating fluctuations, the downwards ones propagate out through the lower boundary in one box-crossing time, leaving only unidirectionally propagating modes. On the other hand, when reflection is active it facilitates the production of additional downward propagating modes. However, these only interact with the propagating upwards modes (since no 2D modes are present), and the interaction time is too short to allow a significant level of turbulence to be maintained.

Thus, it is only in the "reflection plus structures" cases that sustained turbulent heating ensues. Examination of the energy spectra for these cases verifies that the turbulence is broadband, with the suggestion of a (short) inertial range. As in the periodic RMHD simulations, intense transient current sheets are again evident [52].

It is well known that (incompressible) MHD turbulence, and hence turbulent heating, cannot be sustained solely by unidirectionally propagating waves, e.g., [62,69,70,71,72]. The conclusion from the above studies is that two conditions must be satisfied if turbulence in magnetically open regions is to be sustained. These are that (i) there exists some source of downward fluctuations, such as reflection, and (ii) low-frequency "non"-propagating fluctuations, sometimes referred to as structures or quasi-2D turbulence are present. In other words, in the open boundary situation, reflection is no longer the only crucial parameter. It is also necessary to have boundary conditions which support the sustained existence of non-propagating modes.

3.2 Coronal Loops

RMHD-based models for the heating of magnetically closed regions have been in use rather longer than those for open regions, e.g., [73,74,75,76,77,78,79,80,81,82], [83,84,85,86,87,88,89,8,47,65]. Mandrini et al. [90] have given a recent review of loop heating models in general, and Gomez et al. [91] have reviewed models which employ MHD turbulence. Here we note a few points regarding heating in loops and the RMHD approximation. We make no attempt to be comprehensive in our treatment, referring the interested reader to the above reviews for further details.

The common basic idea—often referred to as the Parker model [73]—is that the very high electrical conductivity of the photospheric and coronal plasma allows photospheric motions to drive magnetic stresses in the corona,[16] twisting the field-lines and generating transverse components to the velocity and magnetic field fluctuations. As mentioned in the introduction, the fundamental question is how the energy in these large stresses is transferred to smaller scales, where it can be efficiently dissipated. MHD turbulence is a naturally appealing mechanism for achieving such nonlinear energy transfer [91]. Comparisons of the scaling of the heating laws for loop heating models with coronal observations [90], suggest that models involving the gradual stressing of the magnetic field are generally

[16] The "frozen field" approximation.

in better agreement with the observations than those involving driving by high-frequency fluctuations (aka "wave heating" models), although the latter are not ruled out.

The point we wish to stress here is the similitude between heating models for magnetically closed (loops) and magnetically open (holes) regions when the motions driving the heating are slow. This similarity exists despite the differences between the boundary conditions and the order of magnitude of the involved length and timescales. In effect, a coronal loop driven (at both ends) by low-frequency photospheric motions can be thought of as an open coronal region, where partial or total reflection[17] is imposed at both the upper and lower boundaries. This "end reflection" provides the population of counter-propagating fluctuations needed for sustainment of incompressible MHD turbulence. Features like perpendicular small-scale structures (current sheets), transverse Kolmogorov-like spectra, and intermittent dissipation are strikingly similar in both closed [91] and "open plus reflection" [51] coronal simulations, confirming the analogy.

In closing this section we note that some coronal loop heating studies (e.g., [8,64]), state the RMHD (global) timescale condition as $\tau_{\rm NL} > \tau_{\rm A}$, which superficially appears to contradict the condition stated in Sect. 2.1. This apparent problem is actually a consequence of different definitions of $\tau_{\rm A}$. In the context of coronal loops, $\tau_{\rm A}$ is often defined to be the time it takes an Alfvén wave to traverse the loop lengthwise. Herein, however, this timescale is denoted by $\tau_{\rm box}$, since it is independent of the wave's $\boldsymbol{k}$. The statement that the crossing time for the loop ($\tau_{\rm box}$) is less than the timescale based on the (perpendicular) structure of the fluctuations ($\tau_{\rm NL}$) is closely connected with the nature of the motions perturbing the system. The slow motions of the photospheric plasma induce quasi-static low-frequency perturbations in the loop plasma. It is essentially a matter of definition that such low-frequency motions are associated with waves that "do not fit" within the box, since in an unbounded medium they would have wavevectors $\boldsymbol{k}_{\rm ph}$ which satisfy $k_{\rm ph}L_{\rm box} \ll 1$. Nonetheless, one finds (see [64,89]) that $\tau_{\rm NL}(\boldsymbol{k}_{\rm ph}) \sim \tau_{\rm box}$, so that the RMHD condition $\epsilon_{\rm RMHD}(\boldsymbol{k}_{\rm ph}) < 1$ is still marginally satisfied even for the shortest wavelength (quasi-static) Alfvén modes, for which $\tau_{\rm A}(\boldsymbol{k}) = \tau_{\rm box}$.

4 Conclusions

In this brief review we have attempted to show why reduced magnetohydrodynamics (RMHD) models can be made relevant to applications in coronal heating. In particular, RMHD can capture the physics of propagation, reflection, and cascade of incompressible MHD fluctuations in models of the lower solar atmosphere. The essential reasons behind the viability of RMHD models for the corona are that the Alfvén speed is large, the plasma beta (thermal pressure/magnetic pressure) is small, and the characteristic lengthscales across the large-scale magnetic field are quite likely much smaller than those along it. In

[17] The degree of reflection depends on the nature and strength of the photospheric velocity field.

addition, there is expected to be a plentiful supply [14] of energy to drive the quasi-2D turbulence, in the form of low-frequency Alfvén waves (generated in the photosphere and/or the chromosphere).

Two clear conclusions have emerged from RMHD studies of low-frequency-driven turbulent heating in regions of open magnetic field-lines, such as coronal holes. The first is that some mechanism for generating counter-propagating fluctuations from unidirectional ones is essential to these models. This requirement has been investigated using both *ad hoc* parameterizations of reflection, as well as models in which the reflection rates are computed consistently from the Alfvén speed profile that is implied by the model adopted for the background coronal magnetic field and density. In each case the conclusion is that turbulence, cascade, and efficient heating can be maintained given sufficient reflection to sustain a population of inward propagating-type modes. This population, surprisingly, may be adequate at a level of just a few percent of the total turbulence energy budget. The second feature that has emerged clearly in these models is the importance of the non-propagating structures (2D turbulence) to the sustainment of the turbulent cascade. The physical reason for this is plain enough: all modes other than the non-propagating ones drain out of the system in an Alfvén crossing time for the system. For a model to work without quasi-2D structures being present, turbulence must be fully replenished through nonlinear processes in a time shorter than this. This is possible, but difficult. Models in which non-propagating structure is permitted have the advantage that 2D fluctuations (which are highly turbulent) stay "in place", in spite of propagation effects that affect other modes.

An observation that can be made at this point is that the differences between heating models for open field-line corona and coronal loops may not be so great, when the analogy between the role of the reflection and the pairing of photospheric boundaries (footpoint motions) is taken into account. Each ensures that a supply of counter-propagating fluctuations will be maintained. While the lengthscales and other physical parameters in loops and coronal holes can be quite different, it still may be that a similar style of model is relevant in both cases.

One aspect of heating models we have not addressed here is whether a sufficient supply of fluctuations reaches the coronal base, a problem that impacts most if not all coronal heating models. Although we shall not discuss this in any detail, it is worth noting that only a small fraction of the fluctuation energy that is apparently available in the photosphere is needed [92], and that observations suggest that fluctuations with the required 20–30 $\mathrm{km\,s^{-1}}$ amplitudes may be present at the coronal base [11,93].

Another subject that we have not delved into is the issue of kinetic dissipation [34]. In the RMHD cascade models it is very clear that the dissipation occurs in randomly formed and highly dynamic sheets or filaments of electric current density. This gives rise to phenomena associated with highly turbulent reconnection [94], including turbulent dissipation of energy into heat. In a cascade picture the exact details of the mechanism of dissipation are not expected

to influence the rate of dissipation itself. However, it is a matter of considerable interest to identify the processes that would absorb the flux of cascaded energy. This is especially true in the corona, which is collisionless, so that the dissipative terms typically employed in turbulence calculations are probably inappropriate, particularly those based on scalar (and uniform) resistivity and viscosity. (They do, however, provide the sink at high wavenumber needed for the turbulence picture to function properly.)

A complete picture of coronal dissipation associated with quasi-2D cascade has not yet been developed. However it seems likely that two recently investigated elements will come into play. Studies of laminar spontaneous reconnection (e.g., [95]) indicate that Hall effects become important and that the reconnection zone takes on a characteristic structure determined by electrons and protons at the scale of the ion gyroradius or at the ion inertial scale. Secondly, the parallel electric fields associated with turbulent reconnection should rapidly produce strong electron beams. In their nonlinear phase these beams form electron "phase space holes" that propagate rapidly along the large-scale magnetic field. (These are observed in the geospace environment in conditions that are analogous to the corona in some interesting ways. See, e.g., [96]). Whether or not the scattering of protons by these high-frequency electron phenomena can give rise to the high perpendicular temperatures observed in the corona by spectrometers such as UVCS [9,10] has yet to be determined. If it can, then this mechanism would provide an alternative to the direct cyclotron damping mechanism often invoked for coronal heating [14], and one that is compatible with the perpendicular cascade. In contrast, direct cyclotron absorption appears to require a strong parallel cascade [97], which is difficult to justify based on MHD turbulence theory (see e.g., [34]). In any case, the wealth of possible kinetic activity within and near turbulent driven small-scale reconnection sites should provide ample opportunity for conversion of MHD energy into heat.

We foresee that there is a wide scope of future possibilities to extend, generalize, and improve the current generation of turbulence models for heating the corona. In the evolution towards more realism, prominent improvements would be better boundary conditions, better models for transport effects, and improved models of the background coronal fields themselves. Ultimately, of course, the large-scale coronal fields, and even the accelerated solar wind, should be included dynamically in such models. At that point the entire framework would require generalization, and a self-consistent model would emerge. While turbulence and cascade of MHD-scale fluctuations will remain a feature of a self-consistent coronal heating formalism, it remains to be seen how far in the evolution of these models the convenient and compact RMHD approximation will carry us.

Appendix A: Alternative Derivations of RMHD

The original derivation of the RMHD equations was presented by Strauss [24], motivated by geometries and plasma conditions typical of fusion devices (e.g., low β_{p}, large aspect ratio tokamaks, significant nonlinearity). The treatment

was perturbative with the small expansion parameter being the aspect ratio $\epsilon_{\rm Strauss} = \ell_\perp/\ell_\parallel$, where the ℓ's characterise typical lengthscales perpendicular and parallel to the mean magnetic field, $\boldsymbol{B}_0$. Ordering the fully 3D variables and equations in $\epsilon_{\rm Strauss}$ and then dropping (i) all toroidal effects, and (ii) all terms of the same order as the toroidal ones, yields the RMHD equations.[18]

Subsequently, Montgomery [25] rederived the RMHD equations from a different perspective, based on the assumption of incompressible fluctuations, which although weak in amplitude (relative to a strong $\boldsymbol{B}_0$), were still strongly nonlinear. In this case the small expansion parameter is expressible as $\epsilon_{\rm Mont} = \delta b/B_0$, where δb is the RMS magnetic fluctuation. This derivation also retained the viscous and resistive dissipation terms, and included discussion of the distinctive nature of spectral transfer in RMHD (see also [32]).

The conditions under which low Mach number compressible 3D MHD systems can be described using the equations of 2D or 3D incompressible MHD, $2\frac{1}{2}$D incompressible MHD, or RMHD were carefully considered by Zank and Matthaeus [26], from both physical and rigorously mathematical standpoints. As a small parameter they employed the sonic Mach number, $\epsilon_{\rm ZM} \equiv M_{\rm s} = u_0/c_{\rm s}$, where u_0 is the RMS velocity and $c_{\rm s}$ is the sound speed. As they showed, RMHD is the appropriate leading-order description when the system is characterised by (i) a small aspect ratio, with an applied uniform magnetic field along the "long" dimension, (ii) either $\beta_{\rm p} \ll 1$ or $\beta_{\rm p} \approx 1$, and (iii) elimination of all high-frequency (i.e., acoustic timescale or faster) modes.

Starting from the fully 3D compressible MHD equations with $\beta_{\rm p} \leq 1$, Bhattacharjee et al. [40] have derived a "four-field" system of equations for situations where the mean field $\boldsymbol{B}_0$ is allowed to vary slowly. These equations reduce to those of RMHD in the case that $\boldsymbol{B}_0$ is uniform.

The derivation sketched in Sect. 2 is based on the requirement that the (global) nonlinear timescale is not slower than the wave timescale, and it is easy to see that this involves the ratio of the Strauss and Montgomery small parameters:

$$\frac{\tau_{\rm NL}}{\tau_{\rm A}} = \frac{\ell_\perp}{\delta b}\Big/\frac{\ell_\parallel}{B_0} = \frac{\ell_\perp}{\ell_\parallel}\frac{B_0}{\delta b} = \frac{\epsilon_{\rm Strauss}}{\epsilon_{\rm Mont}} \equiv \epsilon_{\rm RMHD} \lesssim 1. \tag{7}$$

As in the Zank and Matthaeus derivation, this approach emphasizes that it is the *timescales* of the fluctuations which determine whether or not RMHD is an appropriate approximation to use. In this sense these derivations unify those of Strauss and Montgomery.

Recently, several other derivations related to the RMHD equations have been given. Gazol et al. [99,100] showed that when the compressible MHD equations are perturbatively expanded using the *Alfvénic* Mach number as the small parameter, one also obtains equations related to the RMHD ones. Specifically, the transverse dynamics is governed by the usual RMHD equations while the parallel dynamics (with the same timescale) is governed by a generalized derivative

[18] Strauss [24] notes that his motivation for deriving the RMHD equations was as a 3D generalisation of a (nonlinear) 2D approximation derived for tokamaks [98].

nonlinear Schrödinger equation. In related simulation studies, Del Zanna et al. [101,102] have shown that the propagation of Alfvén waves in a 3D compressible medium can lead to the formation of perpendicular structures, which may be connected with RMHD dynamics. Similarly, Laveder et al. [103] showed that in a system governed by the Hall-MHD equations, driving with finite amplitude dispersive Alfvén waves can produce transverse dynamics governed by the RMHD equations. In such cases, RMHD dynamics coexists with the propagation of *small*-scale Alfvén waves, thereby providing an example of the broader validity of the RMHD description.

Note that *none of these derivations of RMHD are linearizations* of the MHD equations: in RMHD the nonlinearities are always important (by construction). Indeed, there is no linearized version of RMHD, since the nonlinear terms cannot become small relative to the the linear (wave propagation) terms without violating the timescale assumptions used to derive the RMHD equations. RMHD fluctuations are intrinsically nonlinear.[19] Nonetheless, as discussed in Appendix B, the non-RMHD modes, which contain the linear wave solutions of the incompressible MHD equations, can still be simulated within the RMHD equations.

Finally, we note that the "critical balance" condition employed in a consideration of the energy spectra in strong incompressible MHD turbulence [29] is equivalent to the modal RMHD condition $\epsilon_{\mathrm{RMHD}}(\boldsymbol{k}) \lesssim 1$, in the sense that critical balance, defined as $\tau_{\mathrm{NL}} = \tau_{\mathrm{A}}$, describes the RMHD boundary (cf. Fig. 7a).

Appendix B: Self-consistency of RMHD Simulations

To obtain the k-space boundary curve which (roughly) separates the RMHD and non-RMHD fluctuations it is necessary to know the functional forms for the dependence of the energy spectrum on $k_\perp$ and k_z. For example, assuming a powerlaw perpendicular spectrum $\sim k_\perp^{-\alpha}$, and an approximately flat parallel spectrum, the modal RMHD condition, $\tau_{\mathrm{NL}}(\boldsymbol{k}) \lesssim \tau_{\mathrm{A}}(\boldsymbol{k})$, becomes $k_\perp \propto k_z^{2/(3-\alpha)}$. For the $\alpha = 5/3$ Kolmogorov case this becomes $k_\perp \propto k_z^{3/2}$. Such a curve is sketched in Fig. 7a. Note that this description of the limits of applicability of RMHD can be no more that an approximation in general, since turbulence may not be driven or steady, as would be required to attain a strict (inertial range) powerlaw; also, the dependence of the spectrum upon k_z is likely to be more complicated than that assumed above.

An interesting point then arises in connection with simulations employing the RMHD equations. Since these equations are valid for RMHD fluctuations, they should not, strictly speaking, be used to evolve non-RMHD modes. However, as Fourier-space simulation domains are often rectangular in shape, the set of excited fluctuations is likely to include both RMHD and non-RMHD modes, as indicated in Fig. 7a. One is then led to consider the self-consistency of such

[19] The exception to this is the subset of RMHD modes which are actually 2D (i.e., they have $k_\parallel = 0$). Hence, for these modes there is no parallel-propagation timescale with which to compare their nonlinear timescale.

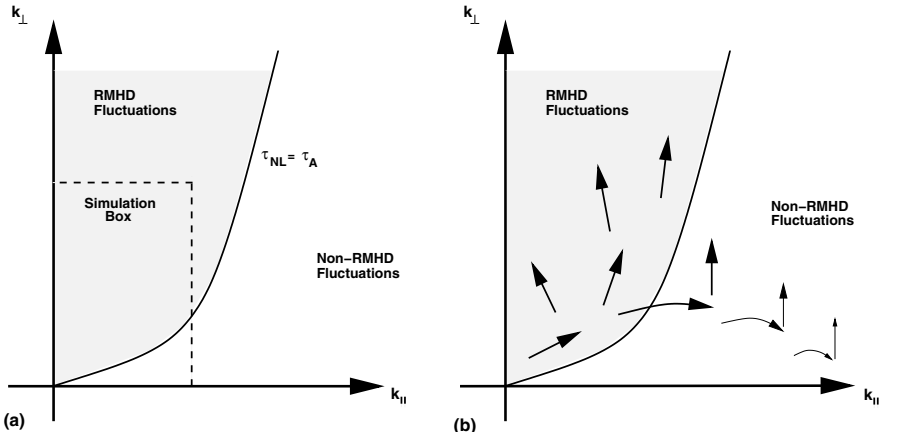


Fig. 7. (a) Cartoon sketch of the (spectral) boundary between the RMHD modes and the non-RMHD modes. The system is assumed to be in approximate steady-state, with an energy spectrum which is Kolmogorov-like in the perpendicular direction and approximately flat in the parallel direction. A typical simulation domain (*dotted*) is also indicated. **(b)** Indications of the direction and strength of spectral transfer for RMHD and non-RMHD modes. The width of the arrows suggests the strength of the transfer

simulations, and in particular whether or not non-RMHD modes behave in an (essentially) correct fashion when simulated via the RMHD equations.

In the case of periodic boundary conditions it appears that the situation is indeed self-consistent. Figure 8 shows results from forced[20] RMHD simulations at resolution $64^2 \times 32$ and Reynolds numbers of 200. Shown are the steady parallel energy spectra $E_z(k_z) = \int E(k_x, k_y, k_z)\,\mathrm{d}k_x \mathrm{d}k_y$, as a function of k_z, for five values of B_0 ranging from $1/2$ to 8. The simulations from which the plots are made all start with identical initial data and Reynolds numbers, with only the value of B_0 different in each run. Note that the values of B_0 are all large, since the physical mean field is given by B_0/ϵ (see Sect. 2.1), where ϵ is considered to have the same fixed value in each simulation.

The solid curve is for the "standard" RMHD situation of $B_0 = 1$. Clearly, larger values of B_0 lead to spectra which are significantly steeper and therefore associated with weaker parallel cascades. There is even some suggestion of a range where the falloff is exponential, especially for the larger values of B_0.[21]

Our interpretation of this result is that the essential physics for the non-RMHD modes is still retained in the RMHD equations. More specifically, counter-propagating Alfvénic wavepackets will only overlap for a time $t \approx \lambda_\parallel/(2V_A)$, where $\lambda_\parallel$ is a parallel lengthscale characteristic of the packets. If this time is

[20] The forcing consists of driving a unidirectional Alfvén wave with a $\boldsymbol{k} = (1, 1, 1)$.

[21] The increases at the highest values of k_z appear to be real properties of the numerical approximation and we suspect they are associated with energy "splashing" at the max k_z "walls", since there is no dissipation in this direction.

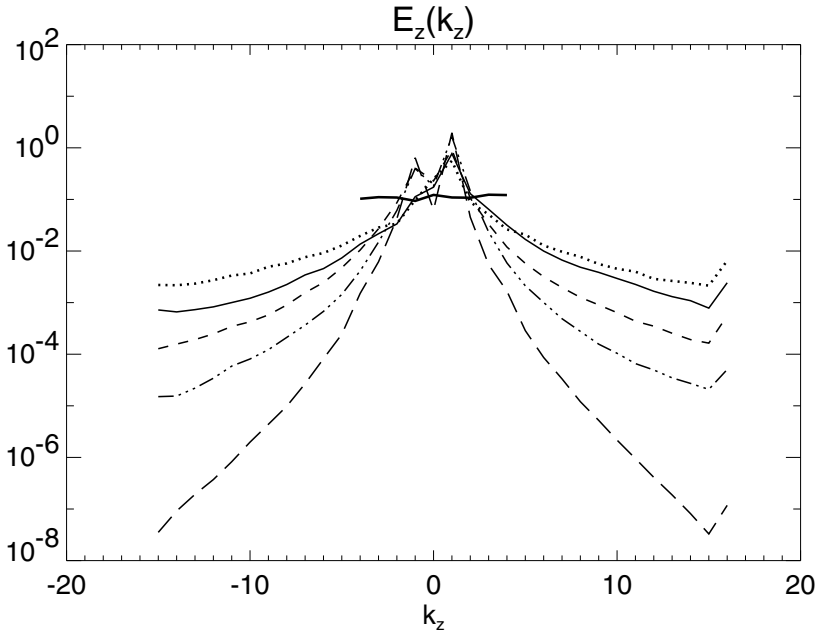


Fig. 8. Steady parallel energy spectra $E_z(k_z)$ for forced $64^2 \times 32$ RMHD simulations with various values of B_0. From top to bottom the values of B_0 are 1/2 (*dots*), 1, 2, 4, and 8 (*long dashes*). Note the steeper spectra for larger B_0, indicating a weakened parallel cascade. The heavy (roughly) horizontal line is the initial condition

shorter than the corresponding $\tau_{\rm NL}$ then the interaction generates destructive interference rather than a spectral cascade.

We close with a physical argument as to why the parallel spectrum could be exponential in k_z.[22] In the interaction of two Fourier modes, most of the excitation will be transferred to higher $\boldsymbol{k}_\perp$, at nearly fixed k_z (see Fig. 7b). However, some excitation will undergo parallel spectral transfer to a new k_z. The process then repeats at this (typically larger) k_z and such a survivalist process could easily produce an exponential spectrum in k_z. A fuller discussion of the topic of this appendix is being prepared for publication.

Acknowledgments

Special thanks to Peter Ulmschneider, for his kind permission to include Table 1, and to Marco Velli, for insightful conversations regarding RMHD and coronal physics. We also acknowledge many useful and instructive conversations with David Montgomery, who suggested the exponential parallel RMHD spectrum

[22] The first clear statement that the parallel spectrum should be exponential while the perpendicular spectrum should be $k^{-5/3}$ for MHD in a strong external magnetic field appears to be due to Montgomery [104, p. 139].

nearly twenty years ago. This work has been supported by grants from the UK PPARC (PPA/G/S/1999/00059), NASA, and the NSF.

References

1. W. Grotrian: Naturwissen **27**, 214 (1939)
2. B. Edlén: Z. Astrophys. **22**, 30 (1942)
3. U. Narain, P. Ulmschneider: Space Sci. Rev. **54**, 377 (1990)
4. U. Narain, P. Ulmschneider: Space Sci. Rev. **75**, 453 (1996)
5. P. Ulmschneider: 'Alfvén wave propagation in the solar corona and inner heliosphere'. In: *Cool Stars, Stellar Systems and the Sun*, ed. by R. Pallavicini, A.K. Dupree (PASP Conf. Series Vol 109, 1996)
6. P. Ulmschneider, E.R. Priest, R. Rosner: *Mechanisms of Chromospheric and Coronal Heating* (Springer, Berlin, 1991)
7. L. Golub, J.M. Pasachoff: *The Solar Corona* (CUP, Cambridge, 1997)
8. M. Velli: 'Coronal heating, nanoflares and MHD turbulence'. In: *Solar Wind Eight*, ed. by D. Winterhalter, J.T. Gosling, S.R. Habbal, W.S. Kurth, M. Neugebauer (AIP, New York, 1996), p. 28
9. J.L. Kohl et al.: Space Sci. Rev. **72**, 29 (1995)
10. J.L. Kohl et al.: Solar Phys. **175**, 613 (1997)
11. J. Chae, U. Schühle, P. Lemaire: Astrophys. J. **505**, 957 (1998)
12. R.R. Grall, W.A. Coles, M.T. Klinglesmith, A.R. Breen, P.J.S. Williams, J. Markkanen, R. Esser: Nature **379**, 429 (1996)
13. S.R. Cranmer, G.B. Field, J.L. Kohl: Astrophys. J. **518**, 937 (1999)
14. W.I. Axford, J.F. McKenzie: 'The solar wind'. In: *Cosmic Winds and the Heliosphere* (Arizona U. Press, 1997), p. 31
15. J.L. Culhane: Adv. Space Res. **19**, 1839 (1998)
16. F. Califano, C. Chiuderi: Physica Scripta **T75**, 197 (1998)
17. F. Califano, C. Chiuderi: Phys. Rev. E **60**, 4701 (1999)
18. A. Barnes: 'Hydromagnetic waves and turbulence in the solar wind'. In: *Solar System Plasma Physics, vol. I*, ed. by E.N. Parker, C.F. Kennel, L.J. Lanzerotti (North-Holland, Amsterdam, 1979), p. 251
19. A. Mangeney, R. Grappin, M. Velli: 'MHD turbulence in the solar wind'. In: *Advances in Solar System Magnetohydrodynamics*, ed. by E.R. Priest, A.W. Hood (CUP, Cambridge, 1991), p. 327
20. C.Y. Tu, E. Marsch: Space Sci. Rev. **73**, 1 (1995)
21. W.H. Matthaeus, J.W. Bieber, G.P. Zank: Rev. Geophys. Supp. **33**, 609 (1995)
22. J.C. Vial, B. Kaldeich-Schürmann, eds.: *Proc. of the 8th SoHO Workshop 'Plasma Dynamics and Diagnostics in the Solar Transition Region and Corona'*, vol. ESA SP-446 (ESA, Noordwijk, The Netherlands, 1999)
23. S. Masuda, T. Kosugi, H. Hara, S. Tsuneta, Y. Ogawara: Nature **371**, 495 (1994)
24. H.R. Strauss: Phys. Fluids **19**, 134 (1976)
25. D.C. Montgomery: Physica Scripta **T2/1**, 83 (1982)
26. G.P. Zank, W.H. Matthaeus: J. Plasma Phys. **48**, 85 (1992)
27. W.H. Matthaeus, S. Ghosh, S. Oughton, D.A. Roberts: J. Geophys. Res. **101**, 7619 (1996)
28. G.K. Batchelor: *The Theory of Homogeneous Turbulence* (CUP, Cambridge, 1970)
29. P. Goldreich, S. Sridhar: Astrophys. J. **438**, 763 (1995)
30. R. Kinney, J.C. McWilliams: Phys. Rev. E **57**, 7111 (1998)

31. S. Oughton, S. Ghosh, W.H. Matthaeus: Phys. Plasmas **5**, 4235 (1998)
32. D.C. Montgomery, L. Turner: Phys. Fluids **24**, 825 (1981)
33. J.V. Shebalin, W.H. Matthaeus, D. Montgomery: J. Plasma Phys. **29**, 525 (1983)
34. R.L. Leamon, W.H. Matthaeus, C.W. Smith, G.P. Zank, D.J. Mullan, S. Oughton: Astrophys. J. **537**, 1054 (2000)
35. R. Grappin: Phys. Fluids **29**, 2433 (1986)
36. S. Galtier, S.V. Nazarenko, A.C. Newell: Nonlin. Process. Geophys. **8**, 1 (2001)
37. V. Carbone, P. Veltri: Geophys. & Astrophys. Fluid Dynam. **52**, 153 (1990)
38. S. Oughton, E.R. Priest, W.H. Matthaeus: J. Fluid Mech. **280**, 95 (1994)
39. W.H. Matthaeus, S. Oughton, S. Ghosh, M. Hossain: Phys. Rev. Lett. **81**, 2056 (1998)
40. A. Bhattacharjee, C.S. Ng, S.R. Spangler: Astrophys. J. **494**, 409 (1998)
41. J. Cho, E.T. Vishniac: Astrophys. J. **539**, 273 (2000)
42. S. Galtier, S.V. Nazarenko, A.C. Newell, A. Pouquet: J. Plasma Phys. **63**, 447 (2000)
43. C.S. Ng, A. Bhattacharjee: Astrophys. J. **465**, 845 (1996)
44. J.G. Doyle, E. OShea, R. Erdelyi, K.P. Dere, D.G. Socker, F.P. Keenan: Solar Phys. **173**, 243 (1997)
45. Y.Q. Hu, R. Esser, S.R. Habbal: J. Geophys. Res. **102**, 14661 (1997)
46. H.P. Warren, J.T. Mariska, K. Wilhelm, P. Lemaire: Astrophys. J. **484**, L91 (1997)
47. G. Einaudi, M. Velli, H. Politano, A. Pouquet: Astrophys. J. **457**, L113 (1996)
48. W.H. Matthaeus, G.P. Zank, S. Oughton: 'Coronal heating by quasi-2D MHD turbulence driven by non-WKB wave reflection'. In: *Solar Wind Nine*, ed. by S. Habbal, R. Esser, J.V. Hollweg, P.A. Isenberg (AIP, Woodbury, NY, 1999), vol. 471, pp. 361–364
49. W.H. Matthaeus, G.P. Zank, S. Oughton, D.J. Mullan, P. Dmitruk: Astrophys. J. **523**, L93 (1999)
50. S. Oughton, W.H. Matthaeus, G.P. Zank, D.J. Mullan: 'Coronal heating via Alfvén waves and 2D MHD turbulence'. In: *Proc. of the 8th SoHO Workshop 'Plasma Dynamics and Diagnostics in the Solar Transition Region and Corona'*, ed. by J.C. Vial, B. Kaldeich-Schürmann (ESA, Noordwijk, The Netherlands, 1999), vol. ESA SP-446, pp. 525–530
51. S. Oughton, W.H. Matthaeus, P. Dmitruk, L.M. Milano, G.P. Zank, D.J. Mullan: Astrophys. J. **551**, 565 (2001)
52. P. Dmitruk, W.H. Matthaeus, L.J. Milano, S. Oughton: Phys. Plasmas **8**, 2377 (2001)
53. P. Dmitruk, L.J. Milano, W.H. Matthaeus: Astrophys. J. **548**, 482 (2001)
54. E.N. Parker: Astrophys. J. **372**, 719 (1991)
55. J. McKenzie, M. Banaszkiewicz, W.I. Axford: Astron. Astrophys. **303**, L45 (1995)
56. C.Y. Tu, E. Marsch: Solar Phys. **171**, 363 (1997)
57. S.R. Spangler, S. Mancuso: Astrophys. J. **530**, 491 (2000)
58. R.L. Moore, Z.E. Musielak, S.T. Suess, C.H. An: Astrophys. J. **378**, 347 (1991)
59. Z.E. Musielak, J.M. Fontenia, R.L. Moore: Phys. Fluids B **4**, 13 (1992)
60. Y. Zhou, W.H. Matthaeus: J. Geophys. Res. **95**, 14 863 (1990)
61. M. Velli: Astron. Astrophys. **270**, 304 (1993)
62. M. Dobrowolny, A. Mangeney, P. Veltri: Phys. Rev. Lett. **45**, 144 (1980)
63. W.H. Matthaeus, D. Montgomery: 'Dynamic alignment and selective decay in MHD '. In: *Statistical Physics and Chaos in Fusion Plasmas*, ed. by C.W.J. Horton, L.E. Reichl (Wiley, New York, 1984), p. 285

64. G. Einaudi, M. Velli: Phys. Plasmas **6**, 4146 (1999)
65. P. Dmitruk, D.O. Gómez: Astrophys. J. **484**, L83 (1997)
66. J.V. Hollweg: Solar Phys. **70**, 25 (1981)
67. Y. Zhou, W.H. Matthaeus: J. Geophys. Res. **95**, 10 291 (1990)
68. J. Maron, P. Goldreich: Astrophys. J. **554**, 1175 (2001)
69. R.H. Kraichnan: Phys. Fluids **8**, 1385 (1965)
70. R. Grappin, U. Frisch, J. Léorat, A. Pouquet: Astron. Astrophys. **105**, 6 (1982)
71. A. Pouquet, U. Frisch, M. Meneguzzi: Phys. Rev. A **33**, 4266 (1986)
72. S. Ghosh, W.H. Matthaeus, D. Montgomery: Phys. Fluids **31**, 2171 (1988)
73. E.N. Parker: Astrophys. J. **174**, 499 (1972)
74. E.N. Parker: Astrophys. J. **264**, 642 (1983)
75. E.N. Parker: Astrophys. J. **330**, 474 (1988)
76. A.A. van Ballegooijen: Astrophys. J. **311**, 1001 (1986)
77. N.F. Otani, H.R. Strauss: Astrophys. J. **325**, 468 (1988)
78. H.R. Strauss, N.F. Otani: Astrophys. J. **326**, 418 (1988)
79. D.O. Gómez, C. Ferro Fontán: Solar Phys. **116**, 33 (1988)
80. D.O. Gómez, C. Ferro Fontán: Astrophys. J. **394**, 662 (1992)
81. Z. Mikić, D.C. Barnes, D.D. Schnack: Astrophys. J. **328**, 830 (1988)
82. Z. Mikić, D.D. Schnack, G. Van Hoven: Astrophys. J. **338**, 1148 (1989)
83. J. Heyvaerts, E.R. Priest: Astrophys. J. **390**, 297 (1992)
84. D.W. Longcope, R.N. Sudan: Astrophys. J. **384**, 305 (1992)
85. D.W. Longcope, R.N. Sudan: Astrophys. J. **437**, 491 (1994)
86. D.W. Longcope, H.R. Strauss: Astrophys. J. **426**, 742 (1994)
87. D.L. Hendrix, G. van Hoven: Astrophys. J. **467**, 887 (1996)
88. P. Dmitruk, D.O. Gómez, E.E. DeLuca: Astrophys. J. **505**, 974 (1998)
89. P. Dmitruk, D.O. Gómez: Astrophys. J. **527**, L63 (1999)
90. C.H. Mandrini, P. Demoulin, J.A. Klimchuk: Astrophys. J. **530**, 999 (2000)
91. D.O. Gómez, P. Dmitruk, L.J. Milano: Solar Phys. **195**, 299 (2000)
92. T.E. Berger, A.M. Title: Astrophys. J. **463**, 365 (1996)
93. D.M. Hassler, G.J. Rottman, E.C. Shoub, T.E. Holzer: Astrophys. J. **348**, L77 (1990)
94. W.H. Matthaeus, S.L. Lamkin: Phys. Fluids **29**, 2513 (1986)
95. M.A. Shay, J.F. Drake, R.E. Denton, D. Biskamp: J. Geophys. Res. **103**, 9165 (1998)
96. R.E. Ergun, C.W. Carlson, J.P. McFadden, F.S. Mozer, L. Muschietti, R. I., R.J. Strangeway: Phys. Rev. Lett. **81**, 826 (1998)
97. S.R. Cranmer: Astrophys. J. **532**, 1197 (2000)
98. M.N. Rosenbluth, D.A. Monticello, H.R. Strauss, R.B. White: Phys. Fluids **19**, 1987 (1976)
99. A. Gazol, T. Passot, P.L. Sulem: Phys. Plasmas **6**, 3114 (1999)
100. A. Gazol, T. Passot, P.L. Sulem: Rev. Mex. A. A. **9**, 80 (2000)
101. L. Del Zanna: Geophys. Rev. Lett. **28**, 2585 (2001)
102. L. Del Zanna, M. Velli, P. Londrillo: Astron. Astrophys. **367**, 705 (2001)
103. D. Laveder, T. Passot, P.L. Sulem: Phys. Plasmas **9**, 305 (2002)
104. D.C. Montgomery: *Lecture Notes on Turbulence* (World Scientific, Singapore, 1989), p. 75. Lecture Notes from the NCAR-GTP Summer School, June 1987

MHD Turbulence: Scaling Laws and Astrophysical Implications

Jungyeon Cho[1], Alex Lazarian[1], and Ethan T. Vishniac[2]

[1] Univ. of Wisconsin, Madison WI53706, USA
[2] Johns Hopkins Univ., Baltimore MD21218, USA

Abstract. Turbulence is the most common state of astrophysical flows. In typical astrophysical fluids, turbulence is accompanied by strong magnetic fields, which has a large impact on the dynamics of the turbulent cascade. Recently, there has been a significant breakthrough on the theory of magnetohydrodynamic (MHD) turbulence. For the first time we have a scaling model that is supported by both observations and numerical simulations. We review recent progress in studies of both incompressible and compressible turbulence. We compare Iroshnikov-Kraichnan and Goldreich-Sridhar models, and discuss scalings of Alfvén, slow, and fast waves. We also discuss the completely new regime of MHD turbulence that happens below the scale at which hydrodynamic turbulent motions are damped by viscosity. In the case of the partially ionized diffuse interstellar gas the viscosity is due to neutrals and truncates the turbulent cascade at ∼parsec scales. We show that below this scale magnetic fluctuations with a shallow spectrum persist and discuss the possibility of a resumption of the MHD cascade after ions and neutrals decouple. We discuss the implications of this new insight into MHD turbulence for cosmic ray transport, grain dynamics, etc., and how to test theoretical predictions against observations.

1 Introduction

Most astrophysical systems, e.g. accretion disks, stellar winds, the interstellar medium (ISM) and intercluster medium are turbulent with an embedded magnetic field that influences almost all of their properties. This turbulence which spans from km to many kpc (see discussion in [2,150,86]) holds the key to many astrophysical processes (e.g., transport of mass and angular momentum, star formation, fragmentation of molecular clouds, heat and cosmic ray transport, magnetic reconnection). Statistics of turbulence is also essential for the cosmic microwave background (CMB) radiation foreground studies [88].

All turbulent systems have one thing in common: they have a large "Reynolds number" ($Re \equiv LV/\nu$; L= the characteristic scale or driving scale of the system, V=the velocity difference over this scale, and ν=viscosity), the ratio of the viscous drag time on the largest scales (L^2/ν) to the eddy turnover time of a parcel of gas (L/V). A similar parameter, the "magnetic Reynolds number", Rm ($\equiv LV/\eta$; η=magnetic diffusion), is the ratio of the magnetic field decay time (L^2/η) to the eddy turnover time (L/V). The properties of the flows on all scales depend on Re and Rm. Flows with $Re < 100$ are laminar; chaotic structures develop gradually as Re increases, and those with $Re \sim 10^3$ are appreciably less

E. Falgarone and T. Passot (Eds.): LNP 614, pp. 56–98, 2003.
© Springer-Verlag Berlin Heidelberg 2003

chaotic than those with $Re \sim 10^7$. Observed features such as star forming clouds and accretion disks are very chaotic with $Re > 10^8$ and $Rm > 10^{16}$.

Let us start by considering incompressible hydrodynamic turbulence, which can be described by the Kolmogorov theory [68]. Suppose that we excite fluid motions at a scale L. We call this scale the *energy injection scale* or the *largest energy containing eddy scale*. For instance, an obstacle in a flow excites motions on scales of the order of its size. Then the energy injected at the scale L cascades to progressively smaller and smaller scales at the eddy turnover rate, i.e. $\tau_l^{-1} \approx v_l/l$, with negligible energy losses along the cascade [1]. Ultimately, the energy reaches the molecular dissipation scale l_d, i.e. the scale where the local $Re \sim 1$, and is dissipated there. The scales between L and l_d are called the *inertial range* and it typically covers many decades. The motions over the inertial range are *self-similar* and this provides tremendous advantages for theoretical description.

The beauty of the Kolmogorov theory is that it does provide a simple scaling for hydrodynamic motions. If the velocity at a scale l from the inertial range is v_l, the Kolmogorov theory states that the kinetic energy ($\rho v_l^2 \sim v_l^2$ as the density is constant) is transferred to next scale within one eddy turnover time (l/v_l). Thus within the Kolmogorov theory the energy transfer rate ($v_l^2/(l/v_l)$) is scale-independent,

$$\frac{v_l^2}{t_{cas}} \sim \frac{v_l^2}{(l/v_l)} = \text{ constant}, \tag{1}$$

and we get the famous Kolmogorov scaling

$$v_l \propto l^{1/3}. \tag{2}$$

The one-dimensional[2] energy spectrum $E(k)$ is the amount of energy between the wavenumber k and $k+dk$ divided by dk. When $E(k)$ is a power law, $kE(k)$ is the energy *near* the wavenumber $k \sim 1/l$. Since $v_l^2 \approx kE(k)$, Kolmogorov scaling implies

$$E(k) \propto k^{-5/3}. \tag{3}$$

Kolmogorov scalings were the first major advance in the theory of incompressible turbulence. They have led to numerous applications in different branches of science (see [112]). However, astrophysical fluids are magnetized and the dynamically important magnetic field should interfere with eddy motions.

Paradoxically, astrophysical measurements are consistent with Kolmogorov spectra (see LPE02 [86]). For instance, interstellar scintillation observations indicate an electron density spectrum very close to $-5/3$ for $10^8 cm$ - $10^{15} cm$ (see [2]). At larger scales LPE02 summarizes the evidence of $-5/3$ velocity power spectrum over pc-scales in HI. Solar-wind observations provide *in-situ* measurements of the power spectrum of magnetic fluctuations and Leamon et al. [93] also obtained a slope of $\approx -5/3$. Is this a coincidence? What properties is the magnetized compressible ISM expected to have? We will deal with these questions, and some related issues, below.

[1] This is easy to see as the motions at the scales of large eddies have $Re \gg 1$.

[2] Dealing with observational data, e.g. in LPE02 [86], we deal with three dimensional energy spectrum $P(k)$, which, for isotropic turbulence, is given by $E(k) = 4\pi k^2 P(k)$.

Our approach here is complementary to that in Vazquez-Semadeni (this volume) and Mac Low (this volume). These reviews discuss attempts to simulate the turbulent ISM in all its complexity by including many physical processes (e.g. heating, cooling, self-gravity) simultaneously. This provides a possibility of comparing observations and simulations (see review by Ostriker, this volume). The disadvantage is that such simulations cannot distinguish between the consequences of different processes. Note, that in studies of turbulence the adaptive mesh does not help as the fine structures emerge through the entire computational volume.

Here we discuss a focused approach which aims at obtaining a clear understanding on the fundamental level, and considering physically relevant complications later. The creative synthesis of both approaches is the way, we think, that studies of astrophysical turbulence should proceed[3]. Certainly an understanding of MHD turbulence in the most ideal terms is a necessary precursor to understanding the complications posed by more realistic physics and numerical effects. For review of general properties of MHD, see a recent book by Biskamp [5].

In what follows, we first consider observational data that motivate our study (Sect. 2), then discuss theoretical approaches to incompressible MHD turbulence (Sect. 3). In Sect. 4 we discuss testing and extending of the Goldreich-Sridhar theory of turbulence, then in Sect. 5 we deal with viscous damping of incompressible turbulence and describe a new regime of MHD turbulence that is present below the viscous cut-off scale. We move to the effects of compressibility in Sect. 6 and discuss implications of our new understanding of MHD turbulence for the problems of dust motion, cosmic ray dynamics, support of molecular clouds, heating of ISM etc in Sect. 7. We propose observational testing of our results in Sect. 8 and present the summary in Sect. 9.

2 Observational Data

Kolmogorov turbulence is the simplest possible model of turbulence. Since it is incompressible and not magnetized, it is completely specified by its velocity spectrum. If a passive scalar field, like "dye particles" or temperature inhomogeneities, is subjected to Kolmogorov turbulence, the resulting spectrum of the passive scalar density is also Kolmogorov (see [95,171]). In compressible and magnetized turbulence this is no longer true, and a complete characterization of the turbulence requires not only a study of the velocity statistics but also the statistics of density and magnetic fluctuations.

[3] Potentially our approach leads to an understanding of the relationship between motions at a given time at small scales (subgrid scales) and the state of the flow at a previous time at some larger, resolved, scale. This could lead to a parametrization of the subgrid scales and to large eddy simulations of MHD.

Direct studies of turbulence[4] have been done mostly for interstellar medium and for the Solar wind. While for the Solar wind *in-situ* measurements are possible, studies of interstellar turbulence require inverse techniques to interpret the observational data.

Attempts to study interstellar turbulence with statistical tools date as far back as the 1950s [59,64,117,172] and various directions of research achieved various degree of success (see reviews by [65,31,2,78,79,86]).

2.1 Solar Wind

Solar wind (see review [48]) studies allow pointwise statistics to be measured directly using spacecrafts. These studies are the closest counterpart of laboratory measurements.

The solar wind flows nearly radially away from the Sun, at up to about 700 km/s. This is much faster than both spacecraft motions and the Alfvén speed. Therefore, the turbulence is "frozen" and the fluctuations at frequency f are directly related to fluctuations at the scale k in the direction of the wind, as $k = 2\pi f/v$, where v is the solar wind velocity [57].

Usually two types of solar wind are distinguished, one being the fast wind which originates in coronal holes, and the slower bursty wind. Both of them show, however, $f^{-5/3}$ scaling on small scales. The turbulence is strongly anisotropic (see [66]) with the ratio of power in motions perpendicular to the magnetic field to those parallel to the magnetic field being around 30. The intermittency of the solar wind turbulence is very similar to the intermittency observed in hydrodynamic flows [58].

2.2 Electron Density Statistics

Studies of turbulence statistics of ionized media (see [157]) have provided information on the statistics of plasma density at scales 10^8-10^{15} cm. This was based on a clear understanding of processes of scintillations and scattering achieved by theorists[5] (see [121,49]). A peculiar feature of the measured spectrum (see [2]) is the absence of slope change at the scale at which the viscosity by neutrals becomes important.

Scintillation measurements are the most reliable data in the "big power law" plot in Armstrong et al. [2]. However there are intrinsic limitations to the scintillations technique due to the limited number of sampling directions, its relevance only to ionized gas at extremely small scales, and the impossibility of getting velocity (the most important!) statistics directly. Therefore with the data one faces the problem of distinguishing actual turbulence from static density structures.

[4] Indirect studies include the line-velocity relationships [75] where the integrated velocity profiles are interpreted as the consequence of turbulence. Such studies do not provide the statistics of turbulence and their interpretation is very model dependent.

[5] In fact, the theory of scintillations was developed first for the atmospheric applications.

Moreover, the scintillation data does not provide the index of turbulence directly, but only shows that the data are consistent with Kolmogorov turbulence. Whether the (3D) index can be -4 instead of -11/3 is still a subject of intense debate [56,121]. In physical terms the former corresponds to the superposition of random shocks rather than eddies.

Additional information on the electron density is contained in the Faraday rotation measures of extragalactic radio sources (see [154,155]). However, there is so far no reliable way to disentangle contributions of the magnetic field and the density to the signal. We feel that those measurements may give us the magnetic field statistics when we know the statistics of electron density better.

2.3 Velocity and Density Statistics from Spectral Lines

Spectral line data cubes are unique sources of information on interstellar turbulence. Doppler shifts due to supersonic motions contain information on the turbulent velocity field which is otherwise difficult to obtain. Moreover, the statistical samples are extremely rich and not limited to discrete directions. In addition, line emission allows us to study turbulence at large scales, comparable to the scales of star formation and energy injection.

However, the problem of separating velocity and density fluctuations within HI data cubes is far from trivial [77,79,84,86]. The analytical description of the emissivity statistics of channel maps (velocity slices) in Lazarian & Pogosyan [84] (see also [79,86] for reviews) shows that the relative contribution of the density and velocity fluctuations depends on the thickness of the velocity slice. In particular, the power-law asymptote of the emissivity fluctuations changes when the dispersion of the velocity at the scale under study becomes of the order of the velocity slice thickness (the integrated width of the channel map). These results are the foundation of the Velocity-Channel Analysis (VCA) technique which provides velocity and density statistics using spectral line data cubes. The VCA has been successfully tested using data cubes obtained via compressible magnetohydrodynamic simulations and has been applied to Galactic and Small Magellanic Cloud atomic hydrogen (HI) data [87,84,159,28]. Furthermore, the inclusion of absorption effects [85] has increased the power of this technique. Finally, the VCA can be applied to different species (CO, H_α etc.) which should further increase its utility in the future.

Within the present discussion a number of results obtained with the VCA are important. First of all, the Small Magellanic Cloud (SMC) HI data exhibit a Kolmogorov-type spectrum for velocity and HI density from the smallest resolvable scale of 40 pc to the scale of the SMC itself, i.e. 4 kpc. Similar conclusions can be inferred from the Galactic data [52] for scales of dozens of parsecs, although the analysis has not been done systematically. Deshpande et al. [28] studied absorption of HI on small scales toward Cas A and Cygnus A. Within the VCA their results can be interpreted as implying that on scales less than 1 pc the HI velocity is suppressed by ambipolar drag and the spectrum of density fluctuations is shallow $P(k) \sim k^{-2.8}$. Such a spectrum [27] can account for the small-scale structure of HI observed in absorption.

2.4 Magnetic Field Statistics

Magnetic field statistics are the most poorly constrained aspect of ISM turbulence. The polarization of starlight and of the Far-Infrared Radiation (FIR) from aligned dust grains is affected by the ambient magnetic fields. Assuming that dust grains are always aligned with their longer axes perpendicular to magnetic field (see the review [80]), one gets the 2D distribution of the magnetic field directions in the sky. Note that the alignment is a highly non-linear process in terms of the magnetic field and therefore the magnetic field strength is not available[6].

The statistics of starlight polarization (see [40]) is rather rich for the Galactic plane and it allows to establish the spectrum[7] $E(K) \sim K^{-1.5}$, where K is a two dimensional wave vector describing the fluctuations over sky patch.[8]

For uniformly sampled turbulence it follows from Lazarian & Shutenkov [83] that $E(K) \sim K^{\alpha}$ for $K < K_0$ and K^{-1} for $K > K_0$, where K_0^{-1} is the critical angular size of fluctuations which is proportional to the ratio of the injection energy scale to the size of the turbulent system along the line of sight. For Kolmogorov turbulence $\alpha = -11/3$.

However, the real observations do not uniformly sample turbulence. Many more close stars are present compared to the distant ones. Thus the intermediate slopes are expected. Indeed, Cho & Lazarian [22] showed through direct simulations that the slope obtained in [40] is compatible with the underlying Kolmogorov turbulence. At the moment FIR polarimetry does not provide maps that are really suitable to study turbulence statistics. This should change soon when polarimetry becomes possible using the airborne SOFIA observatory. A better understanding of grain alignment (see [80]) is required to interpret the molecular cloud magnetic data where some of the dust is known not to be aligned (see [82] and references therein).

Another way to get magnetic field statistics is to use synchrotron emission. Both polarization and intensity data can be used. The angular correlation of polarization data [3] shows the power-law spectrum $K^{-1.8}$ and we believe that the interpretation of it is similar to that of starlight polarization. Indeed, Faraday depolarization limits the depth of the sampled region. The intensity fluctuations were studied in [83] with rather poor initial data and the results were inconclusive. Cho & Lazarian [22] interpreted the fluctuations of synchrotron emissivity [43,44] in terms of turbulence with Kolmogorov spectrum.

---

[6] The exception to this may be the alignment of small grains which can be revealed by microwave and UV polarimetry [80].

[7] Earlier papers dealt with much poorer samples (see [65]) and they did not reveal power-law spectra.

[8] This spectrum is obtained by [40] in terms of the expansion over the spherical harmonic basis Y_{lm}. For sufficiently small areas of the sky analyzed the multipole analysis results coincide with the Fourier analysis.

3 Theoretical Approaches to MHD Turbulence

Here we consider mainly Kolmogorov-type theories. Other theories not discussed in this section include the eddy-damped quasinormal Markovian (EDQNM) approximation [139], the renormalization group technique [39,174], and the direct interaction approximation [70].

3.1 Iroshnikov-Kraichnan Theory

Attempts to describe magnetic turbulence statistics were made by Iroshnikov [61] and Kraichnan [71]. Their model of turbulence (IK theory) is isotropic in spite of the presence of the magnetic field.

We can understand the IK theory as follows.[9] For simplicity, let us suppose that a uniform external magnetic field ($\mathbf{B}_0$) is present. In the incompressible limit, any magnetic perturbation propagates *along* the magnetic field line. To the first order, the speed of propagation is constant and equal to the Alfvén speed $V_A = B_0/\sqrt{4\pi\rho}$, where ρ is the density. Since wave packets are moving along the magnetic field line, there are two possible directions for propagation. If all the wave packets are moving in one direction, then they are stable to nonlinear order[132]. Therefore, in order to initiate turbulence, there must be opposite-traveling wave packets and the energy cascade occurs only when they collide. The IK theory starts from this observation.

The IK theory assumes that, when two opposite-traveling wave packets of size l collide, they lose the following amount of energy to smaller scales:

$$\Delta E \sim (dv^2/dt)\Delta t \sim \mathbf{v}_l \cdot \dot{\mathbf{v}}_l \Delta t \sim v_l(v_l^2/l)\Delta t \sim (v_l^3/l)(l/V_A), \tag{4}$$

where the IK theory assumes that only collisions between similar size packets are important and $\Delta t \sim l/V_A$. The latter means that the duration of the collision is the size of the wave packet divided by the speed of the wave packets. These assumptions look reasonable at first. But, it is important to note that they fail when eddies are anisotropic. That is, if eddies are elongated along the magnetic field line, then Δt is not l/V_A, but $l_\parallel/V_A$, where $l_\parallel$ is the parallel size of the wave packet (or 'eddy').

Equation (4) tells us that the energy change per collision is $v_l^2(v_l/V_A)$, which is only a tiny fraction of v_l^2 when $V_A \gg v_l$. Therefore, in order for the eddy to transfer all the energy to small eddies, the eddy must go through many collisions. When such collisions are incoherent, we require a total $(v^2/\Delta E)^2$ collisions to complete the cascade. This means that the energy cascade time t_{cas} is

$$t_{cas} \sim \left(\frac{v^2}{\Delta E}\right)^2 \Delta t \sim \frac{l}{v_l}\frac{V_A}{v_l}, \tag{5}$$

which means that this new cascade time is (V_A/v_l) times longer than the eddy turnover time (l/v_l). As in the Kolmogorov theory, the IK theory assumes the

[9] We follow arguments in [46].

constancy of energy cascade (1): $(v_l^4)/(lV_A)$ = constant, which, in turn, yields $v_l \propto l^{1/4}$, or,

$$\textit{Iroshnikov-Kraichnan:} \qquad E(k) \propto k^{-3/2}. \tag{6}$$

3.2 Anisotropy

A uniform component to the magnetic field defines a special direction, which will be reflected in the dynamics of turbulent fluctuations. One obvious effect is that it is easy to mix field lines in directions perpendicular to the local mean magnetic field and much more difficult to bend them. The IK theory assumes isotropy of the energy cascade in Fourier space, an assumption which has attracted severe criticism [115,153,114,158,104]. Mathematically, anisotropy manifests itself in the resonant conditions for 3-wave interactions:

$$\mathbf{k}_1 + \mathbf{k}_2 = \mathbf{k}_3, \tag{7}$$

$$\omega_1 + \omega_2 = \omega_3, \tag{8}$$

where $\mathbf{k}$'s are wavevectors and ω's are wave frequencies. The first condition is a statement of wave momentum conservation and the second is a statement of energy conservation. Alfvén waves satisfy the dispersion relation: $\omega = V_A|k_\parallel|$, where $k_\parallel$ is the component of wavevector parallel to the background magnetic field. Since only opposite-traveling wave packets interact, $\mathbf{k}_1$ and $\mathbf{k}_2$ must have opposite signs. Then from equations (7) and (8), either $k_{\parallel,1}$ or $k_{\parallel,2}$ must be equal to 0 and $k_{\parallel,3}$ must be equal to the nonzero initial parallel wavenumber. That is, zero frequency modes are essential for energy transfer [153]. Therefore, in the wavevector space, 3-wave interactions produce an energy cascade which is strictly perpendicular to the mean magnetic field. However, in real turbulence, equation (8) does not need to be satisfied exactly, but only to within an error of order $\delta\omega \sim 1/t_{cas}$ [45]. This implies that the energy cascade is not strictly perpendicular to $\mathbf{B}_0$, although clearly very anisotropic.

It is noteworthy that there has been a claim that "by increasing the magnitude of the mean field in 3-D simulations one finds that the transition from the isotropic 3-D scaling properties toward those observed in 2-D" [6]. (see also [7,175,6] for recent development in 2-D MHD turbulence). This claim has yet to be substantiated, however. We feel that the available numerical simulations [116,20,101] are reasonably consistent with the Goldreich & Sridhar [45] model that we review in the next section. It is also worth noting that the idea of an anisotropic (perpendicular) cascade has been incorporated into the framework of the reduced MHD approximation [161,146,113,182,4].

3.3 Goldreich-Sridhar Theory

We assume throughout this discussion that the rms turbulent velocity at the energy injection scale is comparable to the Alfvén speed of the mean field and consider only scales below the energy injection scale. Consequently, we are not

concerned with the problem of magnetic field generation, or the magnetic dynamo which is considered elsewhere in this volume (see also [111,132,72] for reviews, [107,123,12,15] for numerical calculations, [173,53,15] for suppression of the α dynamo effect in highly conducting fluids, and [73,11,176,100,185] for recent developments).

An ingenious model very similar in its beauty and simplicity to the Kolmogorov model has been proposed by Goldreich & Sridhar [45] (1995; hereinafter GS95) for incompressible MHD turbulence. They pointed out that motions perpendicular to the magnetic field lines mix them on a hydrodynamic time scale, i.e. at a rate $t_{cas}^{-1} \approx k_\perp v_l$, where $k_\perp$ is the wavevector component perpendicular to the local mean magnetic field and $l \sim k^{-1} (\approx k_\perp^{-1})$. These mixing motions couple to the wave-like motions parallel to magnetic field giving a *critical balance* condition

$$k_\parallel V_A \sim k_\perp v_k, \tag{9}$$

where $k_\parallel$ is the component of the wavevector parallel to the local magnetic field. When the typical $k_\parallel$ on a scale $k_\perp$ falls below this limit, the magnetic field tension is too weak to affect the dynamics and the turbulence evolves hydrodynamically, in the direction of increasing isotropy in phase space. This quickly raises the value of $k_\parallel$. In the opposite limit, when $k_\parallel$ is large, the magnetic field tension dominates, the error $\delta\omega$ in the matching conditions is reduced, and the nonlinear cascade is largely in the $k_\perp$ direction, which restores the critical balance.

If conservation of energy in the turbulent cascade applies locally in phase space then the energy cascade rate (v_l^2/t_{cas}) is constant (1): $(v_l^2)/(l/v_l) =$ constant. Combining this with the critical balance condition we obtain an anisotropy that increases with decreasing scale

$$k_\parallel \propto k_\perp^{2/3}, \tag{10}$$

and a Kolmogorov-like spectrum for perpendicular motions

$$v_l \propto l^{1/3}, \quad \text{or, } E(k) \propto k_\perp^{-5/3}, \tag{11}$$

which is not surprising since the magnetic field does not influence motions that do not bend it. At the same time, the scale-dependent anisotropy reflects the fact that it is more difficult for the weaker, smaller eddies to bend the magnetic field.

GS95 shows the duality of motions in MHD turbulence. Those perpendicular to the mean magnetic field are essentially eddies, while those parallel to magnetic field are waves. The critical balance condition couples these two types of motions.

3.4 Weak/Intermediate Turbulence

Let us reconsider the interaction of two wave packets moving oppositely along the mean magnetic field line. As in (4), the energy loss per collision is

$$\Delta E \sim (\dot{v^2})\Delta t \sim (v_l^3/l_\perp)(l_\parallel/V_A), \tag{12}$$

where we explicitly distinguish the parallel ($l_\parallel$) and the perpendicular size ($l_\perp$). The kinetic energy of the eddy is v_l^2. Therefore, the ratio of ΔE to E,

$$\zeta_l \equiv \frac{\Delta E}{v_l^2} \sim \frac{v_l l_\parallel}{V_A l_\perp} = \frac{v_l k_\perp}{V_A k_\parallel}, \tag{13}$$

characterizes the strength of the nonlinear interaction [45]. In the GS95 theory, $\zeta_l \sim 1$. This means that $V_A \approx V$ is required at the energy injection scale when energy injection is isotropic ($k_{\perp,L} \sim k_{\parallel,L}$). When this condition is satisfied, the turbulence is called *strong* turbulence.

There are some astrophysical situations, e.g. the Jovian magnetosphere [148], where the parameter ζ_L is much smaller than the unity over a broad range of length scales. Although as noted above the cascade will evolve in the direction of increasing ζ_l for decreasing l, and may reach the strong turbulent regime on very small scales. In this regime, the parallel cascade is strongly suppressed so the turbulence is qualitatively different from the strong turbulence discussed above. This is the *weak* turbulence regime. We do not discuss this type of turbulence here due to its limited astrophysical applicability and restricted inertial range. For more information, see [129,158,122,46,42]. Note that Galtier et al. [42] obtained $E(k) \sim k_\perp^{-2}$ (see also [46]).

4 Testing and Extending Incompressible Theory

Here we focus on recent direct numerical simulations related to the anisotropic structure of MHD turbulence. A discussion of earlier pioneering numerical simulations of MHD turbulence can be found in [140,107].

4.1 Scaling Laws

Despite its attractiveness, the Goldreich-Sridhar model is a conjecture that requires testing. The first such test was done by Cho & Vishniac [20] who used an incompressible pseudo-spectral MHD code with 256^3 grid points. In their simulations, they used $V_A = B_0/\sqrt{4\pi\rho} \sim V$. Their results for eddy shapes are shown in Fig. 4a, which shows a reasonable agreement with the predicted scale-dependent anisotropy of the turbulence (10) in the inertial range (i.e. the scales between the energy injection scale and the dissipation scale). They also obtained $E(k) \sim k^{-5/3}$ in the inertial range. Yet although the velocity does show the expected scaling, the magnetic field scaling is a bit more uncertain.

A subsequent numerical study by Maron & Goldreich [101] performed with a different code and in a different physical regime, namely, for $V_A \gg V$, also supported the GS95 model and clarified the role of pseudo-Alfvén and Alfvén modes. In particular, they confirmed that the pseudo-Alfvén modes are passively carried down the cascade through interactions with the Alfvén modes. They also showed that passive scalars adopt the same power spectrum as the velocity and magnetic field fluctuations. In addition, they addressed several issues about the

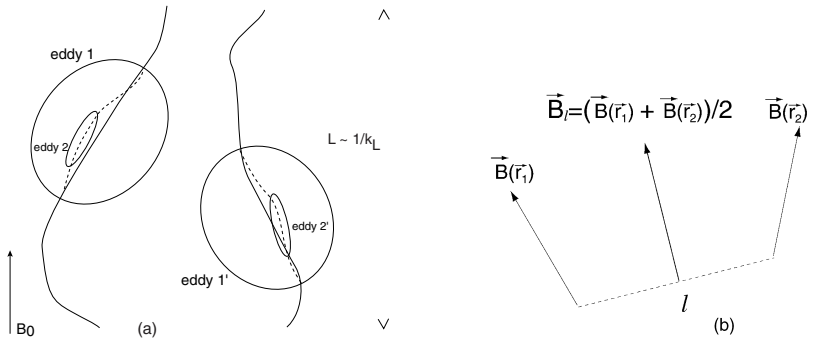


Fig. 1. (a) Eddies and local mean magnetic field. Local mean field is the properly averaged magnetic field near eddies under consideration. Eddies at different locations (e.g. eddies 1 & 1′) can have different local mean fields. Eddies of different sizes (e.g. eddies 1 & 2) can also have different local mean fields. When we do not take into account the local mean field, calculations hardly reveal true eddy structures. From [23]. (b) An example of calculating the local mean field. The direction of the local mean field $\mathbf{B}_l$ is obtained by the pair-wise average: $\mathbf{B}_l = (\mathbf{B}(\mathbf{r}_1) + \mathbf{B}(\mathbf{r}_2))/2$.

imbalanced cascade. Overall, they obtained results reasonably consistent with the GS95 model, but their energy spectra scale slightly differently: $E(k) \propto k^{-3/2}$. They attributed this result to intermittency.

Müller & Biskamp [116] studied MHD turbulence numerically in the regime of $V \gg V_A$ and obtained a Kolmogorov spectrum: $E(k) \sim k^{-5/3}$. They numerically studied the scaling exponents and obtained $\zeta_2 \approx 0.7$ and $\zeta_3 \approx 1$ (see the definition of the scaling exponents in the next section). The value of ζ_2 is consistent with the energy spectrum.

Other related recent numerical simulations include Matthaeus et al. [104] and Milano et al. [109]. Matthaeus et al. [104] showed that the anisotropy of low frequency MHD turbulence scales linearly with the ratio of perturbed and total magnetic field strength b/B $(= b/(b^2 + B_0^2)^{1/2})$.[10]

[10] In fact we can derive the GS95 scaling using this result. Although their analysis was based on comparing the strength of a uniform background field and the magnetic perturbations on all scales, we can reinterpret this result by assuming that the strength of random magnetic field at a scale l is b_l, and that the background field is the sum of all contributions from larger scales. Then Matthaeus et al.'s result becomes a prediction that the anisotropy $(k_\parallel/k_\perp)$ is proportional to (b_l/B). We can take the total magnetic field strength $B \sim$ constant as long as the background field is stronger than the perturbations on all scales. Since $b_l \sim (kE(k))^{1/2} \sim k_\perp^{-1/3}$, we obtain an anisotropy $(k_\parallel/k_\perp)$ proportional to $k_\perp^{-1/3}$, and $k_\parallel \propto k_\perp^{2/3}$. In this interpretation, smaller eddies are more elongated because they have a smaller b_l/B ratio.

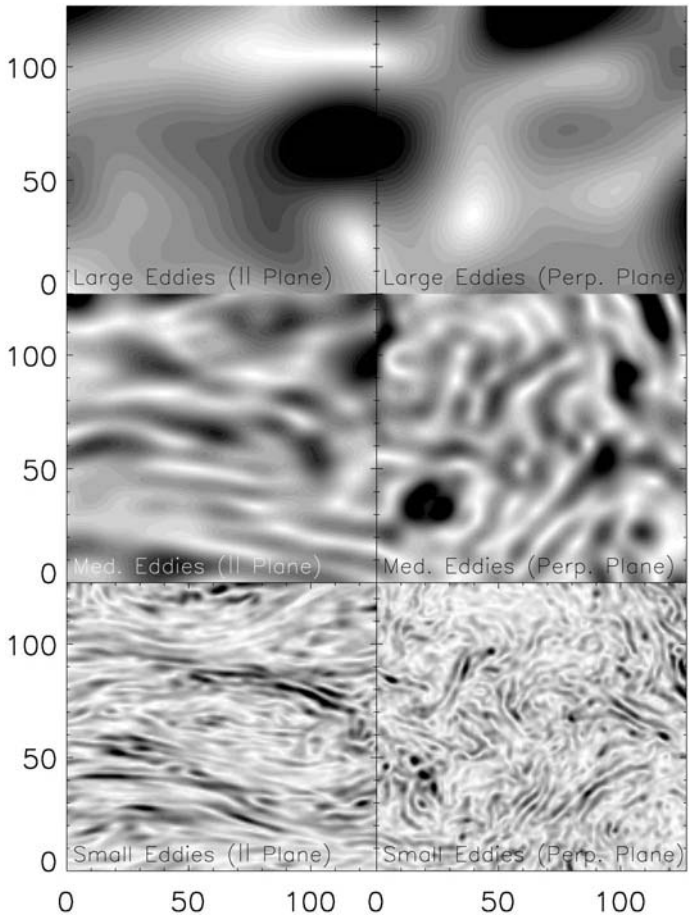


Fig. 2. Cross-sections of the data cube. (*Left panels*) $|\mathbf{b}_l|$ in a plane $\parallel$ to $\mathbf{B}_0$. (*Right panels*) $|\mathbf{b}_l|$ in a plane $\perp$ to $\mathbf{B}_0$. In the left panels, $\mathbf{B}_0$ is along the horizontal axis. *Large scale eddies* are obtained from the Fourier components with $1 \leq k < 4$. *Medium scale eddies* are obtained from the Fourier components with $4 \leq k < 16$. *Small scale eddies* are obtained from the Fourier components with $16 \leq k < 64$. The small scale eddies show a high degree of elongation in the parallel plane. However, they do not show a systematic behavior in the perpendicular plane.

All in all, numerical simulations so far have been largely, but not perfectly, consistent with the GS95 theory, e.g. the Kolmogorov-type scaling and the scale-dependent anisotropy ($k_\parallel \propto k_\perp^{2/3}$), and helped to extend it. An important point, which was not included in some of the earlier work, is that the scale dependent anisotropy can be measured only in a local coordinate frame which is aligned with the *locally averaged* magnetic field direction [20]. The necessity of using a local frame is due to the fact that eddies are aligned along the local mean magnetic fields, rather than the global mean field $\mathbf{B}_0$. Since smaller scale eddies

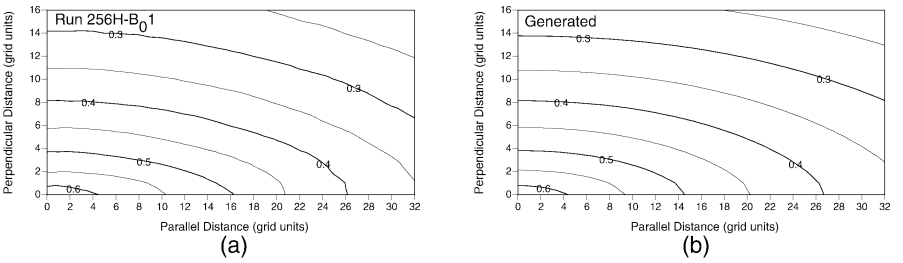


Fig. 3. (a) Velocity correlation function (VCF) from a simulation. The contours represent shape of different size eddies. The smaller contours (or, eddies) are more elongated. (b) VCF generated from (14). From [23].

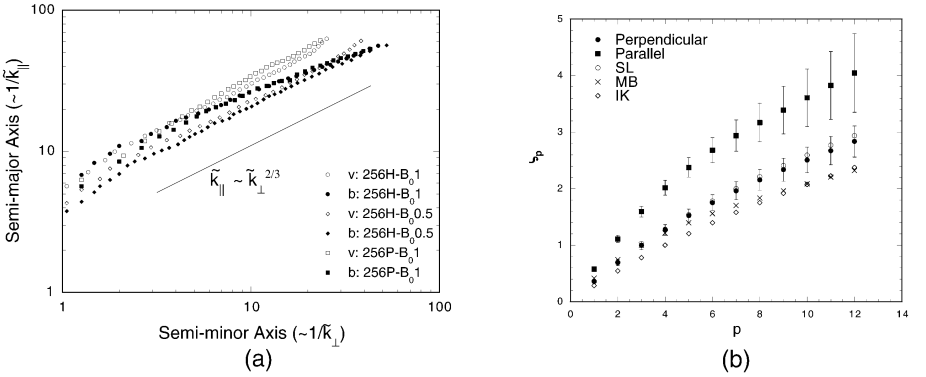


Fig. 4. (a) Semi-major axis and semi-minor axis of contours of the VCF shown in Fig. 3a. The results support $k_\parallel \propto k_\perp^{2/3}$. From [20]. (b) Scaling exponents. From [23].

are weaker, and more anisotropic, measurements of eddy shape based on a global coordinate system are always dominated by the largest eddies in the simulation. Figure 1 illustrates the concept of a local frame and one way to identify it. Further research in Cho, Lazarian, & Vishniac [23] (2002a: hereinafter CLV02a) showed that in the local system of reference the mixing motions perpendicular to the magnetic field have statistics identical to hydrodynamic turbulence (cf. Müller & Biskamp [116]).

Figure 2 shows the shapes of eddies of different sizes. Left panels show an increased anisotropy as we move from the top (large eddies) to the bottom (small eddies). The horizontal axes of the left panels are parallel to $\mathbf{B_0}$. Structures in the perpendicular plane do not show a systematic elongation.

Figure 3a and Fig. 4a quantify some of these results. The contours of the correlation function obtained in [20] are shown in Fig. 3a and are consistent with the predictions of the GS95 model. Figure 4a shows that the semi-major axis ($1/k_\parallel$) is proportional to the 2/3 power of the semi-minor axis ($1/k_\perp$), implying

that $k_\parallel \propto k_\perp^{2/3}$. While the one dimensional energy spectrum follows Kolmogorov spectrum, $E(k) \propto k^{-5/3}$, CLV02a showed that the 3D energy spectrum can be fit by

$$P(k_\perp, k_\parallel) = (B_0/L^{1/3}) k_\perp^{-10/3} \exp\left(-L^{1/3}\frac{k_\parallel}{k_\perp^{2/3}}\right), \tag{14}$$

where B_0 is the strength of the mean field and L is the scale of energy injection. The velocity correlation from the 3D spectrum provides an excellent fit to the numerical data (Fig. 3b). This allows practical applications illustrated in Sect. 5.

4.2 Intermittency

Intermittency refers to the non-uniform distribution of dissipative structures. Intermittency has an important dynamical consequence: it affects the energy spectrum. Highly intermittent turbulent structures were invoked by Falgarone et al. [37] and Joulain et al. [63] as the primary location of endothermic interstellar chemical reactions.

Maron & Goldreich [101] studied the intermittency of dissipation structures in MHD turbulence using the fourth order moments of the Elsässer fields and the gradients of the fields. Their simulations show strong intermittent structures. CLV02a used a different, but complementary, method to study intermittency, based on the higher order longitudinal structure functions. They found that by this measure the intermittency of velocity field in MHD turbulence across local magnetic field lines is as strong as, but not stronger than, hydrodynamic turbulence.

In fully developed hydrodynamic turbulence, the (longitudinal) velocity structure functions $S_p =< ([\mathbf{v}(\mathbf{x}+\mathbf{r}) - \mathbf{v}(\mathbf{x})] \cdot \hat{\mathbf{r}})^p >\equiv< \delta v_L^p(\mathbf{r}) >$ are expected to scale as r^{ζ_p}. For example, the classical Kolmogorov phenomenology (K41) predicts $\zeta_p = p/3$. The (exact) result for p=3 is the well-known 4/5-relation: $< \delta v_L^3(\mathbf{r}) >= -(4/5)\epsilon r$, where ϵ is the energy injection rate (or, energy dissipation rate) (see e.g. [41]). On the other hand, considering intermittency, She & Leveque ([152]; hereinafter S-L) proposed a different scaling relation: $\zeta_p^{SL} = p/9 + 2[1 - (2/3)^{p/3}]$. Note that the She-Leveque model also implies $\zeta_3 = 1$.

So far in MHD turbulence, to the best of our knowledge, there is no rigorous intermittency theory which takes into account scale-dependent anisotropy. Politano & Pouquet [135] have developed an MHD version of the She-Leveque model:

$$\zeta_p^{PP} = \frac{p}{g}(1-x) + C\left(1 - (1 - x/C)^{p/g}\right), \tag{15}$$

where C is the co-dimension of the dissipative structure, g is related to the scaling $v_l \sim l^{1/g}$, and x can be interpreted as the exponent of the cascade time $t_{cas} \propto l^x$. (In fact, g is related to the scaling of Elsässer variable z=$v \pm b$: $z_l \sim l^{1/g}$.) In the framework of the IK theory, where $g = 4$, $x = 1/2$, and $C = 1$ when the dissipation structures are sheet-like, their model of intermittency

becomes $\zeta_p^{IK} = p/8 + 1 - (1/2)^{p/4}$. On the other hand, Müller & Biskamp [116] performed numerical simulations on decaying isotropic MHD turbulence and obtained Kolmogorov-like scaling ($E(k) \sim k^{-5/3}$ and $t \sim l^{2/3}$) and sheet-like dissipation structures, which implies $g = 3$, $C = 1$, and $x = 2/3$. From (15), they proposed that

$$\zeta_p^{MB} = p/9 + 1 - (1/3)^{p/3}. \tag{16}$$

The intermittency results from [23] are shown in Fig. 4b. The filled circles represent the scaling exponents of longitudinal velocity structure functions in directions *perpendicular* to the local mean magnetic field. It is surprising that the scaling exponents are so close to the original (i.e. hydrodynamic) S-L model. This raises an interesting question. In the simulations of CLV02a, $t_{cas} \propto l^{2/3}$ and $E(k) \propto k^{-5/3}$ scaling is observed. It is observed that MHD turbulence has sheet-like dissipation structures [137]. Therefore, the parameters for CLV02a simulations should be the same as those of Müller & Biskamp's (i.e. $g = 3$, $C = 1$, and $x = 2/3$) rather than suggesting $C = 2$. We believe this difference stems from the different simulation settings (Müller & Biskamp's turbulence is isotropic and CLV02a's is anisotropic) and the way the turbulence is analyzed (global versus local frame). In fact, we expect the small-scale behavior of MHD turbulence should not depend on whether the largest scale fields are uniform or have the same scale of organization as the largest turbulent eddies. Nevertheless, given the limited dynamical range available in these simulations, it would not be surprising if the scale of the magnetic field has an impact on the intermittency statistics. It is not clear how scale-dependent anisotropy changes the intermittency model in (15) and we will not discuss this issue further. Instead, we simply stress that a striking similarity exists between ordinary hydrodynamic turbulence and MHD turbulence in perpendicular directions, which further supports the picture of the GS95 turbulence.[11]

In Fig. 4b, we also plot the scaling exponents (represented by filled squares) of longitudinal velocity structure functions *along* directions of the local mean magnetic field. Although we show only the exponents of longitudinal structure functions, those of transverse structure functions follow a similar scaling law. Evidently intermittency along the local mean magnetic field directions is completely different from the scaling predicted by previous (isotropic) models. Roughly speaking, the scaling exponents along the directions of local magnetic field are 1.5 times larger than those of perpendicular directions. This has an obvious similarity to the scaling of eddy shapes.

The second order exponent ζ_2 is related to the the 1-D energy spectra: $E(k_\perp) \propto k_\perp^{-(1+\zeta_2)}$. Previous 2-D driven MHD calculations for $B_0 = 0$ by Politano, Pouquet, & Carbone [136] also found $\zeta_2 \sim 0.7$. However, Biskamp, & Schwarz [7] obtained $\zeta_2 \sim 0.5$ from decaying 2-D MHD calculations with $B_0 = 0$. The result of CLV02a suggests that ζ_2 is closer to 2/3, rather than to 1/2. (It is not clear whether or not the scaling exponents follow the original S-L model exactly. At the same time, our calculation shows that the original S-L model can

[11] MG01 attributed the deviation of their spectrum from the Kolmogorov-type to the intermittency present in their simulation.

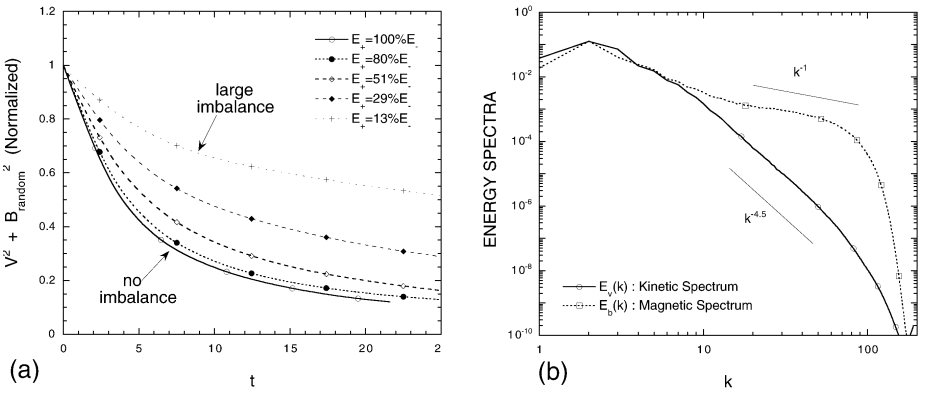


Fig. 5. (a) Imbalanced Decay. When imbalance is large, turbulence decays slow. From CLV02a. (b) Viscous damped regime. A new inertial range emerges below the viscous cut-off at $k \sim 7$. From Cho, Lazarian, & Vishniac (2002b).

be a good approximation for our scaling exponents. The S-L model predicts that $\zeta_2 \sim 0.696$.) This is equivalent to our earlier claim that our result supports the scaling law $E(k_\perp) \propto k_\perp^{-5/3}$ at least for velocity. For the parallel directions, the results support $E(k_\parallel) \propto k_\parallel^{-2}$ although the uncertainty is large.

5 Damping of Turbulence

5.1 Imbalanced Cascade

Turbulence plays a critical role in molecular cloud support and star formation and the issue of the time scale of turbulent decay is vital for understanding these processes. If MHD turbulence decays quickly, then serious problems face the researchers attempting to explain important observational facts, e.g. turbulent motions seen within molecular clouds without star formation [119] and rates of star formation [106]. Earlier studies attributed the rapid decay of turbulence to compressibility effects [99]. GS95 predicts and numerical simulations, e.g. CLV02a, confirm that MHD turbulence decays rapidly even in the incompressible limit. This can be understood if mixing motions perpendicular to magnetic field lines are considered. As we discussed earlier, such eddies, as in hydrodynamic turbulence, decay in one eddy turnover time.

Below we consider the effect of imbalance [103,165,50,60,101] on the turbulence decay time scale. Duality of the MHD turbulence means that the turbulence can be described by colliding wave packets. 'Imbalance' means that the flux of wave packets traveling in one direction is significantly larger than those traveling in the other direction. In the ISM, many energy sources are localized both in space and time. For example, in terms of energy injection, stellar outflows are essentially point energy sources. With these localized energy sources, it is natural that interstellar turbulence be typically imbalanced.

Here we show results of the CLV02a study that demonstrate that imbalance does extend the lifetime of MHD turbulence (Fig. 5a). We used a run on a grid of 144^3 to investigate the decay time scale. For initial conditions we took a data cube from a driven turbulence simulation. The initial data cube contains both upward (denoted as +) and downward moving waves (denoted as −). To adjust the degree of initial imbalance, we either increased or decreased the energy of the upward moving components and, by turning off the forcing terms, let the turbulence decay. Note that the initial energy is normalized to 1. The y-axis is the normalized total (=up + down) energy.

The dependence of the turbulence decay time on the degree of imbalance is an important finding. To what degree the results persist in the presence of compressibility is the subject of our current research. It is obvious that results of CLV02a are applicable to incompressible, namely, Alfvén motions.[12] We show in Sect. 6.3 that the Alfvén motions are essentially decoupled from compressible modes. As the result we expect that the turbulence decay time may be substantially longer than one eddy turnover time provided that the turbulence is imbalanced.

5.2 Ion-Neutral Damping: A New Regime of Turbulence

In hydrodynamic turbulence viscosity sets a minimal scale for motion, with an exponential suppression of motion on smaller scales. Below the viscous cutoff the kinetic energy contained in a wavenumber band is dissipated at that scale, instead of being transferred to smaller scales. This means the end of the hydrodynamic cascade, but in MHD turbulence this is not the end of magnetic structure evolution. For viscosity much larger than resistivity, $\nu \gg \eta$, there will be a broad range of scales where viscosity is important but resistivity is not. On these scales magnetic field structures will be created by the shear from non-damped turbulent motions, which amounts essentially to the shear from the smallest undamped scales. The created magnetic structures would evolve through generating small-scale motions. As a result, we expect a power-law tail in the energy distribution, rather than an exponential cutoff. To our best knowledge, this is a completely new regime for MHD turbulence.

In partially ionized gas neutrals produce viscous damping of turbulent motions. In the Cold Neutral Medium (see Draine & Lazarian [34] for a list of the idealized phases) this produces damping on the scale of a fraction of a parsec. The magnetic diffusion in those circumstances is still negligible and exerts an influence only at the much smaller scales, $\sim 100km$. Therefore, there is a large

[12] In the long run, the imbalance will be defeated by the parametric instability, which develops through formation of density inhomogeneities within the beam of Alfvén waves [26,47,62,25]. However, this instability takes many wave periods to be established. A similar argument can be applied when we consider completely imbalanced cascade. That is, even in the completely imbalanced cascade decay of energy can occur due to non-linear steepening of waves. But, this will be very slow for Alfvén waves.

range of scales where the physics of the turbulent cascade is very different from the GS95 picture.

Cho, Lazarian, & Vishniac [24] have explored this regime numerically. Here we used a grid of 384^3 and a physical viscosity for velocity damping. The kinetic Reynolds number is around 100. With this Reynolds number, viscous damping occurs around $k \sim 7$, or about $\sim 1/7$ of the width of the computational box. We minimized magnetic diffusion through the use of a hyper-diffusion term of order 3. To test for possible "bottle neck" effects we also did simulations with normal magnetic diffusion and reproduced similar results but only over a reduced dynamical range. The bottleneck effect is a common feature in numerical hydrodynamic simulations with hyperviscosity. See [8] for MHD simulations.

In Fig. 5b, we plot energy spectra. The spectra consist of several parts. First, the peak of the spectra corresponds to the energy injection scale. Second, for $2 < k < 7$, kinetic and magnetic spectra follow a similar slope. This part is more or less a severely truncated inertial range for undamped MHD turbulence. Third, the magnetic and kinetic spectra begin to decouple at $k \sim 7$. Fourth, after $k \sim 20$, a new *damped-scale inertial range* emerges. In the new inertial range, magnetic energy spectrum follows

$$\textit{New Regime:} \qquad E_b(k) \propto k^{-1}, \tag{17}$$

implying considerable magnetic structures below the viscous damping scale. The velocity power spectrum steepens in this regime, but does not fall exponentially.

Figure 6a shows that the small magnetic structures are highly intermittent in the viscous-damped regime. Here we obtained the small scale magnetic field by eliminating Fourier modes with $k < 20$. We can see that the typical radius of curvature of field lines in the plane is much larger than the typical perpendicular scale for field reversal. The typical radius of curvature of field lines corresponds to the viscous damping scale, indicating that stretched structures are results of the shearing motions at the viscous damping scale. There is no preferred direction for these elongated structures. A similar plot for ordinary (*not* viscously damped) MHD turbulence (Fig. 6b) shows much less intermittent structures. We remove Fourier modes with $k < 20$ also in Fig. 6b.

A theoretical model for this new regime and its consequences for stochastic reconnection [89] will be found in an upcoming paper (Lazarian, Vishniac, & Cho 2002). Here we summarize the main points of the model. We begin by noting that the strong intermittency seen in this regime suggests a new parameter, f_l, which is the volume filling fraction of structures with scales comparable to l. This in turn implies that we need to distinguish between volume averaged means and typical values of velocity and magnetic field perturbations in fraction of space where they are concentrated. We denote the latter with a '$\hat{}$', so that $v_l^2 = f_l \hat{v}_l^2$, and $b_l^2 = f_l \hat{b}_l^2$. This model does not include information about the range of field strengths or velocities with structure on a scale l in the bulk of the volume, aside from assuming that they are sufficiently weak that they do not contribute to any global averages.

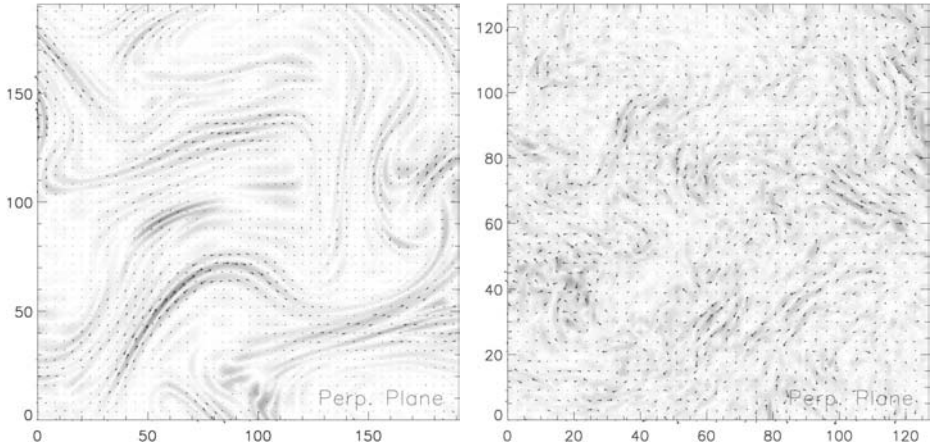


Fig. 6. (*Left*) Viscous-damped turbulence. Strength of magnetic field in a plane perpendicular to $\mathbf{B}_0$. Arrows are magnetic fields in the plane. Only a part of the plane is shown. Note highly intermittent structures. From [24]. (*Right*) Same as *Left*, but for ordinary (not viscous-damped) MHD turbulence. Structures are less intermittent.

The second new fundamental parameter in this model is the eddy turn over rate at the damping scale, k_d, i.e.

$$\tau_s^{-1} \sim k_d v_d \sim k_d^2 \nu_n, \tag{18}$$

where v_d is the velocity at $l_d \sim 1/k_d$ and ν_n is the viscosity of the plasma due to neutral particles, which differs from the viscosity of neutral fluid by the ratio of atomic to total densities.

Since motions on smaller scales are strongly damped, the cascade of magnetic energy to smaller scales is due to motions on the damping scale. This implies that

$$b_l^2/\tau_s \sim \text{ constant}, \tag{19}$$

so that b_l is maintained approximately at the level of magnetic field at the damping scale, b_d.

This folding and refolding, perpendicular to the mean field direction, has a weak effect on the field line curvature. (This result is also seen in the simulations.) The decrease in the structure length l is due to an increase in the magnetic field gradient perpendicular to the mean field direction. The resulting magnetic pressure gradients are balanced by plasma pressure gradients. Thus on the scales below the viscous cutoff, the tension forces are balanced by viscous drag, i.e.

$$\frac{\rho \nu_n}{l^2} \hat{v}_l \sim \max[k_d \hat{b}_l, k_\parallel B_0] \hat{b}_l. \tag{20}$$

Finally, this dynamic equilibrium can be maintained only if the small scale motions are strong enough to counteract the shear, τ_s^{-1}. In other words, $\hat{v}_l/l \sim \tau_s^{-1}$.

Combining these results we see that $f_l \sim k_d l$, $\hat{b}_l \sim b_d (k_d l)^{-1/2}$, and $v_l \propto l^{3/2}$. These scaling laws are at least qualitatively consistent with the simulation results, although the velocity power spectrum may be slightly steeper than the model prediction.

Unfortunately, a realistic treatment of the ISM requires an explicit recognition of the two fluid nature of the partially ionized plasma, rather than simply representing neutral drag with an effective viscosity. Here we are beyond the reach of available simulations, and need to rely on an extension of the scaling arguments given above.

First, at sufficiently small values of l, the ambipolar diffusion time ($\sim k^2 t_{in} (b_l^2/\rho_{tot})(\rho_n/\rho_i)$, where t_{in} = ion-neutral collision time) will become less than τ_s, and the magnetic pressure gradients will be supported entirely by the ionized particle pressure. This means that only a fraction, $\sim \rho_i/\rho_{tot}$, of the energy will continue the cascade to smaller scales.

Next, for some ion-neutral collision time, t_{in}, there will be some decoupling scale, l_c, where

$$\rho \nu_n / l_c^2 \sim \rho_i / t_{in}. \tag{21}$$

At smaller scales the ions will be dragged through a more or less uniform neutral background. The argument above needs to be modified by replacing the viscous drag coefficient in (20) with the right-hand-side of (21). This gives

$$f_l \propto l^{-1}, \tag{22}$$

$$\hat{b}_l \propto l^{1/2}, \quad \text{and} \quad v_l \propto l^{1/2}. \tag{23}$$

Finally, at some sufficiently small scale the filling factor will rise to unity, and the gradients in the magnetic field will become strong enough that neutral drag can be ignored. These conditions are satisfied simultaneously when

$$k_d l \sim t_{in}/\tau_s. \tag{24}$$

Below this scale we expect to see a resumption of the turbulent cascade, now involving only the ionized component of the plasma, down to scales where plasma resistivity and viscosity finally dissipate the remaining energy. Since the longest cascade time for this regime is $t_{in} \ll \tau_s$, we expect this small scale turbulence to be intermittent, with a duty cycle $\sim t_{in}/\tau_s$.

All the consequences of the new regime of the MHD turbulence have not yet been appreciated, but we expect that it will have a substantial impact on our understanding of the interstellar physics. Moreover, the treatment given above actually applies only when ρ_i/ρ is not very small. Otherwise the decoupling scale can be larger than the viscous damping scale.

6 Compressible Turbulence

For the rest of the review, we consider MHD turbulence of a single conducting fluid. While the GS95 model describes incompressible MHD turbulence well,

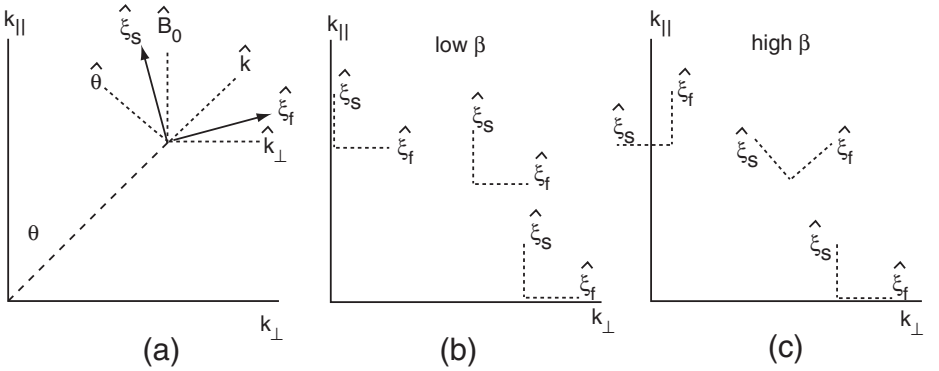


Fig. 7. (a) Directions of fast and slow basis vectors. $\hat{\xi}_f$ and $\hat{\xi}_s$ represent the directions of displacement of fast and slow modes, respectively. In the fast basis ($\hat{\xi}_f$) is always between $\hat{\mathbf{k}}$ and $\hat{\mathbf{k}}_\perp$. In the slow basis ($\hat{\xi}_s$) lies between $\hat{\theta}$ and $\hat{\mathbf{B}}_0$. Here, $\hat{\theta}$ is perpendicular to $\hat{\mathbf{k}}$ and parallel to the wave front. All vectors lie in the same plane formed by $\mathbf{B}_0$ and $\mathbf{k}$. On the other hand, the displacement vector for Alfvén waves (not shown) is perpendicular to the plane. (b) Directions of basis vectors for a very small β drawn in the same plane as in (a). The fast bases are almost parallel to $\hat{\mathbf{k}}_\perp$. (c) Directions of basis vectors for a very high β. The fast basis vectors are almost parallel to $\mathbf{k}$. The slow waves become pseudo-Alfvén waves.

no universally accepted theory exists for compressible MHD turbulence despite various attempts (e.g., [56]). Earlier numerical simulations of compressible MHD turbulence covered a broad range of astrophysical problems, such as the decay of turbulence (e.g. [98,160]) or turbulent modeling of the ISM (see recent review [170]; see also [133,168,134,169] for earlier 2D simulations and [125,126,131,67,9] for recent 3D simulations). In what follows, we concentrate on the fundamental properties of compressible MHD.

6.1 Alfvén, Slow, and Fast Modes

Let us start by reviewing different MHD waves. In particular, we describe the Fourier space representation of these waves. The real space representation can be found in papers on modern shock-capturing MHD codes (e.g. [13,147]). For the sake of simplicity, we consider an isothermal plasma. Figure 7 and Fig. 8 give schematics of slow and fast waves. For slow and fast waves, $\mathbf{B}_0$, $\mathbf{v_k}$ ($\propto \xi$), and $\mathbf{k}$ are in the same plane. On the other hand, for Alfvén waves, the velocity of the fluid element $(\mathbf{v_k})_A$ is orthogonal to the $\mathbf{B}_0 - \mathbf{k}$ plane.

As before, the Alfvén speed is $V_A = B_0/\sqrt{4\pi\rho_0}$, where ρ_0 is the average density. Fast and slow speeds are

$$c_{f,s} = \left[\frac{1}{2}\left\{a^2 + V_A^2 \pm \sqrt{(a^2+V_A^2)^2 - 4a^2V_A^2\cos^2\theta}\right\}\right]^{1/2}, \qquad (25)$$

where θ is the angle between $\mathbf{B}_0$ and $\mathbf{k}$. See Table 1 for the definition of other variables. When β ($\beta = P_g/P_B{=}2a^2/V_A^2$; $P_g{=}$ gas pressure, $P_B{=}$ magnetic pres-

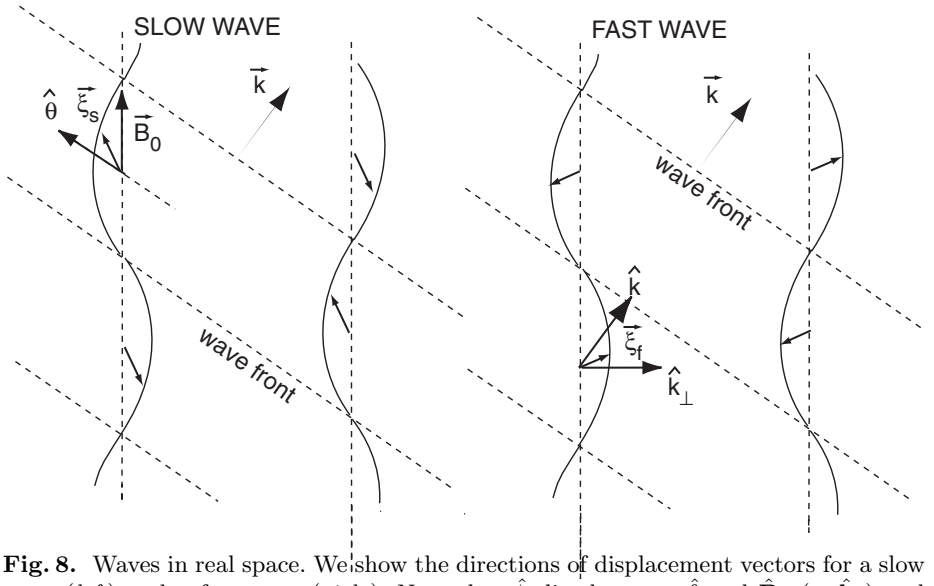


Fig. 8. Waves in real space. We show the directions of displacement vectors for a slow wave (*left*) and a fast wave (*right*). Note that $\hat{\xi}_s$ lies between $\hat{\theta}$ and $\hat{\mathbf{B}}_0$ $(= \hat{\mathbf{k}}_\parallel)$ and $\hat{\xi}_f$ between $\hat{\mathbf{k}}$ and $\hat{\mathbf{k}}_\perp$. Again, $\hat{\theta}$ is perpendicular to $\hat{\mathbf{k}}$ and parallel to the wave front. Note also that, for the fast wave, for example, density (inferred by the directions of the displacement vectors) becomes higher where field lines are closer, resulting in a strong restoring force, which is why fast waves are faster than slow waves.

Table 1. Notations for compressible turbulence

Notation	Meaning
a, c_s, c_f, V_A	sound, slow, fast, and Alfvén speed
δV, $(\delta V)_s$, $(\delta V)_f$, $(\delta V)_A$	random (rms) velocity Previously we used V for the rms velocity
v_l, $(v_l)_s$, $(v_l)_f$, $(v_l)_A$	velocity at scale l
$\mathbf{v_k}$, $(\mathbf{v_k})_s$, $(\mathbf{v_k})_f$, $(\mathbf{v_k})_A$	velocity vector at wavevector $\mathbf{k}$
$\hat{\mathbf{B}}_0$ $(=\hat{\mathbf{k}}_\parallel)$, $\hat{\mathbf{k}}_\perp$, $\hat{\mathbf{k}}$, $\hat{\theta}$, ...	unit vectors
ξ_s, ξ_f	displacement vectors

sure; hereinafter $\beta =$ average $\beta \equiv \bar{P}_g/\bar{P}_B$) goes to zero, we have

$$c_f \approx V_A,$$
$$c_s \approx a\cos\theta. \tag{26}$$

Figure 7 shows directions of displacement (or, directions of velocity) vectors for these three modes. We will call them the basis vectors for these modes. The Alfvén basis is perpendicular to both $\hat{\mathbf{k}}$ and $\hat{\mathbf{B}}_0$, and coincides with the azimuthal vector $\hat{\phi}$ in a spherical-polar coordinate system. Here hatted vectors

are unit vectors. The fast basis $\hat{\xi}_f$ lies *between* $\hat{\mathbf{k}}$ and $\hat{\mathbf{k}}_\perp$:

$$\hat{\xi}_f \propto \frac{1-\sqrt{D}+\beta/2}{1+\sqrt{D}-\beta/2}\left[\frac{k_\perp}{k_\parallel}\right]^2 k_\parallel \hat{\mathbf{k}}_\parallel + k_\perp \hat{\mathbf{k}}_\perp, \tag{27}$$

where $D=(1+\beta/2)^2-2\beta\cos^2\theta$, and β is the averaged β ($=\bar{P}_g/\bar{P}_B$). The slow basis $\hat{\xi}_s$ lies *between* $\hat{\theta}$ and $\hat{\mathbf{B}}_0$ ($=\hat{\mathbf{k}}_\parallel$):

$$\hat{\xi}_s \propto k_\parallel \hat{\mathbf{k}}_\parallel + \frac{1-\sqrt{D}-\beta/2}{1+\sqrt{D}+\beta/2}\left[\frac{k_\parallel}{k_\perp}\right]^2 k_\perp \hat{\mathbf{k}}_\perp. \tag{28}$$

The two vectors $\hat{\xi}_f$ and $\hat{\xi}_s$ are mutually orthogonal. Proper normalizations are required for both bases to make them unit-length.

When β goes to zero (i.e. the magnetically dominated regime), $\hat{\xi}_f$ becomes parallel to $\hat{\mathbf{k}}_\perp$ and $\hat{\xi}_s$ becomes parallel to $\hat{\mathbf{B}}_0$ (Fig. 7b). The sine of the angle between $\hat{\mathbf{B}}_0$ and $\hat{\xi}_s$ is $(\beta/2)\sin\theta\cos\theta$. When β goes to infinity (i.e. gas pressure dominated regime)[13], $\hat{\xi}_f$ becomes parallel to $\hat{\mathbf{k}}$ and $\hat{\xi}_s$ becomes parallel to $\hat{\theta}$ (Fig. 7c). This is the incompressible limit. In this limit, the slow mode is sometimes called the pseudo-Alfvén mode [45].

6.2 Theoretical Considerations

Here we address the issue of mode coupling in a low β plasma. It is reasonable to suppose that in the limit where $\beta \gg 1$ turbulence for Mach numbers ($M_s = \delta V/a$) less than unity should be largely similar to the exactly incompressible regime. Thus, Lithwick & Goldreich [96] conjectured that the GS95 relations are applicable to slow and Alfvén modes with the fast modes decoupled. They also mentioned that this relation can carry on for low β plasmas. For $\beta \gg 1$ and $M_s > 1$, we are in the regime of hydrodynamic compressible turbulence for which no theory exists, as far as we know.

In the diffuse interstellar medium β is typically less than unity. In addition, it is ~ 0.1 or less for molecular clouds. There are a few simple arguments suggesting that MHD theory can be formulated in the regime where the Alfvén Mach number ($\equiv \delta V/V_A$) is less than unity, although this is not a universally accepted assumption. Alfvén modes describe incompressible motions. Arguments in GS95 are suggestive that the coupling of Alfvén to fast and slow modes will be weak. Consequently, we expect that in this regime the Alfvén cascade should follow the GS95 scaling. Moreover the slow waves are likely to evolve passively [96]. For $a \ll V_A$ their nonlinear evolution should be governed by Alfvén modes so that we expect the GS95 scaling for them as well. The phase velocity of Alfvén waves and slow waves depend on a factor of $\cos\theta$ and this enables modulation of the slow waves by the Alfvén ones. However, fast waves do not have this factor and therefore cannot be modulated by the changes of the magnetic field direction

[13] In this section, we assume that external mean field is strong (i.e. $V_A > (\delta V)$) but finite, so that $\beta \to \infty$ means the gas pressure $\bar{P}_g \to \infty$.

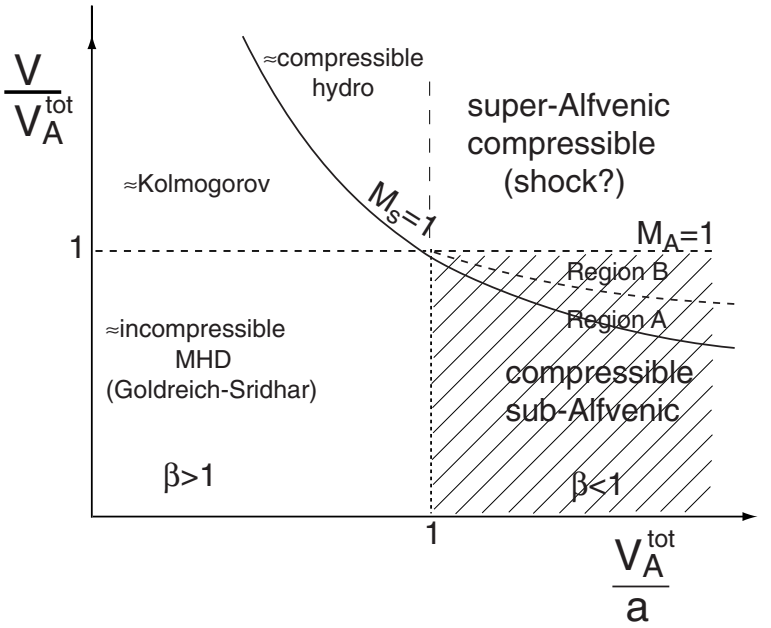


Fig. 9. Different regimes of MHD turbulence. We consider the compressible sub-Alfvénic regime (shaded region). In this figure, V_A^{tot} represents the total Alfvén speed ($= \sqrt{B_0^2 + \delta B^2}/\sqrt{4\pi\rho}$). $V \equiv V_{flow} = \delta V$. Cho & Vishniac [19] argued that, even when the external field is weak, small scales can follow the GS95-like scaling in incompressible MHD regime. Similarly, for compressible case, the small scales of super-Alfvénic compressible turbulence are expected to fall in the sub-Alfvénic compressible regime. Moreover, winding of magnetic field by turbulence increases the magnetic field energy and the super-Alfvénic turbulence becomes more and more magnetically dominated with $V/V_A \to 1$. Region B is the region where, within the density fluctuations, the velocities can get super-Alfvénic. In the figure, we used (33) to determine the borderline between region A and B.

associated with Alfvén waves. The coupling between the modes is through the modulation of the local Alfvén velocity and therefore is weak.

For Alfvén Mach number (M_A) larger than unity a shock-type regime is expected. However, generation of magnetic field by turbulence [19] is expected for such a regime. It will make the steady state turbulence approach $M_A \sim 1$.[14] Therefore in Cho & Lazarian [21] we consider turbulence in the limit $M_s > 1$, $M_A < 1$, and $\beta < 1$ (Fig. 9). For these simulations, we mostly used $M_s \sim 2.2$, $M_A \sim 0.7$, and $\beta \sim 0.2$. The Alfvén speed of the mean external field is similar to the rms velocity ($V_A = 1, \delta V \sim 0.7, a = \sqrt{0.1}$), and we used an isothermal equation of state.

Although the scaling relations presented below are applicable to sub-Alfvénic turbulence, we cautiously speculate that small scales of super-Alfvénic turbu-

[14] We suspect that simulations that show super-Alfvénic turbulence is widely spread in the ISM might not evolve for a long enough time to reach the steady state.

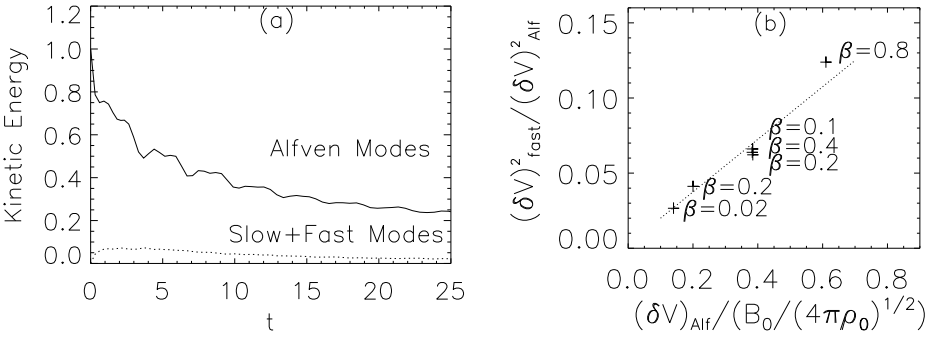


Fig. 10. (a) Decay of Alfvénic turbulence. The generation of fast and slow waves is not efficient. $\beta \sim 0.2$, $M_s \sim 3$. (b) The ratio of $(\delta V)_f^2$ to $(\delta V)_A^2$. The stronger the external field (B_0) is, the more suppressed the coupling is. The ratio is not sensitive to β. From [21]

lence might follow similar scalings. Boldyrev, Nordlund, & Padoan [10] obtained energy spectra close to $E(k) \sim k^{-1.74}$ in solenoidally driven super-Alfvénic supersonic turbulence simulations. The spectra are close to the Kolmogorov spectrum ($\sim k^{-5/3}$), rather than shock-dominated spectrum ($\sim k^{-2}$). This result might imply that small scales of super-Alfvénic MHD turbulence can be described by our sub-Alfvénic model presented below, which predicts Kolmogorov-type spectra for Alfvén and slow modes.

6.3 Coupling of MHD Modes and Scaling of Alfvén Modes

Alfvén modes are not susceptible to collisionless damping (see [156,110] and references therein) that damps slow and fast modes. Therefore, we mainly consider the transfer of energy from Alfvén waves to compressible MHD waves (i.e. to the slow and fast modes).

In Cho & Lazarian [21], we carry out simulations to check the strength of the mode-mode coupling. We first obtain a data cube from a driven compressible numerical simulation with $B_0/\sqrt{4\pi\rho_0} = 1$. Then, after turning off the driving force, we let the turbulence decay. We go through the following procedures before we let the turbulence decay. We first remove slow and fast modes in Fourier space and retain only Alfvén modes. We also change the value of $\mathbf{B}_0$ preserving its original direction. We use the same constant initial density ρ_0 for all simulations. We assign a new constant initial gas pressure P_g [15]. After doing all these procedures, we let the turbulence decay. We repeat the above procedures for different values of B_0 and P_g. Figure 10a shows the evolution of the kinetic energy

[15] The changes of both B_0 and P_g preserve the Alfvén character of perturbations. In Fourier space, the mean magnetic field ($\mathbf{B}_0$) is the amplitude of $\mathbf{k} = \mathbf{0}$ component. Alfvén components in Fourier space are for $\mathbf{k} \neq \mathbf{0}$ and their directions are parallel/anti-parallel to $\hat{\xi}_A$ ($= \hat{\mathbf{B}}_0 \times \hat{\mathbf{k}}_\perp$). The direction of $\hat{\xi}_A$ does not depend on the magnitude of B_0 or P_g.

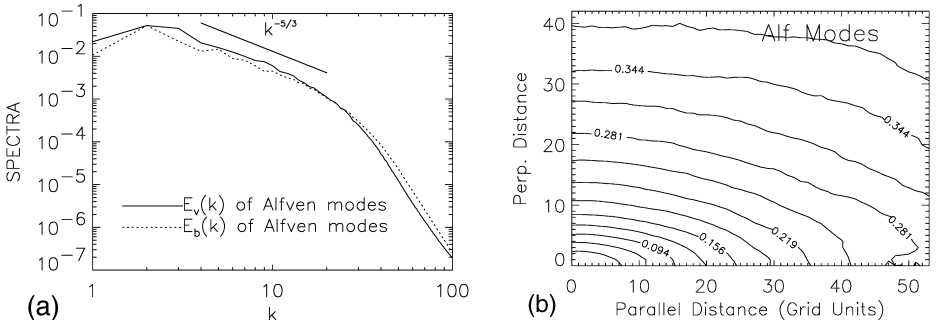


Fig. 11. (a) Alfvén spectra follow a Kolmogorov-like power law. (b) The second-order structure function ($SF_2 =< \mathbf{v}(\mathbf{x}+\mathbf{r}) - \mathbf{v}(\mathbf{x}) >$) for Alfv́en velocity shows anisotropy similar to the GS95. Conturs represent eddy shapes. From [21].

of a simulation. The solid line represents the kinetic energy of Alfvén modes. It is clear that Alfvén waves are poorly coupled to the compressible modes, and do not generate them efficiently [16]. Therefore, we expect that Alfvén modes will follow the same scaling relation as in the incompressible case. Figure 10b shows that the coupling gets weaker as B_0 increases:

$$\frac{(\delta V)_f^2}{(\delta V)_A^2} \propto \frac{(\delta V)_A}{B_0}. \tag{29}$$

The ratio of $(\delta V)_s^2$ to $(\delta V)_A^2$ is proportional to $(\delta V)_A^2/B_0^2$.

This marginal coupling is in good agreement with a claim in GS95, as well as earlier numerical studies where the velocity was decomposed into a compressible component $\mathbf{v}_C$ and a solenoidal component $\mathbf{v}_S$. The compressible component is curl-free and parallel to the wave vector $\mathbf{k}$ in Fourier space. The solenoidal component is divergence-free and perpendicular to $\mathbf{k}$. The ratio $\chi = (\delta V)_C/(\delta V)_S$ is an important parameter that determines the strength of any shock [133,141]. Porter, Woodward, & Pouquet [138] performed a hydrodynamic simulation of decaying turbulence with an initial sonic Mach number of unity and found that χ^2 evolves toward ~ 0.11. Matthaeus et al. [105] carried out simulations of decaying weakly compressible MHD turbulence [183] and found that $\chi^2 \sim O(M_s^2)$, where M_s is the sonic Mach number. In [10] a weak generation of compressible components in solenoidally driven super-Alfvénic supersonic turbulence simulations was obtained.

Figure 11 shows that the spectrum and the anisotropy of Alfvén waves in this limit are compatible with the GS95 model:

$$\textit{Spectrum of Alfvén Modes:} \qquad E(k) \propto k_\perp^{-5/3}, \tag{30}$$

[16] As correctly pointed out by Zweibel (this volume) there is always residual coupling between Alfvén and compressible modes due to steepening of Alfvén modes. However, this steepening happens on time-scales much longer than the cascading time-scale.

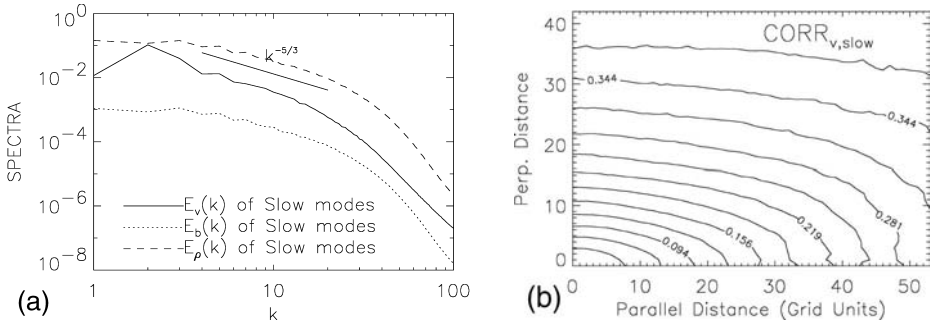


Fig. 12. (a) Slow spectra also follow a Kolmogorov-like power law. (b) Slow modes show anisotropy similar to the GS95 theory. From [21].

and scale-dependent anisotropy $k_\parallel \propto k_\perp^{2/3}$ that is compatible with the GS95 theory.

6.4 Scaling of the Slow Modes

Slow waves are somewhat similar to pseudo-Alfvén waves (in the incompressible limit). First, the directions of displacement (i.e. ξ_s) of both waves are similar when anisotropy is present. The vector ξ_s is always between $\hat{\theta}$ and $\hat{\mathbf{k}}_\parallel$. In Fig. 7, we can see that the angle between $\hat{\theta}$ and $\hat{\mathbf{k}}_\parallel$ gets smaller when $k_\parallel \ll k_\perp$. Therefore, when there is anisotropy (i.e. $k_\parallel \ll k_\perp$), $\hat{\xi}_s$ of a low β plasma becomes similar to that of a high β plasma. Second, the angular dependence in the dispersion relation $c_s \approx a\cos\theta$ is identical to that of pseudo-Alfvén waves (the only difference is that, in slow waves, the sound speed a is present instead of the Alfvén speed V_A).

Goldreich & Sridhar [46] argued that the pseudo-Alfvén waves are slaved to the shear-Alfvén (i.e. ordinary Alfvén) waves in the presence of a strong $\mathbf{B}_0$, meaning that the energy cascade of pseudo-Alfvén modes is primarily mediated by the shear-Alfvén waves. This is because pseudo-Alfvén waves do not provide efficient shearing motions. Similar arguments are applicable to slow waves in a low β plasma [21] (see also [96] for high-β plasmas). As a result, we conjecture that slow modes follow a scaling similar to the GS95 model [21]:

$$\textit{Spectrum of Slow Modes:} \qquad E^s(k) \propto k_\perp^{-5/3}. \tag{31}$$

Figure 12a shows the spectra of slow modes. For velocity, the slope is close to $-5/3$. Figure 12b shows the contours of equal second-order structure function (SF_2) of slow velocity, which are compatible with $k_\parallel \propto k_\perp^{2/3}$ scaling.

In low β plasmas, density fluctuations are dominated by slow waves [21]. From the continuity equation $\dot{\rho} = \rho\nabla\cdot\mathbf{v}$

$$\omega\rho_k = \rho_0\mathbf{k}\cdot\mathbf{v}_k, \tag{32}$$

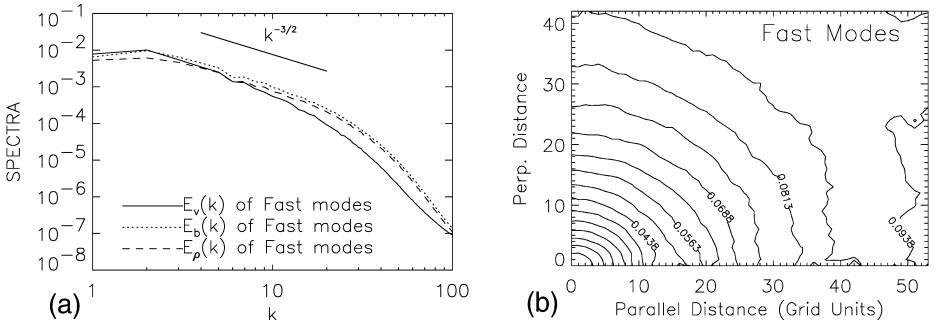


Fig. 13. (a) The power spectrum of fast waves is compatible with the IK spectrum. (b) The magnetic second-order structure function of fast modes shows isotropy.From [21].

we have, for slow modes, $(\rho_k)_s \sim \rho_0 (v_k)_s / a$. Hence, this simple argument implies

$$\left(\frac{\delta\rho}{\rho}\right)_s = \frac{(\delta V)_s}{a} \sim M_s, \tag{33}$$

where we assume that $(\delta V)_s \sim (\delta V)_A$ and M_s is the Mach number. On the other hand, only a small amount of magnetic field is produced by the slow waves. Similarly, using the induction equation ($\omega \mathbf{b}_k = \mathbf{k} \times (\mathbf{B}_0 \times \mathbf{v}_k)$), we have

$$\frac{(\delta B)_s}{(\delta V)_s} \sim \frac{a}{B_0} = O(\sqrt{\beta}), \tag{34}$$

which means that equipartition between kinetic and magnetic energy is not guaranteed in low β plasmas. In fact, in Fig. 12a, the power spectrum for density fluctuations has a much larger amplitude than the magnetic field power spectrum. Since density fluctuations are caused mostly by the slow waves and magnetic fluctuation is caused mostly by Alfvén and fast modes, we *do not* expect a strong correlation between density and magnetic field, which agrees with the ISM simulations [130,126,170].

6.5 Scaling of the Fast Modes

Figure 13 shows fast modes are isotropic. The resonance conditions for interacting fast waves are:

$$\omega_1 + \omega_2 = \omega_3, \tag{35}$$

$$\mathbf{k}_1 + \mathbf{k}_2 = \mathbf{k}_3. \tag{36}$$

Since $\omega \propto k$ for the fast modes, the resonance conditions can be met only when all three $\mathbf{k}$ vectors are colinear. This means that the direction of energy cascade is *radial* in Fourier space, and we expect an isotropic distribution of energy in Fourier space.

Using the constancy of energy cascade and uncertainty principle, we can derive an IK-like energy spectrum for fast waves. The constancy of cascade rate reads

$$\frac{v_l^2}{t_{cas}} = \frac{k^3 v_k^2}{t_{cas}} = \text{ constant.} \tag{37}$$

On the other hand, t_{cas} can be estimated as

$$t_{cas} \sim \frac{v_k}{(\mathbf{v}\cdot\nabla\mathbf{v})_\mathbf{k}} \sim \frac{v_k}{\sum_{\mathbf{p}+\mathbf{q}=\mathbf{k}} k v_p v_q}. \tag{38}$$

If contributions are random, the denominator can be written by the square root of the number of interactions ($\sqrt{\mathcal{N}}$) times strength of individual interactions ($\sim k v_k^2$) [17]. Here we assume locality of interactions: $p \sim q \sim k$. Due to the uncertainly principle, the number of interactions becomes $\mathcal{N} \sim k(\Delta k)^2$, where Δk is the typical transversal (i.e. not radial) separation between two wave vectors $\mathbf{p}$ and $\mathbf{q}$ (with $\mathbf{p}+\mathbf{q}=\mathbf{k}$). Therefore, the denominator of (38) is $(k(\Delta k)^2)^{1/2} k v_k^2$. We obtain an independent expression for t_{cas} from the uncertainty principle ($t_{cas}\Delta\omega \sim 1$ with $\Delta\omega \sim \Delta k(\Delta k/k)$). From this and (38), we get $t_{cas} \sim t_{cas}^{1/2}/(k^2 v_k)$, which yields

$$t_{cas} \sim 1/k^4 v_k^2. \tag{39}$$

Combining (37) and (39), we obtain $v_k^2 \sim k^{-7/2}$, or $E^f(k) \sim k^2 v_k^2 \sim k^{-3/2}$. This is very similar to acoustic turbulence, turbulence caused by interacting sound waves [180,181,97]. Zakharov & Sagdeev [181] found $E(k) \propto k^{-3/2}$. However, there is debate about the exact scaling of acoustic turbulence. Here we cautiously claim that our numerical results are compatible with the Zakharov & Sagdeev scaling:

$$\textit{Spectrum of Fast Modes:} \qquad E^f(k) \sim k^{-3/2}. \tag{40}$$

Magnetic field perturbations are mostly affected by fast modes [21] when β is small:

$$\textit{Fast:} \qquad \frac{(\delta B)_f}{(\delta V)_A} \sim \frac{(\delta V)_f}{(\delta V)_A}, \tag{41}$$

if $(\delta V)_A \sim (\delta V)_s$.

The turbulent cascade of fast modes is expected to be slow and in the absence of collisionless damping they are expected to propagate in turbulent media over distances considerably larger than Alfvén or slow modes. This effect is difficult to observe in numerical simulations with $\Delta B \sim B_0$. A modification of the spectrum in the presence of the collisionless damping is presented in [178].

[17] To be exact, the strength of individual interactions is $\sim k v_k^2 \sin\theta$, where θ is the angle between $\mathbf{k}$ and $\mathbf{B}_0$. Thus marginal anisotropy is expected. It will be investigated elsewhere.

7 Astrophysical Implications

Many astrophysical problems require some knowledge of the scaling properties of turbulence. Therefore we expect a wide range of applications of the established scaling relations. Here we show how recent breakthroughs in understanding MHD turbulence affect a few selected issues.

7.1 Cosmic Ray Propagation

The propagation of cosmic rays is mainly determined by their interactions with electromagnetic fluctuations in interstellar medium. The resonant interaction of cosmic ray particles with MHD turbulence has been repeatedly suggested as the main mechanism for scattering and isotropizing cosmic rays. In these analysis it is usually assumed that the turbulence is *isotropic* with a Kolmogorov spectrum (see [151]). How should these calculations be modified?

The essence of the mechanism is rather simple. Particles moving with velocity v interact with a resonant Alfvén wave of frequency $\omega = k_\parallel v\mu + n\Omega$ ($n = \pm 1, 2...$), where $\Omega = \Omega_0/\gamma$ is the gyrofrequency of relativistic particles, μ is the cosine of the pitch angle. From the resonant condition above, we know that the most important interaction occurs at $k_\parallel \sim \Omega/v\mu \sim (\mu r_L)^{-1}$, where r_L is Larmor radius of the high-energy particles.

The calculations in [178] that made use of tensor (14) provided the scattering efficiency of anisotropic Alfvénic turbulence. The results are compared in Fig. 14a with the predictions of the scattering on isotropic Kolmogorov-type magnetic fluctuations and also with earlier calculations by Chandran [17]. The latter used rather *ad hoc* form of the tensor to describe magnetic fluctuations within the Goldreich-Sridhar theory. We see from Fig. 14a that the scattering is substantially suppressed, compared to the Kolmogorov turbulence that is usually used for scattering calculations. This happens, first of all, because most turbulent energy in GS95 turbulence goes to $k_\perp$ so that there is much less energy left in the resonance point $k_\parallel = (\mu r_L)^{-1}$. Furthermore, $k_\perp \gg k_\parallel$ means $k_\perp \gg r_L^{-1}$ so that cosmic ray particles cover many eddies during one gyration. This random walk decreases the scattering efficiency by a factor of $(\Omega/k_\perp v_\perp)^{\frac{1}{2}} = (r_L/l_\perp)^{\frac{1}{2}}$, where $l_\perp$ is the turbulence scale perpendicular to magnetic field.

Thus the gyroresonance with Alfvénic turbulence is not an effective scattering mechanism for cosmic rays if turbulence is injected on the large scales, since the degree of anisotropy increases on smaller scales. However, if energy is injected isotropically at small scales, the resulting turbulence would be more isotropic and scattering will be more efficient. Scattering by undamped fast modes is more efficient than the Kolmogorov theory would predict. Yan & Lazarian [178] performed calculations taking into account the collisionless damping of fast modes and showed that the gyroresonance scattering by fast modes is the dominant scattering mechanism.

There is another important property of turbulence that was neglected in earlier work. When cosmic rays stream at a velocity much larger than Alfvén velocity, they can excite resonant MHD waves, which in turn scatter cosmic rays.

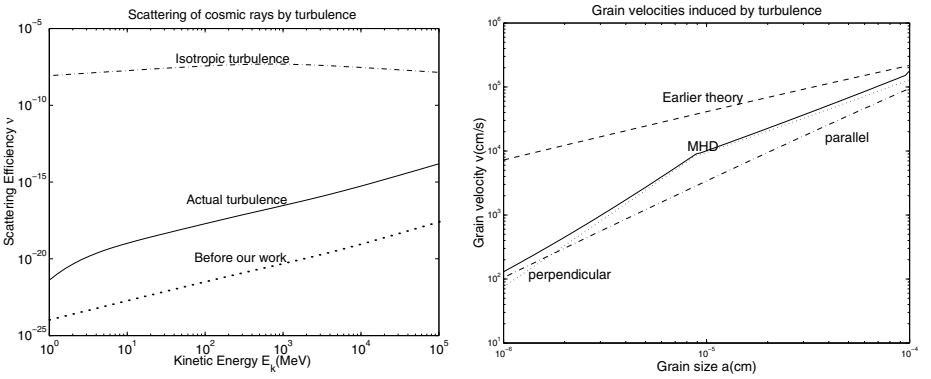


Fig. 14. Applications. (By Lazarian & Yan). *Left:* Cosmic ray scattering for realistic MHD turbulence is reduced substantially compared to scattering by isotropic turbulence, but still larger than estimates in Chandran [17]. *Right:* The dust acceleration by turbulence is reduced compared to the accepted estimates in Draine [33].

This is the 'streaming instability'. It is usually assumed that this instability can confine cosmic rays with energies less than 100GeV [16]. However, this is true only in an idealized situation when there is no background MHD turbulence. As noted earlier, the rates of turbulent decay are very fast and excited perturbations should vanish quickly. In [178] we find that the streaming instability is only applicable to particles with energies $< 0.15GeV$, which is less than the energy of most cosmic ray particles. This result casts doubt on the self-confinement mechanism discussed by previous authors.

All these findings tend to support the alternative picture of cosmic ray diffusion advocated by Jokipii (see [69]). In this picture cosmic rays follow magnetic field lines, but the magnetic field wanders. The rate of this wandering can be calculated from the established turbulence scaling laws.

Knowledge of the scattering rates is essential for understanding both the first order and the second order Fermi acceleration. The first order Fermi acceleration, may be important for a wide range of phenomena from clusters of galaxies and gamma-ray bursts (GRBs) to solar flares. Results obtained in [178] where the discovered properties of MHD turbulence are used proved to be very different from earlier estimates.

7.2 Grain Dynamics

Turbulence induces relative dust grain motions and leads to grain-grain collisions. These collisions determine grain size distribution, which affects most dust properties, including absorption and H_2 formation. Unfortunately, as in the case of cosmic rays, earlier work appealed to hydrodynamic turbulence to predict grain relative velocities (see [74,166,33,124,167,38]).

The differences between the hydrodynamic and MHD calculations stem from (a) grain charges, which couple grains to the magnetic field, (b) the anisotropy

of MHD cascade, and (c) the direct interaction of charged grains with magnetic perturbations. Effects (a) and (b) are considered in Lazarian & Yan [92], while (c) is considered in Yan & Lazarian (2002b; in preparation). As consequence the picture of grain dynamics is substantially altered.

Consider grain charge first. If a grain's Larmor period $\tau_L = 2\pi m_{gr}c/qB$ is shorter than the gas drag time t_{drag}, grain perpendicular motions are constrained by magnetic field. Their velocity dispersion is determined by the turbulence eddy whose turnover period is $\sim \tau_L$ instead of the drag time [33].

Accounting for the anisotropy of MHD turbulence it is convenient to consider separately grain motions parallel and perpendicular to magnetic field. The perpendicular motion is influenced by the Alfvén modes, which have a Kolmogorov spectrum. The parallel motion is subjected to compressible modes which scale as $v_\parallel \propto k_\parallel^{-1/2}$. In addition we should account for viscous forces. When the eddy turnover time is of the order of $t_{damp} \sim \nu_n^{-1} k_\perp^{-2}$, the turbulence is viscously damped. Thus grains sample only a part of the eddy before gaining the velocity of the ambient gas if τ_L or $t_{drag} < t_{damp}$. The results are shown in Fig. 14b.

The direct interaction of the charged grains with turbulent magnetic field results in a stochastic acceleration that can potentially provide grains with supersonic velocities.

7.3 Turbulence in HII Regions

Lithwick & Goldreich [96] addressed the issue of the origin of density fluctuations within HII regions. There the gas pressure is larger than the magnetic pressure (the 'high beta' regime) and they conjectured that fast waves, which are essentially sound waves, would be decoupled from the rest of the cascade. They found that density fluctuations are due to the slow mode and the entropy mode, which are passively mixed by shear Alfvén waves and follow a Kolmogorov spectrum. They also found that slow mode density fluctuations are proportional to $1/\sqrt{\beta}$. On the other hand, the entropy mode density fluctuations are suppressed when cooling is faster than the cascade time. Lithwick & Goldreich [96] also gave detailed discussions about density fluctuations on various scales in the ISM, e.g. proton gyro-radius. These results are important as radio-wave scintillation observations can constrain the nature of MHD turbulence in the ISM, especially in the HII regions. Lithwick & Goldreich [96] argued that the turbulent cascade survives ion-neutral damping only when a high degree of ionization is present. However, the study by Cho, Lazarian & Vishniac [24] suggests that the magnetic fluctuations protrude below the damping scale and the results of [96] should be revised.

7.4 Tiny-Scale Atomic Structures

The intermittent small-scale structures in Sect. 5.2 should have important implications for transport processes (heat, cosmic rays, etc.) in partially ionized plasmas. We also speculate that they might have some relation to the tiny-scale atomic structures (TSAS). Heiles [54] introduced the term TSAS for the

mysterious H I absorbing structures on scales from thousands to tens of AU, discovered by Dieter, Welch & Romney [32]. Analogs are observed in NaI and CaII [108,36,1] and in molecular gas [102]. Recently Deshpande, Dwarakanath & Goss [28] analyzed channel maps of opacity fluctuations toward Cas A and Cygnus A. They found that the amplitudes of density fluctuations at scales less than 0.1 pc are far larger than expected from extrapolation from larger scales, possibly explaining TSAS. This study, however, cannot answer what confines those presumably overpressured (but very quiescent!) blobs of gas. Deshpande [27] related those structures to the shallow spectrum of interstellar turbulence.

Figure 5b indicates that while velocity decreases rapidly, but *not* exponentially, below the viscous damping scale, the magnetic field fluctuations persist, thereby providing nonthermal pressure support. Magnetic structures perpendicular to the mean magnetic field are compensated by pressure gradients. Our calculations so far are produced using incompressible code [24]. In the case of compressible media, we expect the pressure fluctuations to entail density fluctuations reminiscent of the Deshpande et al. [28] observations.

The calculations in Cho, Lazarian & Vishniac [24] are applicable on scales from the viscous damping scale (determined by equating the energy transfer rate with the viscous damping rate; ~ 0.1 pc in the Warm Neutral Medium with $n = 0.4$ cm^{-3}, T= 6000 K) to the ion-neutral decoupling scale (the scale at which viscous drag on ions becomes comparable to the neutral drag; $\ll 0.1$ pc). Below the viscous scale the fluctuations of magnetic field obey the damped regime shown in Fig. 5b and produce density fluctuations. For typical Cold Neutral Medium gas, the scale of neutral-ion decoupling decreases to $\sim$ 70AU, and is less for denser gas. TSAS may be created by strongly nonlinear MHD turbulence!

A simple technique of estimating magnetic field was suggested by Chandrasekhar & Fermi [18] (see also review by Ostriker, this volume). According to it, the fluctuations of magnetic field that can be measured from polarization maps are related to velocity fluctuations measured through Doppler broadening $\delta b/\sqrt{4\pi\rho} \sim \delta v$. The existence of the damped regime of MHD turbulence suggests that this technique is not applicable to very small scales in partially ionized gas.

7.5 Magnetic Reconnection

Magnetic reconnection is the fundamental process that allows magnetic fields to change their topology, despite being 'frozen' into highly conducting plasmas. It is the key process for solar flares, the magnetic dynamo, the acceleration of energetic particles, etc. According to the Lazarian & Vishniac model [89] (see also review [90] and Vishniac, Lazarian, & Cho, this volume) of stochastic reconnection, this process is controlled by the turbulent wandering of magnetic field. The exact properties of the turbulent cascade are especially important for the viscously damped regime present in partially ionized gas. However, it is shown in Lazarian, Vishniac & Cho [91] that the reconnection rates are sufficiently high in this case. The implications of the finding for the removal of magnetic flux during star formation is to be evaluted yet.

7.6 Support and Compression of Molecular Clouds

To understand the dynamics of molecular clouds and star formation it is necessary to understand turbulence. In a recent review [106] McKee pointed out that the fast damping of MHD turbulence observed in numerical simulations is difficult to reconcile with the fact that "a GMC such as G216, which has no visible star formation, can have a level of turbulence that exceeds that in the Rosette molecular cloud, which has an embedded OB association". He pointed out that the conclusions obtained on the basis of numerics should be treated with caution as they do not resolve the microscales.

In typical astrophysical conditions the sources of turbulence are localized both in space (stars) and time (stellar outflows; supernovae), and the outgoing waves have much larger amplitudes than the background waves (we call this situation "imbalanced cascade"). Figure 5a shows that the turbulent damping could be substantially reduced in this situation. Moreover, even in a balanced regime, we expect fast modes to be subjected to slow non-linear damping.

At the same time, it is worth mentioning, that turbulence can not only support, but also compress molecular clouds. Clouds can be compressed by external turbulence feeding into them and depositing energy and momentum. Myers & Lazarian [120] explained observed infalling motions of molecular gas surrounding dense cores [164,94] in this way, based on ion-neutral damping. The infall rate is proportional to the rate of turbulence damping. Therefore, fast non-linear damping associated with the Alfvénic turbulence should enhance the infall.

7.7 Heating of Diffuse Ionized Gas

The "Diffuse Ionized Gas" (DIG), or equivalently the "Reynolds layer" within the Milky Way, is detected by rather faint but ubiquitous Galactic Hα emission [144,145]. Such emission is found in several other spirals as well [142,143,127,128]. In the Galaxy, the Reynolds layer contains a substantial portion of the H^+ in the ISM. Current models generally involve photoionization from the OB stars, although how the Lyman continuum radiation from OB stars can penetrate the neutral H layer remains controversial.

The observations show strong [S II] λ6717 and [N II] λ6583 that increase relative to Hα with distance z above the planes of various galaxies, including the Milky Way [145]. The only reasonable conclusion is that there is an additional source of heating in the ISM that dominates over photoionization heating at low densities. It has been proposed that carbonaceous molecules provide the excess heating through the photoelectric effect [177], but this explanation is not unique. Heating by turbulence, surely present, may dominate. Minter & Spangler [110] suggested a heating rate that is adequate to explain the [S II]/Hα and [N II]/Hα ratios, but did not take nonlinear interactions into account, thereby underestimating the heating. A new study that would capitalize on the new understanding of MHD turbulence (damping of Alfvén and fast modes, imbalance etc.) is on our agenda.

8 Observational Tests

Comparing numerics with observations is a challenging problem. Ostriker (this volume) discusses PDF, clump identification, and linewidth-scale relations as possible diagnostics and outlines problems with any of those approaches. A use of spectral line data cubes and application to it of different techniques (e.g. spectral correlation function, principal component analysis, wavelets, etc.) can be found in the review [79]. Here we shall concentrate on comparing spectra from observations with our theoretical expectations (see also review [86]).

8.1 Is the Big Power Law Real?

We have mentioned above that observations suggest that the Kolmogorov power law should span from AU to kpc scales. Kolmogorov scaling is exactly what one would expect from the GS95 picture when the observations sample magnetic field in the system of reference aligned with the *mean magnetic field* (see Fig. 15 for examples of observable quantities). Indeed, it is obvious from Fig. 1 that in this system of reference (i.e. global system of reference) the locally defined scalings of $k_\parallel$ with $k_\perp$ are not valid. Indeed, one can easily see that the fluctuations perpendicular to the *local* direction of magnetic field dominate both the statistics measured perpendicular to the *mean* magnetic field and parallel to it. As a result, in the reference system aligned with the *mean* magnetic field $k'_\parallel \propto k'_\perp$ (see Fig. 15a,b,c) and according to equation (11) $E(k) \sim k^{-5/3}$ will be measured.

Ambiguities in measurements reviewed in [2], however, make it uncertain whether or not the Big Power Law should be taken at face value. Still, we note that agreement between an observed power spectrum and theoretical expectations is far more significant than just an approximate fit between observations and numerics. The latter is definitely not unique and is *a priori* suspect in view of the huge difference in terms of Re and Rm between any numerical simulation and the ISM.

To test the Big Power Law properly, it is important to extend the theory of Velocity Channel Analysis (VCA) [84,86] by including self-absorption in the analysis of turbulent emission lines. It can then be applied to regions of HI in the inner Galaxy [30] or CO [163,162][18]. Lazarian & Pogosyan [85] provided results consistent with observations. For instance, the study predicts that in the presence of absorption the emitted power in the line is proportional to k^{-3}, exactly what is seen in Dickey et al. [30] for HI in the inner region of the Galactic plane.

An application of the VCA to different emission lines (e.g. Hα, [N II], [S II] etc.) would help to answer the question of whether or not ISM turbulence is

[18] Brunt & Heyer [14] applied Principal Component Analysis (see [55]), to simulated CO data and found empirical relations between the statistics of velocities, eigenvectors and eigenimages. However, they note that somehow their relation does not depend on the absorption coefficient for the limited range of absorptions they tested. Unfortunately, their analysis does not seem to be applicable to Galactic HI and the applicability of their technique to the correlated velocity and density fields is unclear.

a large-scale cascade with various phases of the ISM interconnected through a dynamically important magnetic field. A contrasting possibility is that various phases form their own cascades.

HI is rather smoothly distributed across the sky. Therefore, the effects of image edges does not pose a problem for the statistical analysis. Results by Stanimirovic (private communication) show that the Fourier analysis of the SMC image and a more laborious wavelet analysis[19] provide identical results. However, when dealing with molecular clouds we might expect that the cloud edges become important. Therefore the incorporation of wavelets (see [184]) within the VCA is a natural step to make. The only difference would be to apply the wavelets instead of Fourier transforms to the channel maps.

Synchrotron fluctuations and fluctuations of polarized radiation arising from aligned dust should be used to study magnetic field statistics. As we mentioned earlier Cho & Lazarian [22] has shown that those fluctuations are consistent with Kolmogorov scaling. More studies in this direction are necessary. The fact that those fluctuctuations interfere with the CMB studies garantees that in the near future we shall have a lot of relevant data.

One should remember, however, that the measured power index of fluctuations may not correspond to the spectral index of the underlying turbulence. For instance, it is shown in [159] that while the actual turbulence in SMC is close to being Kolmogorov, depending on the thickness of the slice, the spectral index of intensity fluctuations within channel maps span from ~ -2.8 to ~ -3.4. Similarly, it is shown in [22] that for Kolmogorov turbulence the spectral index of observable fluctuations may vary from ~ -1 to -3.7 depending on how observations sample turbulence.

8.2 Does Turbulence Reveal Magnetic Field Direction?

Anisotropy of Alfvénic turbulence is a definite prediction of the GS95 theory. We mentioned earlier that the scale-dependent anisotropy can only be revealed in the local frame of reference, which in practical terms require direct measurements, e.g. with spacecraft. Measurements of the Solar wind magnetic turbulence have failed so far to reveal the differential scaling of the turbulence in terms of $k_{\|}$ and $k_{\perp}$, but these measurements are inconclusive [57]. If measurements are performed in a global system of reference, as is the case with observations, they should reveal anisotropy in the direction of the mean magnetic field.

In isotropic turbulence, correlations depend only on the distance between sampling points. Contours of equal correlation are circular in this case. The presence of a magnetic field introduces anisotropy and these contours become elongated with a symmetry axis given by the magnetic field. To study turbulence anisotropy, we can measure contours of equal correlation corresponding to the data within various velocity bins. The results obtained with simulated data are shown in Fig. 15.

[19] Wavelet analysis involves determining the deviation of each pixel from a weighted average of the pixels at a particular projected distance from it.

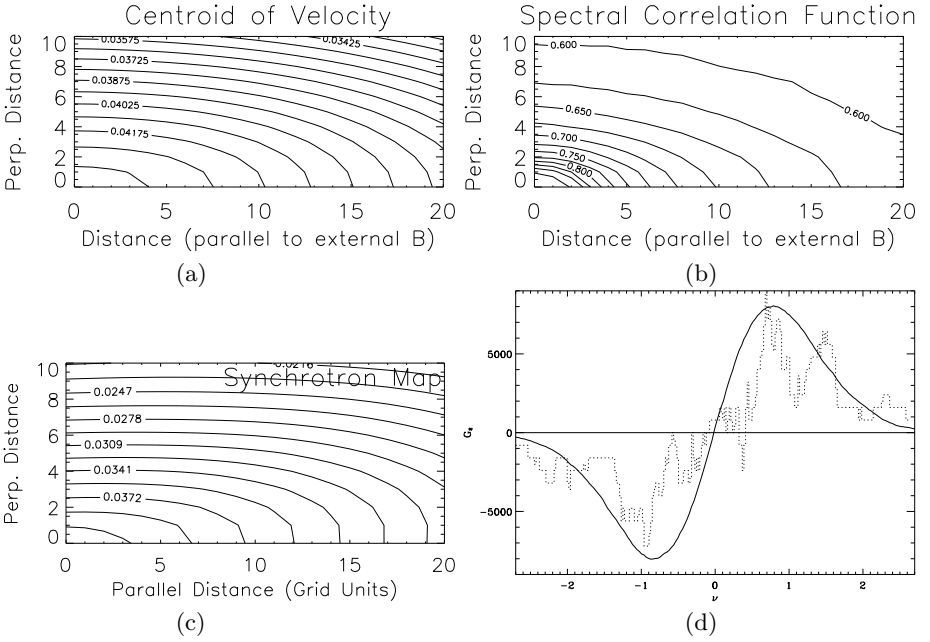


Fig. 15. *Observational test of the synthetic data by Cho, Esquivel & Lazarian.* Contours of equal correlation obtained with Centroids of Velocity (*(a)*) and with Spectral Correlation Function (SCF)(*(b)*). The direction of anisotropy reveals the direction of projected magnetic field. Combined with the anisotropy analysis, the SCF (introduced by Alyssa Goodman) is likely to become even more useful tool. *(c)* Contours of equal correlation obtained for synthetic synchrotron intensity map. *(d)* shows the 2D genus of the Gaussian distribution (smooth analytical curve) against the genus for the isothermal compressible MHD simulations with Mach number ∼2.5 (dotted curve).

Since the degree of anisotropy is related to the strength of the magnetic field, studies of anisotropy can provide the means to analyze magnetic fields. It is important to study different data sets and channel maps for the anisotropy. Optical and infrared polarimetry can benchmark the anisotropies in correlation functions. We hope that anisotropies will reveal magnetic field structure within dark clouds where grain alignment and therefore polarimetry fails (see [80] for a review of grain alignment).

Not only velocity statistics can be used for such an analysis. Lazarian & Shutenkov [83] (see review [76]) showed that the mean magnetic field must lead to anisotropies in the synchrotron statistics. Lazarian & Chibisov [81] pointed out that using HI regions as screens for radiation at the decameter wavelength it should be possible to study the 3D distribution of the magnetic field. Figure 15c shows the anisotropy of synchrotron statistics available through simulations.

8.3 How Else Can We Compare Observations and Simulations?

Velocity and density power spectra do not provide a complete description of turbulence. Intermittency (variations in the strength of the turbulent cascade) and its topology in the presence of different phases are not described by the power spectrum. Use of the higher moments is possible (see discussion of the 3 point statistics, or bispectrum, in Lazarian [79]), but is limited by the noise in the observational data.

"Genus analysis" is a good tool for studying the topology of turbulence (see the review [86]). This tool has already been successfully applied to cosmology [51]. Consider an area on the sky with contours of projected density. The 2D genus, $G(\nu)$, is the difference between the number of regions with a projected density higher than ν and those with densities lower than ν. Figure 15d shows the 2D genus as the function of ν for a Gaussian distribution of densities (completely symmetric curve), for MHD isothermal simulations with Mach number $\sim$2.5. It is shown in [86] that the genus of the Small Magellanic Cloud is very different from that in Fig. 15d, while the spectra in both cases are similar.

9 Summary

Recently there have been significant advances in the field of MHD turbulence:

1. The first self-consistent model (GS95) of incompressible MHD turbulence that is supported by both numerical simulations and observations is now available. The major predictions of the model are scale-dependent anisotropy ($k_\parallel \propto k_\perp^{2/3}$) and a Kolmogorov energy spectrum ($E(k) \propto k^{-5/3}$).
2. There have been substantial advances in understanding compressible MHD. Simulations of compressible MHD turbulence show that there is a weak coupling between Alfvén waves and compressible MHD waves and that the Alfvén modes follow the Goldreich-Sridhar scaling. Fast modes, however, decouple and exhibit isotropy.
3. Contrary to general belief, in typical interstellar conditions, magnetic fields can have rich structures below the scale at which motions are damped by the viscosity created by neutral drag (the ambipolar diffusion damping scale).
4. These advances will have a dramatic impact on our understanding of many fundamental interstellar processes, like cosmic-ray propagation, grain dynamics, turbulent heating and molecular cloud stability.
5. New techniques, e.g. VCA, allow observational tests of the theory.

Acknowledgments

We thank Peter Goldreich, John Mathis, Steven Shore, Enrique Vazquez-Semadeni, and Huirong Yan for helpful discussions. AL and JC acknowledge the support of the NSF through grant AST-0125544. ETV acknowledges NSF grant AST-0098615. This work was partially supported by National Computational Science Alliance under AST000010N and AST010011N and utilized the NCSA SGI/CRAY Origin2000. AL thanks the LOC for their financial support. The authors thank the editors for their patience with the manuscript.

References

1. S. M. Andrews, D. M. Meyer, J. T. Lauroesch: Astrophys. J. **552**, L73 (2001)
2. J. W. Armstrong, B. J. Rickett, S. R. Spangler: Astrophys. J. **443**, 209 (1995)
3. C. Baccigalupi, C. Burigana, F. Perrotta, G. De Zotti, L. La Porta, D. Maino, M. Maris, R. Paladini: Astron. Astrophys. **372** 8 (2001)
4. A. Bhattacharjee, C. S. Ng, S. R. Spangler: Astrophys. J. **494**, 409 (1998)
5. D. Biskamp: *Nonlinear Magnetohydrodynamics* (Cambridge Uinversity Press, Cambridge, 1993)
6. D. Biskamp: Phys. Plasmas, **9**(4), 1486 (2002)
7. D. Biskamp, E. Schwarz: Phys. Plasmas, **8**(7), 3282 (2001)
8. D. Biskamp, E. Schwarz, A. Celani: Phys. Rev. Lett. **81** 4855 (1998)
9. S. Boldyrev: Astrophys. J. **569**, 841 (2002)
10. S. Boldyrev, A. Nordlund, P. Padoan Astrophys. J. **573**, 678 (2002)
11. A. Brandenburg: Astrophys. J. **550**, 824 (2001)
12. A. Brandenburg, R. L. Jennings, A. Nordlund: J. Fluid Mech. **306**, 325 (1996)
13. M. Brio, C. Wu: J. Comput. Phys. **75**, 500 (1988)
14. C. Brunt, M. Heyer: Astrophys. J. **566**, 276 (2002)
15. F. Cattaneo, D. W. Hughes: Phys. Rev. E. **54**, R4532 (1996)
16. C. Cesarsky: Annu. Rev. Astro. Astrophys. **18**, 289 (1980)
17. B. Chandran: Phys. Rev. Lett. **85**(22), 4656 (2001)
18. S. Chandrasekhar, E. Fermi: Astrophys. J. **118**, 113 (1953)
19. J. Cho, E. T, Vishniac: Astrophys. J. **538**, 217 (2000a)
20. J. Cho, E. T. Vishniac: Astrophys. J. **539**, 273 (2000b)
21. J. Cho, A. Lazarian: Phys. Rev. Lett. accepted (2002) astro-ph/0205282
22. J. Cho, A. Lazarian: Astrophys. J.,**575**, L63 (2002) astro-ph/0205284
23. J. Cho, A. Lazarian, E. T. Vishniac: Astrophys. J. **564**, 291 (2002a) (CLV02a)
24. J. Cho, A. Lazarian, E. T. Vishniac: Astrophys. J. **566**, L49 (2002b)
25. L. Del Zanna, M. Velli, P. Londrillo: Astron. Astrophys. **367**, 705, (2001)
26. N. F. Derby: Astrophys. J. **224**, 1013, (1978)
27. A. A. Deshpande: Monthly Not. Roy. Astron. Soc. **317**, 199 (2000)
28. A. A. Deshpande, K. S. Dwarakanath, W. M. Goss: Astrophys. J. **543**, 227 (2000)
29. P. H. Diamond, G. G. Craddock: Comments on the Plasma Physics of Controlled Fusion, **13**, 287 (1990)
30. J. M. Dickey, N. M. McClure-Griffiths, S. Stanimirovic, B. M. Gaensler, A. J. Green: Astrophys. J. **561**, 264 (2001)
31. R. L. Dickman: in *Protostars and Planets II*, ed. by D. C. Black, M. S. Mathews (Tucson: Univ. Arizona Press, 1985) p.150
32. N. H. Dieter, W. J. Welch, J. D. Romney: Astrophys. J. **206**, L113 (1976)
33. B. T. Draine: 1985, in *Protostars and Planets II*, ed. by D. C. Black, M. S. Matthews (Tucson: Univ. Arizona Press, 1985) p. 621
34. B. T. Draine, Lazarian, A.: Astrophys. J. **512**, 740 (1999)
35. B. G. Elmegreen: in *From Darkness to Light*, ed. by T. Montmerle, P. Andre (ASP Conf. Series, 2002, p. 255) (astro-ph:0010582)
36. M. D. Faison, W. M. Goss: Astron. J. **121**, 2706 (2001)
37. E. Falgarone, G. Pineau des Forets, E. Roueff: Astron. Astrophys. **300**, 870 (1995)
38. E. Falgarone, J. L. Puget, Astron. Astrophys. **293**, 840 (1995)
39. J. Fournier, P. Sulem, A. Pouquet: J. Phys. A, **15**, 1393 (1982)
40. P. Fosalba, A. Lazarian, S. Prunet, J. A. Tauber: Astron. J. **564**, 762 (2002)

41. U. Frisch: *Turbulence: the legacy of A.N. Kolmogorov*, (Cambridge Univ. Press, New York, 1995)
42. S. Galtier, S. V. Nazarenko, A. C. Newell, A. Pouquet: J. Plasma Phys. **63**, 447 (2000)
43. G. Giardino, A.J. Banday, P. Fosalba, K.M. Górski, J.L. Jonas, W. O'Mullane, J. Tauber: Astron. Astrophys. **371** 708 (2001)
44. G. Giardino, A.J. Banday, K.M. Górski, K. Bennett, J.L. Jonas, J. Tauber: Astron. Astrophys. astro-ph/0202520 (2002)
45. P. Goldreich, S. Sridhar: Astrophys. J. **438**, 763 (1995) (GS95)
46. P. Goldreich, S. Sridhar: Astrophys. J. **485**, 680 (1997)
47. M. L. Goldstein: Astrophys. J. **219**, 700 (1978)
48. M. L. Goldstein, D. A. Roberts: Annu. Rev. Astron. Astrophys. **33**, 283, (1995)
49. J. Goodman, R. Narayan: Monthly Not. Roy. Astron. Soc. **214**, 519 (1985)
50. S. Ghosh, W. M. Matthaeus, D. C. Montgomery: Phys. Fluids, **31**, 2171 (1988)
51. J. R. Gott, C. Park, R. Juskiewicz, W. Bies, F. Bouchet, A. Stebbins: Astrophys. J. **352**, 1 (1990)
52. D. A. Green: Monthly Not. Roy. Astron. Soc. **262**, 328 (1993)
53. A. V. Gruzinov, P. H. Diamond: Phys. Rev. Lett. **72**, 1671 (1994)
54. C. Heiles: Astrophys. J. **481**, 193 (1997)
55. M. H. Heyer, F. P. Schloerb: Astrophys. J. **475**, 173 (1997)
56. J. C.Higdon: Astrophys. J. **285**, 109 (1984)
57. T. S. Horbury: in *Plasma Turbulence and Energetic particles*, ed. by M. Ostrowski, R. Schlickeiser (Cracow, Poland, 1999) p.28
58. T. S. Horbury, A. Balogh: Nonlin. Proc. Geophys. **4**, 185 (1997)
59. S. von Horner: Zs.F. Ap. **30**, 17 (1951)
60. M. Hossain, P. C. Gray, D. H. Pontius, W. H. Matthaeus: Phys. Fluids, **7**, 2886 (1995)
61. P. Iroshnikov: Astron. Zh. **40**, 742 (1963) (English: Sov. Astron. **7**, 566 (1964))
62. V. Jayanti, J. V. Hollweg: J. Geophys. Res. **98**, 13247 (1993)
63. K. Joulain, E. Falgarone, G. Pineau des Forets, D. Flower: Astro. Astrophys. **340**, 241 (1998)
64. J. Kampé de Fériet: in: *Gas Dynamics of Cosmic Clouds* (Amsterdam: North-Holland, 1955) p.134
65. S. A. Kaplan, S. B. Pickelner: *The Interstellar Medium* (Harvard Univ. Press, 1970)
66. L. Klein, R. Bruno, B. Bavassano, H. Rosenbauer: J. Geophys. Res. **98**, 17461 (1993)
67. R. S. Klessen: Astrophys. J. **556**, 837 (2001)
68. A. Kolmogorov: Dokl. Akad. Nauk SSSR, **31**, 538 (1941)
69. J. Kota, J. R. Jokipii: Astrophys. J. **531**, 1067 (2000)
70. R. Kraichnan: J. Fluid Mech. **5**, 497 (1959)
71. R. Kraichnan: Phys. Fluids **8**, 1385 (1965)
72. F. Krause, & K.H. Radler: *Mean-Field Magnetohydrodynamics and Dynamo Theory* (Oxford: Pergamon Press 1980)
73. R. M. Kulsrud, S. W. Anderson: Astrophys. J. **396**, 606 (1992)
74. T. Kusaka, T. Nakano, C. Hayashi: Prog. Theor. Phys.**44**, 1580 (1970)
75. R. B. Larson: Monthly Not. Roy. Astron. Soc. **194**, 809 (1981)
76. A. Lazarian: Astron. and Astrophys. Transactions, **3**, 33 (1992)
77. A. Lazarian: Astron. Astrophys. **293**, 507 (1995)

78. A. Lazarian: in *Interstellar Turbulence*, ed. by J. Franco, A. Carraminana (Cambridge Univ. Press, 1999a) p.95 (astro-ph/9804024)
79. A. Lazarian: in *Plasma Turbulence and Energetic Particles*, ed. by M. Ostrowski, R. Schlickeiser (Cracow, 1999b) p.28, (astro-ph/0001001)
80. A. Lazarian: 2000, in *Cosmic Evolution and Galaxy Formation*, ASP v.215, ed. by J. Franco, E. Terlevich, O. Lopez-Cruz, I. Aretxaga (Astron. Soc. Pacific,2000) p.69 (astro-ph/0003414)
81. A. Lazarian, G. Chibisov: Sov. Astron. Lett. **17**(3), 208 (1991)
82. A. Lazarian, A. Goodman, P. Myers: Astrophys. J. **490**, 273 (1997)
83. A. Lazarian, V. R. Shutenkov: Sov. Astron. Lett. **16**(4), 297 (1990)
84. A. Lazarian, D. Pogosyan: Astrophys. J. **537**, 720L (2000)
85. A. Lazarian, D. Pogosyan: (2002), in preparation
86. A. Lazarian, D. Pogosyan, A. Esquivel: 2002, in *Seeing Through the Dust*, ed. by R. Taylor, T. Landecker, A. Willis (ASP Conf. Series, 2002), in press (astro-ph/0112368) (LPE02)
87. A. Lazarian, D. Pogosyan, E. Vazquez-Semadeni, B. Pichardo: Astrophys. J. **555**, 130 (2001)
88. A. Lazarian, S. Prunet: in *Astrophysical Polarized Backgrounds*, ed. by S. Cecchini, S. Cortiglioni, R. Sault, C. Sbarra (2002) (astro-ph/0111214)
89. A. Lazarian, E. T. Vishniac: Astrophys. J. **517**, 700 (1999)
90. A. Lazarian, E. T. Vishniac: in Astrophysical Plasmas: Codes, Models, and Observations, eds. by J. Arthur, N. Brickhouse, & J. Franco, RevMexAA (Serie de Conferencias), Vol. **9**, 55
91. A. Lazarian, E. T. Vishniac, J. Cho: 2002, in preparation
92. A. Lazarian, H. Yan: Astrophys. J. **566**, L105 (2002)
93. R. J. Leamon, C. W. Smith, N. F. Ness, W. H. Matthaeus: J. Geophys. Res. **103**, 4775 (1998)
94. C. W. Lee, P. C. Myers, M. Tafalla: Astrophys. J. Supp. **136**, 703 (2001)
95. M. Lesieur: *Turbulence In Fluids* (Dordrecht: Kluwer, 1990)
96. Y. Lithwick, P. Goldreich: Astrophys. J. **562**, 279 (2001)
97. V. S. L'vov, Y. V. L'vov, A. Pomyalov: Phys. Rev. E, **61**, 2586 (2000)
98. M. Mac Low: Phys. Rev. Lett. **80**, 2754 (1998)
99. M. Mac Low: Astrophys. J. **524**, 169 (1999)
100. J. Maron, S. Cowley: astro-ph/0111008 (2001)
101. J. Maron, P. Goldreich: Astrophys. J. **554**, 1175 (2001)
102. A. P. Marscher, E. M. Moore, T. M. Bania: Astrophys. J. **419**, L101 (1993)
103. W. M. Matthaeus, M. L. Goldstein, D. C. Montgomery: Phys. Rev. Lett. **51**, 1484 (1983)
104. W. M. Matthaeus, S. Oughton, S. Ghosh, M. Hossain: Phy. Rev. Lett. **81**, 2056 (1998)
105. W. M. Matthaeus, S. Ghosh, S. Oughton, D. A. Roberts: J. Geophys. Res. **101**, 7619 (1996)
106. C. F. McKee: in *The Origin of Stars and Planetary Systems*, ed. by J. L. Charles, D. K. Nikolaos (Dordrecht: Kluwer, 1999) p.29
107. M. Meneguzzi, U. Frish, A. Pouquet: Phys. Rev. Lett. **47**, 1060 (1981)
108. D. M. Meyer, J. C. Blades: Astrophys. J. **464**, L179 (1996)
109. L. J. Milano, W. H. Matthaeus, P. Dmitruk, D. C. Montgomery: Phys. Plasmas, **8**(6), 2673 (2001)
110. A. Minter, S. Spangler: Astrophys. J. **485**, 182 (1997)

111. H. K. Moffatt: *Magnetic Field Generation in Electrically Conducting Fluids* (Cambridge: Cambridge Univ. Press, 1978)
112. A. S. Monin, A. A. Yaglom: *Statistical Fluid Mechanics: Mechanics of Turbulence*, Vol. 2 (Cambridge: MIT Press, 1975)
113. D. C. Montgomery: Physica Scripta, **T2/1**, 83 (1982)
114. D. C. Montgomery, W. H. Matthaeus: Astrophys. J. **447**, 706 (1995)
115. D. C. Montgomery, L. Turner: Phys. Fluids **24**(5), 825 (1981)
116. W.-C. Müller, D. Biskamp: Phys. Rev. Lett. **84**(3) 475 (2000)
117. G. Munch: Rev. Mod. Phys. **30**, 1035 (1958)
118. P. C. Myers: Astrophys. J. **270**, 105 (1983)
119. P. C. Myers: in *The Origin of Stars and Planetary Systems*, ed. by J. L. Charles, D. K. Nikolaos (Dordrecht: Kluwer, 1999) p.67
120. P. C. Myers, A. Lazarian: Astrophys. J. Lett. **507**, 157 (1998)
121. R. Narayan, J. Goodman: Monthly Not. Roy. Astron. Soc. **238**, 963 (1989)
122. C. S. Ng, A. Bhattacharjee: Astrophys. J. **465**, 845 (1996)
123. A. Nordlund, A. Brandenburg, R. Jennings, M. Rieutord, J. Ruokolainen, R. Stein, I. Tuominen: Astrophys. J. **392**, 647 (1992)
124. V. Ossenkopf: Astron. Astrophys. **280**, 617 (1993)
125. E. C. Ostriker, C. F. Gammie, J. M. Stone: Astrophys. J. **513**, 259 (1999)
126. E. C. Ostriker, J. M. Stone, C. F. Gammie: Astrophys. J. **546**, 980 (2001)
127. B. Otte, R.-J. Dettmar: Astron. Astrophys. **343**, 705 (1999)
128. B. Otte, R. J. Reynolds, J. S. Gallagher III, A. M. N. Ferguson: Astrophys. J. **560**, 207O (2001)
129. S. Oughton, E. R. Priest, W. H. Matthaeus: J. Fluid Mech. **280**, 95 (1994)
130. P. Padoan, A. Nordlund: Astrophys. J. **526**, 279 (1999)
131. P. Padoan, A. Goodman, B. T. Draine, M. Juvela, A. Nordlund, O. Rognvaldsson: Astrophys. J. **559**, 1005 (2001)
132. E. N. Parker: *Cosmical magnetic fields: Their origin and their activity* (Oxford University Press, New York, 1979)
133. T. Passot, A. Pouquet, P. Woodward: Astron. Astrophys. **197**, 228 (1988)
134. T. Passot, E. Vazquez-Semadeni, A. Pouquet: Astrophys. J. **455**, 536 (1995)
135. H. Politano, A. Pouquet: Phys. Rev. E **52**(1), 636 (1995)
136. H. Politano, A. Pouquet, V. Carbone: Europhys. Lett. 43(5), 516 (1995)
137. H. Politano, A. Pouquet, P. L. Sulem: Phys. Plasmas **2**(8), 2931 (1995)
138. D. Porter, P. Woodward, A. Pouquet: Phys. Fluids **10**, 237 (1998)
139. A. Pouquet, U. Frish, J. Léorat: J. Fluid Mech. **77**, 321 (1976)
140. A. Pouquet, G. S. Patterson: J. Fluid Mech. **85**, 305 (1978)
141. A. Pouquet: in *Interstellar Turbulence*, ed. by J. Franco, A. Carraminana (Cambridge Univ. Press, 1999) p.87
142. R. J. Rand, S. R. Kulkarni, J. J. Hester: Astrophys. J. **352**, 1 (1990)
143. R. J. Rand: Pub. Astro. Soc. Aust. **15**, 106 (1998)
144. R. J. Reynolds: Astrophys. J. **333**, 341 (1988)
145. R. J. Reynolds, L. M. Haffner, S. L. Tufte: Astrophys. J. **525**, 21 (1999)
146. M. N. Rosenbluth, D. A. Monticello, H. R. Strauss, R. B. White: Phys. Fluids **19**, 1987 (1976)
147. D. Ryu, T. W. Jones: Astrophys. J. **442**, 228 (1995)
148. J. Saur, H. Politano, A. Pouquet, W. H. Matthaeus: American Geophysical Union, Fall Meeting (2001)
149. J. M. Scalo: Astrophys. J. **277**, 556 (1984)

150. J. M. Scalo: in *Interstellar Processes*, ed. by D. J. Hollenbach, H. A. Thronson (Dordrecht: Reidel, 1987) p.349
151. A. Schlickeiser, J. A. Miller: Astrophys. J. **492**, 352 (1998)
152. Z.-S. She, E. Leveque: Phys. Rev. Lett. **72**(3), 336 (1994) (S-L)
153. J. V. Shebalin, W. H. Matthaeus, D. C. Montgomery: J. Plasma Phys. **29**, 525 (1983)
154. J. H. Simonetti, J. M. Cordes: in *Radio wave scattering in the interstellar medium; Proceedings of the AIP Conference* (American Institute of Physics, New York, 1988) p. 134
155. J. H. Simonetti: Astrophys. J. **386**, 170 (1992)
156. S. R. Spangler: Astrophys. J. **376**, 540 (1991)
157. S. R. Spangler, C. R. Gwinn: Astrophys. J. **353**, L29 (1990)
158. S. Sridhar, P. Goldreich: Astrophys. J. **432**, 612 (1994)
159. S. Stanimirovic, A. Lazarian: Astrophys. J. **551**, L53 (2001)
160. J. M. Stone, E. C. Ostriker, C. F. Gammie: Astrophys. J. **508**, L99 (1998)
161. H. R. Strauss: Phys. Fluids, **19**, 134 (1976)
162. J. Stutzki: in *Plasma Turbulence and Energetic particles*, ed. by M. Ostrowski, R. Schlickeiser (Cracow, 1999) p.48
163. J. Stutzki, F. Bensch, A. Heithausen, V. Ossenkopf, M. Zielinsky: Astron. Astrophys. **336**, 697 (1998)
164. M. Tafalla, D. Mardones, P. C. Myers, P. Caselli, R. Bachiller, B. J. Benson: Astrophys. J. **504**, 900 (1998)
165. A. Ting, W. M. Matthaeus, D. C. Montgomery: Phys. Fluids, **29**, 3261 (1986)
166. H. J. Volk, F. C. Jones, G. E. Morfill, S. Roser: Astron. Astrophys. **85**, 316 (1980)
167. S. J. Weidenschilling, T. V. Ruzmaikina: Astrophys. J. **430**, 713 (1994)
168. E. Vazquez-Semadeni, T. Passot, A. Pouquet: Astrophys. J. **441**, 702 (1995)
169. E. Vazquez-Semadeni, T. Passot, A. Pouquet: Astrophys. J. **473**, 881 (1996)
170. E. Vazquez-Semadeni: in *Seeing Through the Dust*, ed. by R. Taylor, T. Landecker, A. Willis (ASP: San Francisco, 2002) (astro-ph/0201072)
171. A. Warhaft: Annu. Rev. Fluid Mech. **32**, 203 (2000)
172. O. C. Wilson, G. Munch, E. M. Flather, M. F. Coffeen: Astrophys. J. Supp. **4**, 199 (1959)
173. S. I. Vainstein, F. Cattaneo: Astrophys. J. **393**, 165 (1992)
174. M. K. Verma: Phys. Plasmas, **6**(5), 1455, (1999)
175. M. K. Verma, G. Dar, V. Eswaran: Phys. Plasmas, **9**(4), 1484 (2002)
176. E. T. Vishniac, J. Cho: Astrophys. J. **550**, 752 (2001)
177. J. C. Weingartner, B. T. Draine: Astrophys. J. Supp. **134**, 263 (2001)
178. H. Yan, A. Lazarian: Phys. Rev. Lett. submitted (2002) astro-ph/0205285
179. H. Yan, A. Lazarian: in preparation (2002)
180. V. E. Zakharov: Sov. Phys. JETP, **24**, 455 (1967)
181. V. E. Zakharov, A. Sagdeev: Sov. Phys. Dokl. textbf15, 439 (1970)
182. G. P. Zank, W. H. Matthaeus: J. Plasma Phys. **48**, 85 (1992)
183. G. P. Zank, W. H. Matthaeus: Phys. Fluids A **5**(1), 257 (1993)
184. M. Zielinsky, J. Stutzki: Astron. Astrophys. **347**, 630 (1999)
185. E. G. Zweibel, C. Heiles: Nature, **385**, 131 (1997)

Numerical Simulations of Magnetic Fields in Astrophysical Turbulence

Ellen G. Zweibel[1], Fabian Heitsch[1], and Yuhong Fan[2]

[1] JILA, University of Colorado, Boulder CO 80309, USA
[2] High Altitude Observatory, NCAR, Boulder, CO 80307, USA

Abstract. The generation and evolution of astrophysical magnetic fields occurs largely through the action of turbulence. In many situations, the magnetic field is strong enough to influence many important properties of turbulence itself. Numerical simulation of magnetized turbulence is especially challenging in the astrophysical regime because of the high magnetic Reynolds numbers involved, but some aspects of this difficulty can be avoided in weakly ionized systems.

1 Introduction

The interaction of magnetic fields with turbulence is a basic feature of many astrophysical systems. Important, basic MHD turbulence problems common to many fields of astrophysics include the nature of MHD turbulence itself, the dynamo problem, the effects of magnetic fields on turbulent transport and turbulent mixing, the effects of small scale turbulence on large scale dynamics, and the formation and evolution of current sheets.

Although progress on all of these problems has been made analytically, numerical simulations are an increasingly powerful means of approaching them. Accurate simulation of astrophysical magnetic fields under turbulent conditions presents extreme challenges of its own. The main reason is the smallness of the magnetic diffusivity, which leads naturally to the formation of thin current layers. Most of the Ohmic dissipation, and most of the change in magnetic topology, takes place in these layers. It is important to understand these diffusive effects in order to answer questions such as: How do dynamos amplify magnetic fields on large scales but not small scales? What determines the ratio of magnetic flux to mass in self gravitating regions? How is magnetic energy dissipated in a turbulent plasma?

The purpose of this review is to isolate some important problems related to magnetized turbulence in astrophysics, with emphasis on their computational aspects. The subject is vast, and we make no claim to be comprehensive. We cite literature up to early 2002.

In Sect. 2, we write down the magnetic induction equation, derive some of its basic properties, and discuss the use of conservation laws in testing MHD codes. As an illustration, we carry out such tests on the ZEUS code. In Sect. 3, we discuss the parameter space for astrophysical MHD. In Sect. 4, we discuss some results on the dynamo problem, the turbulence problem, the formation of singularities, and the role of turbulence in large scale dynamics. In Sect. 5,

E. Falgarone and T. Passot (Eds.): LNP 614, pp. 101–126, 2003.
© Springer-Verlag Berlin Heidelberg 2003

we discuss lightly ionized media, for which ambipolar drift yields an effective diffusivity which is much higher than the Ohmic value. Although the two forms of diffusion are not equivalent, ambipolar drift is a partial solution to the problem of small diffusivity. Section 6 is a summary and discussion of future prospects.

2 The Magnetic Induction Equation: Theory and Tests

2.1 Induction Equation and Consequences

According to Faraday's law, also known as the magnetic induction equation

$$\frac{\partial \boldsymbol{B}}{\partial t} = -c\boldsymbol{\nabla} \times \boldsymbol{E}. \tag{1}$$

Throughout this paper, we will assume $\boldsymbol{E}$ is given by

$$c\boldsymbol{E} = -\boldsymbol{v} \times \boldsymbol{B} + \sigma^{-1} c\boldsymbol{J} = -\boldsymbol{v} \times \boldsymbol{B} + \lambda_\Omega \boldsymbol{\nabla} \times \boldsymbol{B}, \tag{2}$$

where σ is the electrical conductivity, $\lambda_\Omega \equiv c^2/4\pi\sigma$ is the magnetic diffusivity, and in the last step we have used Ampere's law, neglecting the displacement current.

The first and second terms on the RHS of (2) represent inductive and resistive effects, respectively. In order to compare these terms, we introduce a characteristic speed V_0 and a characteristic lengthscale L_0, and write $\boldsymbol{v}$ and $\boldsymbol{r}$ in terms of dimensionless velocities and coordinates; $\boldsymbol{v} \equiv V_0 \boldsymbol{u}$; $\boldsymbol{r} \equiv L_0 \boldsymbol{s}$. Equation (2) can then be written as

$$c\boldsymbol{E} = -V_0 \left(\boldsymbol{u} \times \boldsymbol{B} - R_m^{-1} \boldsymbol{\nabla_s} \times \boldsymbol{B} \right), \tag{3}$$

where the dimensionless parameter R_m is defined by

$$R_m \equiv \frac{L_0 V_0}{\lambda_\Omega}. \tag{4}$$

In some problems it is convenient to set V_0 to a typical Alfvén speed $v_A \equiv B/(4\pi\rho)^{1/2}$, in which case R_m is known as the Lundquist number and denoted by S.

The case $\lambda_\Omega \propto S^{-1} \equiv 0$ is called *ideal* MHD. It is clear from (1) and (2) that the limit $S \to \infty$ is a singular limit, in the sense that at this limit the order of the magnetic induction equation drops from second to first. We therefore expect that as $S \to \infty$, thin boundary layers, or current sheets, will form. This should be anticipated when choosing numerical schemes (Sect. 3.3).

The high S limit describes most astrophysical problems. The ideal form of the magnetic induction equation can be solved exactly in terms of fluid trajectories. Suppose that the fluid at position $\boldsymbol{r}$ at time t was at position $\boldsymbol{r}_0$ at time 0. Define the deformation matrix $\boldsymbol{D}$ by

$$D_{ij} = \frac{\partial r_i}{\partial r_{0j}}. \tag{5}$$

It can then be shown that the magnetic field $\boldsymbol{B}(\boldsymbol{r},t)$ is related to the initial field $\boldsymbol{B}_0(\boldsymbol{r}_0(\boldsymbol{r},t),0)$ by

$$\boldsymbol{B}(\boldsymbol{r},t)=\frac{\boldsymbol{D}\cdot\boldsymbol{B}_0(\boldsymbol{r}_0(\boldsymbol{r},t),0)}{|\boldsymbol{D}|}, \tag{6}$$

where $|\boldsymbol{D}|$ is the determinant of $\boldsymbol{D}$. Equation (6) is known as the Cauchy solution of the magnetic induction equation. It is the basis for the numerical technique used by Kinney et al. [53] to simulate 2D MHD turbulence.

Conservation laws can be used to test MHD codes. The Cauchy solution embodies magnetic flux conservation. This basic property is in practice difficult to test, because in turbulent flow, fluid elements follow complex paths. The Cauchy solution only makes sense as long as fluid trajectories do not intersect, but this cannot be guaranteed in a finite difference code, which is always somewhat diffusive. Therefore, we turn to a globally conserved quantity: magnetic helicity.

The helicity H of a fixed volume V of fluid with magnetic vector potential $\boldsymbol{A}$ and magnetic field $\boldsymbol{B}=\boldsymbol{\nabla}\times\boldsymbol{A}$ is

$$H\equiv\int_V d^3r\boldsymbol{A}\cdot\boldsymbol{B}. \tag{7}$$

Uncurling (1) leads to an evolution equation for $\boldsymbol{A}$

$$\frac{\partial\boldsymbol{A}}{\partial t}=-c\boldsymbol{E}+\boldsymbol{\nabla}\phi, \tag{8}$$

where ϕ is a free gauge function. According to (1) and (8), the rate of change of H is

$$\frac{dH}{dt}=-2c\int_V d^3r\boldsymbol{E}\cdot\boldsymbol{B}+\int_S d\boldsymbol{S}\cdot\left(\boldsymbol{B}\phi-c\boldsymbol{E}\times\boldsymbol{A}\right), \tag{9}$$

where we have integrated once by parts and used Gauss's theorem. We now take periodic boundary conditions on V, so that the surface integral vanishes, and use (2). Equation (9) then becomes

$$\frac{dH}{dt}=-2\int_V d^3r\lambda_\Omega\boldsymbol{B}\cdot\boldsymbol{\nabla}\times\boldsymbol{B}. \tag{10}$$

Equation (10) shows that helicity is conserved in an ideal medium. The rate at which helicity varies is a global measure of the magnetic diffusivity λ_Ω. Diffusion can cause helicity growth as well as decay.

We obtain a slightly different conservation law if we assume that V is comoving instead of fixed. In this case, we find that in an ideal medium, H is conserved within V as long as $\boldsymbol{B}\cdot d\boldsymbol{S}\equiv 0$. However, in view of the difficulty of following comoving volumes in a turbulent fluid, it is more useful to treat V as fixed, and to take it as the computational domain.

Finally, we derive an equation for magnetic energy W

$$W\equiv\int_V d^3r\frac{B^2}{8\pi}. \tag{11}$$

Taking the scalar product of (1) with $\boldsymbol{B}$, integrating over space, and assuming periodic boundary conditions yields

$$\frac{dW}{dt} = -\int_V d^3r \boldsymbol{v} \cdot \frac{\boldsymbol{J} \times \boldsymbol{B}}{c} - \int_V d^3r \lambda_\Omega \frac{|\boldsymbol{\nabla} \times \boldsymbol{B}|^2}{4\pi}. \tag{12}$$

The first term on the RHS of (12) represents the work done by the flow on the field, and appears with opposite sign in the evolution equation for kinetic energy. The second term represents energy loss by Ohmic decay.

2.2 Helicity Conservation in the ZEUS Code

The ZEUS-3D code [99,100] solves the equations of ideal, compressible MHD using a finite difference scheme and a von Neumann artificial viscosity to capture shocks. The MHD induction equation is followed using the method of consistent transport along characteristics [46]. ZEUS-3D is publically available, and has been of great service in the astrophysical community.

If there were no numerical dissipation in the ZEUS code, magnetic helicity would be strictly conserved (see (10)). Here, we investigate helicity conservation in the ZEUS code in two applications.

The first problem is the evolution of a twisted magnetic flux tube which is unstable to the kink mode. Initially, the helicity of the tube is entirely in the twist of the field about the axis. Theory predicts that the instability drives the system to a new equilibrium state in which some of the helicity is carried by a writhing deformation of the tube axis, and the total magnetic energy is reduced while the total helicity is fixed.

The simulations confirm the broad outline of this picture. The growth rate of the kink during the linear phase agrees with an analytical calculation, and significant motion occurs only during the kinking phase, while magnetic energy is being released. However, both magnetic energy and magnetic helicity decline steadily once the system reaches equilibrium, as shown in Fig. 1 for computations with 80^3 and 160^3 gridpoints. During the dynamical kink phase, the energy declines much faster than the helicity, confirming the importance of both dynamical and resistive processes in the evolution of magnetic energy.

We have used (10) to estimate the mean pseudo- magnetic diffusivity $\langle \lambda_\Omega \rangle$, i.e. the mean numerical magnetic diffusivity, by writing the integral on the right hand side as

$$\int_V d^3r \lambda_\Omega \boldsymbol{B} \cdot \boldsymbol{\nabla} \times \boldsymbol{B} = \langle \lambda_\Omega \rangle \int_V d^3r \boldsymbol{B} \cdot \boldsymbol{\nabla} \times \boldsymbol{B}, \tag{13}$$

which can be regarded as the definition of $\langle \lambda_\Omega \rangle$. The result is shown in Fig. 2, where $\langle \lambda_\Omega \rangle$ is plotted for the two simulations in units of $\Delta x v_{A0}$, where Δx is the grid scale. The near coincidence of the two curves shows that the numerical diffusion is linear in Δx. During the dynamical phase, $\langle \lambda_\Omega \rangle$ is enhanced over its "quiet" value $\epsilon_0 \Delta x v_{A0}$ by about a factor of 3, perhaps suggesting that numerical resistivity, like numerical viscosity, is proportional to velocity.

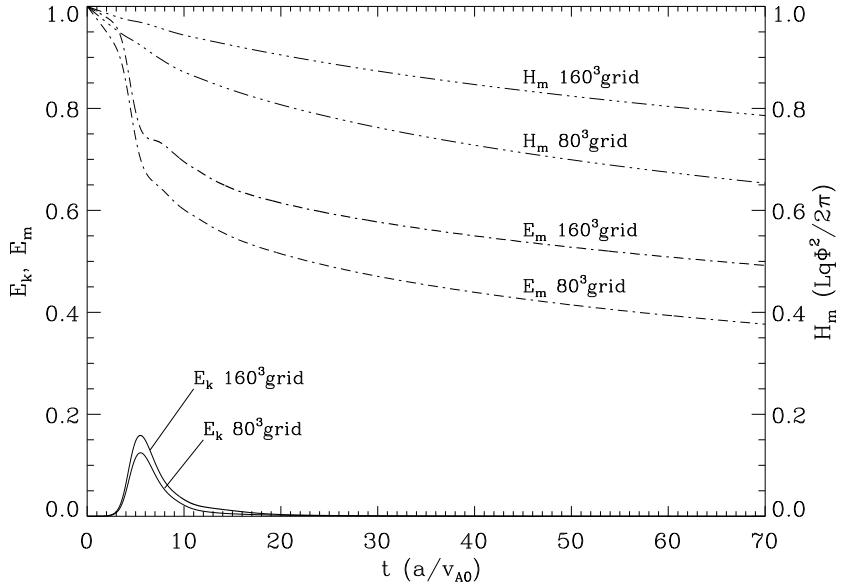


Fig. 1. Kinetic and magnetic energy, and helicity against time for the twisted magnetic flux tube. t is given in units of the Alfvén crossing time for the flux tube with diameter a. Helicity is given in units of $Lq\Phi^2/2\pi$, where $Lq/2\pi$ is the number of rotations by 2π of the twisted field lines about the tube axis over the domain length L, and Φ is the total axial flux of the tube.

The results shown in Fig. 2 allow us to estimate the Lundquist number S. According to (4), S can be written in terms of ϵ_0, the number of gridpoints N, and the magnetic lengthscale L_B in units of the box size L as

$$S = \frac{L_B}{L}\frac{N^{1/3}}{\epsilon_0}. \tag{14}$$

In these computations, the tube has an initially Gaussian magnetic profile with $L_B/L = 0.1$. Taking $\epsilon_0 \sim 0.01$, which is representative of the "quiet" value shown in Fig. 2 and $N^{1/3} \sim 100$, we see that $S \lesssim 10^3$.

We have also computed the evolution of H in models of molecular clouds [47,48], with initially uniform magnetic fields, stirred by supersonic turbulence. In these models H is initially zero, and should remain so, although the helicity density $\boldsymbol{A} \cdot \boldsymbol{B}$ can vary arbitrarily between positive and negative values from point to point. Figure 3 shows H relative to $E_m L$ for two different runs at various times, measured in units of the sound crossing time. In these units, the errors in H are at the 1% level, and, reassuringly, there is no evidence that H drifts steadily away from zero. The variance of H around zero is smaller for the run at Alfven Mach number $M_A = 0.7$ than for the run at $M_A = 6.7$, presumably because the field becomes less tangled if it is relatively strong.

Figure 4 shows mass density and helicity density, for the weak field model, in a plane perpendicular to the mean magnetic field. The figure shows that helicity

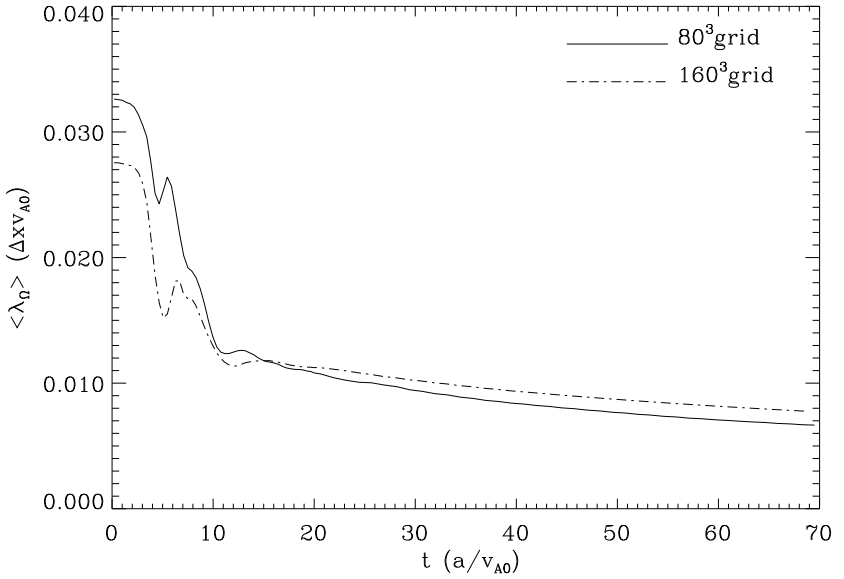


Fig. 2. Mean magnetic diffusivity against time for the twisted magnetic flux tube.

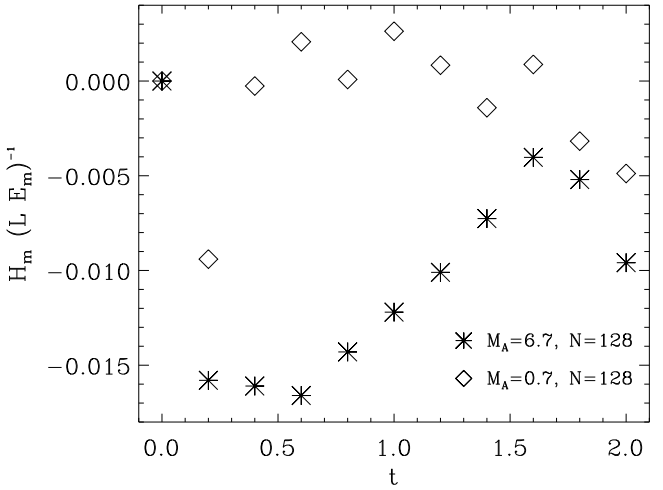


Fig. 3. Total magnetic helicity against time for two models of turbulent molecular clouds at $M_A = 0.7$ and $M_A = 6.7$.

density varies strongly with position, and that it is well correlated with mass density (the correlation is not as good in the strong field case).

The density compressions in these models are associated with shock fronts. Helicity density can also increase across a shock. For example, consider a locally sheared, force free magnetic field of the form

$$\boldsymbol{B} = B_0 \left(\hat{x} \cos k_0 z + \hat{y} \sin k_0 z\right). \tag{15}$$

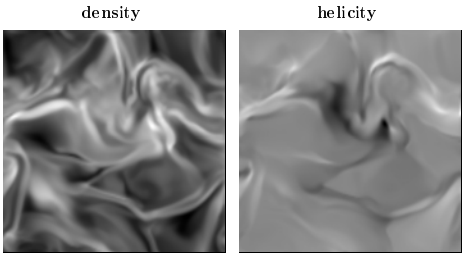


Fig. 4. Density (left panel) and helicity (right panel) for the weak field model with $M_A = 6.7$.

The vector potential is

$$\boldsymbol{A} = -k_0^{-1}\boldsymbol{B}, \tag{16}$$

so the helicity density is $-k_0^{-1}B_0^2$. Suppose the fluid is compressed in the $\hat{z}$ direction, so that the initial and final coordinates z_0 and z are related by $z = \alpha z_0$, with $\alpha < 1$. It can be shown from (6), or from intuitive arguments, that $\boldsymbol{B}$ still has the form (15), but with $B_0 \to \alpha^{-1}B_0$ and $k_0 \to \alpha^{-1}k_0$. The helicity density then increases by a factor α^{-1}, which is the same factor by which the mass density increases.

3 Parameter Space

The dimensionless parameters of astrophysical turbulence define the scope of the problem and should influence the choice of numerical technique. Here we define and give quantitative expressions for some important parameters.

3.1 Macroscopic Parameters

The ratio of gas pressure to magnetic pressure is usually denoted by $\beta \equiv 8\pi P/B^2$, which is closely related to the ratio of the sound speed v_S to the Alfvén speed v_A; $v_S^2/v_A^2 = \gamma\beta/2$, where γ is the ratio of specific heats. If β is either very small or very large, the acoustic and Alfvén frequencies are very different, with consequences for choice of simulation technique.

The ratios of the mean flow speed v to v_S and v_A, respectively, are denoted by the sonic and Alfvén Mach numbers M and M_A. When $M \gg 1$, compressibility effects are important and the flow is pervaded by shocks, requiring an algorithm which treats them accurately.

The importance of self gravity is measured by the Jeans length λ_J. Turbulent pressure effectively increases λ_J. This is captured by a heuristic expression for λ_J in a turbulent medium

$$\lambda_J \equiv \left[\frac{\pi v_S^2\left(1 + M^2(\lambda_J)\right)}{G\rho}\right]^{1/2}, \tag{17}$$

where $M^2(\lambda_J)$ is the Mach number of the turbulence at wavelengths less that λ_J; turbulence at longer wavelengths does not contribute to pressure support [22,12,90]. Equation (17) reverts to the usual definition of the Jeans length as $M \to 0$. Truelove et al. [106] have shown that unless the size of the grid, Δx, remains less than about 0.25 of the local λ_J, the system is subject to spurious local fragmentation.

3.2 Microscopic Parameters

Complete expressions for plasma transport coefficients are given by [13]. We are primarily concerned here with the magnetic and viscous diffusivities, which are conveniently given in terms of the electron and ion collision times τ_e and τ_i

$$\tau_e = \frac{2.9 \times 10^{-2}}{(\Lambda/10)} \frac{T^{3/2}}{n_e Z} s; \quad \tau_i = \frac{1.7}{(\Lambda/10)} \frac{A^{1/2} T^{3/2}}{n_e Z^3} s. \tag{18}$$

In (18), T is given in degrees K, Λ is the Coulomb logarithm ($\Lambda = 9.4 - 1.15 \log n_e + 3.45 \log T$ for $T < 5.8 \times 10^5 K$ and $\Lambda = 15.9 - 1.15 \log n_e + 2.3 \log T$ for $T > 5.8 \times 10^5 K$), and A and Z are the ionic atomic number and charge, respectively. The electron density n_e is expressed in cgs units; cm^{-3}.

The magnetic diffusivity λ_Ω introduced in (2) can also be expressed in terms of the electron skin depth $\delta_e \equiv c(m_e/4\pi n_e e^2)^{1/2} = 5.4 \times 10^5 n_e^{-1/2} \mathrm{cm}$ and τ_e as

$$\lambda_\Omega = \frac{\delta_e^2}{\tau_e} = 9.9 \times 10^{12} \, (\Lambda/10) \, \frac{Z}{T^{3/2}} cm^2 s^{-1}. \tag{19}$$

This numerical expression allows us to estimate the Lundquist number S

$$S \equiv \frac{L_B v_A}{\lambda_\Omega} = 0.022 \frac{L_B T^{3/2} B}{n_e^{1/2} (\Lambda/10)}, \tag{20}$$

where the last expression holds in a hydrogen plasma. Under astrophysical conditions, S is enormous: for example, in ionized interstellar gas with $T = 10^4$, $n_e = 1$, $B = 3 \times 10^{-6} G$, and $L_B = 1$ pc, $S = 10^{17}$. With the exception of extremely dense, cold environments such as protostellar disks, or systems with extremely small lengthscales, such as the outer layers of accreting neutron stars, λ_Ω is always much less than the numerical diffusivity arising from discretization.

The viscous diffusivity, or kinematic viscosity ν, is

$$\nu = \frac{kT}{m_i} \tau_i = \frac{1.4 \times 10^8}{(\Lambda/10)} \frac{T^{5/2}}{n_e A^{1/2} Z^3} cm^2 s^{-1}. \tag{21}$$

The Reynolds number R with respect to an organized flow V on a lengthscale L is defined as LV/ν, and is almost always extremely large, confirming our expectation that astrophysical flow can be structured over a wide range of scales.

According to (19) and (21), the ratio of viscous to magnetic diffusivity, which is known as the magnetic Prandtl number P_r, is

$$P_r \equiv \frac{\nu}{\lambda_\Omega} = \frac{1.4 \times 10^{-5}}{(\Lambda/10)^2} \frac{T^4}{n_e A^{1/2} Z^4}. \tag{22}$$

Typically, $P_r \gg 1$; in the example of fully ionized interstellar gas introduced above, $P_r = 10^{11}$. Therefore, the kinetic energy spectrum should be truncated on much larger scales than the magnetic fluctuation spectrum [55,66,24] The computations by Maron & Cowley [66] show that the essentially constant shear of the velocity field on scales near the resistive scale generates tightly folded, or hairpin-like, magnetic structures.

If the ion cyclotron frequency $\omega_{ci} \sim 1.8 \times 10^4 BZ/A$ and the ion collision time τ_i satisfy the condition $\omega_{ci}\tau_i \gg 1$, the viscosity is highly anisotropic with respect to the magnetic field. The viscous force acting on shear flow parallel to the magnetic field is reduced by a factor of $(\omega_{ci}\tau_i)^2$, while the parallel viscosity remains the same (see [13] for a full description). Using (18), we see that $\omega_{ci}\tau_i$ is indeed large; in our numerical example, it is nearly 10^5, implying very strong suppression of viscosity perpendicular to the magnetic field. Maron & Cowley [66] have implemented the full tensor viscosity in their simulations of turbulent dynamos.

3.3 Implications for Numerical Techniques

In order for a numerical method to include diffusive processes, we would ideally require two things, namely (a) that it can resolve the physically important scales and (b) that it can handle the disparate time scales. Clearly, meeting each of these requirements alone is already a non-trivial task.

Resolving the physically interesting scales means separating the intrinsic numerical diffusion existing in any scheme from its physical counterpart. Simultaneously, it is desirable to cover scales much larger than the diffusive scales as well, in order to embed the problem in a realistic environment. Currently, studies are restricted to one of these regimes, for obvious reasons.

There are various ways to parametrize the small-scale physics in the large scale approach. Shock capturing is a familiar example (for a detailed discussion see e.g. [58]). Many finite-difference codes apply a von-Neumann-Richtmyer artificial viscosity [78], spreading out the shock across several support points and not affecting the solution in smooth regions. Artificial viscosity mimicks physical viscosity in that it satisfies the Rankine-Hugoniot conditions, but on a much larger scale, which renders the numerical scheme very robust at low computational cost. It operates only where the flow is compressive.

Galsgaard & Nordlund [36] implemented an artificial resistivity algorithm similar to artifical viscosity in their simulations of formation and dissipation of current sheets in the solar corona; see also Caunt & Korpi [21].

"Current sheet capturing" is intrinsically more complex than shock capturing. As we discuss in Sect. 4.3, the energy dissipated by a shock is independent of the magnitude of the diffusivity, but this does not hold for a current sheet. Moreover, the topological evolution of the magnetic field is determined primarily by processes in current sheets, and this imposes additional requirements on how they are treated. Detailed, small scale studies could be useful for developing a parameterization of the effects of diffusion on energy balance and topology. This

is reminiscent of the approach taken in large eddy simulations, which numerically resolve the largest scales while treating the unresolved scales with a subgrid model (e.g. [18,27,93]).

Godunov-type methods [41] avoid broadening the shocks over many zones. They solve the Riemann problem defined by a set of advection equations for the physical variables and the physical states on both sides of the discontinuity. The upstream and downstream states are connected by a sequence of shocks or rarefaction waves, each of which satisfies the Rankine-Hugoniot conditions. As implementation of a full Riemann solver would involve determining the full wave structure and all wave propagation speeds, a variety of approximate Riemann solvers has been designed to simplify the process by disregarding certain wave types or linearizing the problem (see [58] for details). Riemann solvers utilize the wave nature of hyperbolic equations, however, the diffusive terms are parabolic, so that they cannot be included in the Riemann solver directly.

An alternative approach has been taken in the so-called BGK-schemes [7], using the collisionless Boltzmann equation as a model of gas dynamics (see [89], and [104] for MHD).

Needless to say, the $P_r \gg 1$ regime cannot be captured by numerical codes which solve the ideal fluid equations ($\eta = \nu = 0$), since the numerical diffusivities for magnetic field and velocity are about the same size and occur at about the same scale, i.e. the grid scale. This is an argument for including viscosity and resistivity in the equations, in which case the physical diffusivities must be larger than their numerical counterparts, and their scales must be well separated. This is the approach taken by Maron & Cowley [66], who used a spectral code to study turbulence with widely separated, but resolved, viscous and resistive scales.

Several efforts have been made to connect the spatial scales by adaptive mesh refinement techniques [5,4] for large eddy simulations [27], general hyperbolic systems [6], incompressible non-ideal MHD [34] and for compressible MHD [110,2]. However, turbulent flows by definition connect scales over the whole domain, which complicates finding a suitable refinement criterion and may in the end yield only modest savings of time.

We have seen in Sect. 3.1 that timescales in astrophysical problems can differ by several orders of magnitude. Including diffusive processes aggravates the problem, as their time scales are generally orders of magnitude longer than the dynamical times. Any disparity in timescales leads to a stiff problem. Explicit methods – often the first choice because of their low time and memory needs – advance the solution by a time step Δt which is given by the so-called Courant - Friedrichs - Levy (CFL) condition [28]

$$\Delta t \leq \frac{\Delta x}{c} \tag{23}$$

with the distance between two support points Δx and a characteristic propagation speed c. Disparate characteristic speeds then lead to severe time step restrictions. Moreover, the slowly varying components may introduce numerical errors [50].

One solution to the problem is to choose a short timestep for the fast processes, and update the variables controlled by slow processes less frequently. Mac Low et al. [63] used this so-called subcycling in their implementation of ambipolar diffusion in the ZEUS code. Depending on the problem, the critical processes often can be broken out and treated implicitly, while the remaining equations can be treated explicitly [64]. Fully implicit schemes (e.g. [51]) can be unconditionally stable and independent of the time step chosen. Their main limitation lies in the tremendous computational needs in the multi-dimensional case. For a discussion of implicit methods and for a comparison between explicit and implicit shock capturing methods see [50,109]. For a discussion of higher order finite difference schemes and their application to MHD turbulence simulations see [14].

As this brief discussion shows, a great variety of techniques can be brought to bear on MHD turbulence problems. Which technique is best depends very much on the problem at hand.

4 Results from Theories and Simulations

Theory offers an interpretive framework for simulations, while numerical experiments test turbulence theory, and can inspire new developments. Here, we discuss amplification of a field by turbulence in the context of dynamo theory, magnetic fluctuations, the formation of current sheets, and the dynamical effects of turbulence on large scales. The first three of these problems are closely related. The fourth is topical in view of recent simulations of star forming regions.

4.1 Turbulent Amplification of a Weak Field

Amplification of a weak magnetic field by turbulence is one of the main components of dynamo theory. A successful astrophysical dynamo theory must demonstrate that a dominant, large scale magnetic field can be generated by small scale turbulence, on a timescale which is virtually independent of the Lundquist number S in the limit $S \to \infty$. This is a highly nonlinear problem, which involves multiple scales. It probably cannot be solved without numerical simulations, but in view of the extreme parameters involved it is unlikely to yield to brute force alone.

The growth rate of magnetic energy in a periodic domain is given by (12). Although the resistive term appears as a sink, some resistivity is necessary to effect irreversible topological change. Amplification occurs through work done by the flow. In general, there is also a surface integral $-c \int \boldsymbol{dS} \cdot \boldsymbol{E} \times \boldsymbol{B}$ on the RHS of (12), which represents an energy flux through the boundaries.

Equation (12) says nothing about the structure of the field, and, in particular, whether magnetic energy is concentrated at large or small scales. If we formally set $\lambda_\Omega \equiv 0$, the ratio of fieldstrength to line length is preserved by incompressible motions. The only way to lengthen fieldlines in a finite volume is to wind or tangle them. Therefore, we expect that most of the energy will be in the *rms*

field, rather than in the mean field. For example, if the protogalactic magnetic field were 10^{-17} G [40], the fieldlines would have been lengthened by more than a factor of 10^{11} in the course of amplifying the field to its present strength. Yet, the large scale and small scale components of the galactic magnetic field are observed to be roughly equal [113]. It is not enough to argue that magnetic forces will prevent the field from becoming tangled on small scales. If the lines aren't somehow lengthened, the field won't be strengthened.

Of course, we are interested in $\lambda_\Omega \to 0$, not $\lambda_\Omega \equiv 0$. The fieldlines are allowed to break, and the question is whether they do so in a way which prevents power from piling up at the resistive scale and peaking instead at the large scale. Because S is so large, topological constraints are strong, and we expect resistive boundary layers to form.

Let's see how theories and simulations of dynamo theory address this problem. The most influential and detailed dynamo theory is the so-called mean field theory, which was proposed by Parker [83], and extended and formalized by Steenbeck, Krause & Rädler [97]. Mean field theory is based on the idea that a large scale magnetic field $\langle \boldsymbol{B} \rangle$ can be generated from the average inductive electric field $\langle \boldsymbol{E} \rangle$ associated with small scale, helical velocity and magnetic fluctuations $\boldsymbol{\delta v}$, $\boldsymbol{\delta B}$. The dynamo property of small scale, helical fluctuations is known as the α effect. There is also diffusive transport of $\langle \boldsymbol{B} \rangle$, which is known as the β effect. Although α and β are in general tensors, it suffices for our purposes to write them as scalars. The induction equation for the mean field is

$$\frac{\partial \langle \boldsymbol{B} \rangle}{\partial t} = -c \boldsymbol{\nabla} \times \langle \boldsymbol{E} \rangle, \tag{24}$$

where

$$\langle c\boldsymbol{E} \rangle + \langle \boldsymbol{V} \rangle \times \langle \boldsymbol{B} \rangle = \lambda_\Omega \boldsymbol{\nabla} \times \langle \boldsymbol{B} \rangle - \langle \boldsymbol{\delta v} \times \boldsymbol{\delta B} \rangle \equiv \alpha \langle \boldsymbol{B} \rangle + (\lambda_\Omega + \beta)\, \boldsymbol{\nabla} \times \langle \boldsymbol{B} \rangle + ... \tag{25}$$

In (25), $\langle \boldsymbol{V} \rangle$ is the large scale velocity field, and the "..." represent additional terms involving moments of the turbulent fluctuations and successively higher derivatives of $\langle \boldsymbol{B} \rangle$ (see Moffatt [74]). Note the correspondence between (24) and (25) and (1) and (2).

If α and β are independent of $\langle \boldsymbol{B} \rangle$, (24) is linear in $\langle \boldsymbol{B} \rangle$. Physically, this corresponds to neglecting the back reaction of the magnetic field on the turbulent velocity, so the linear case is sometimes called the kinematic case.

The linear version of (24) has solutions which vary exponentially in time. At large shear and low wavenumber k, the growth rate is approximately the geometric mean of the shear rate V' and the turbulent frequency $k\alpha$.

Standard mean field dynamo theory does not directly address the evolution of the small scale field, and the role of resistivity, on which the fate of the small scale field depends, is left somewhat implicit. Usually λ_Ω is dropped from (24), since β is assumed to be much larger. A calculation by Moffatt [74] of the α effect due to helical Alfvén waves demonstrates that $\alpha \equiv 0$ unless there is a phase difference between $\boldsymbol{\delta v}$ and $\boldsymbol{\delta B}$. In Moffatt's model, the phase lag arises from magnetic diffusivity. At small λ_Ω, $\alpha \propto \lambda_\Omega$, implying slow dynamo action at large R_m.

In more general calculations of α, the diffusive damping time $(k^2\lambda_\Omega)^{-1}$ is replaced by the correlation time τ of the turbulence. At large R_m, this is much shorter than the resistive time, and implies large growth rates for the dynamo. But β does not distinguish between converting large scale field to small scale field and actually destroying field, so the behavior of the small scale fields are obscured by this treatment. Kulsrud & Anderson [55] used the tools of mean field theory itself to follow the growth of the small scale field. They showed that at large R_m, the rms field $\langle B^2\rangle^{1/2}$ grows much faster than the mean field $\langle \boldsymbol{B}\rangle$, and is dominated by small scale fluctuations with a $k^{3/2}$ power spectrum.

A completely different perspective on kinematic dynamo theory emerges from solutions of the full magnetic induction equation for prescribed, chaotic flow. These studies, which are fully described up to 1995 in the volume by Childress & Gilbert [23], use a combination of numerical simulations at large R_m and analytical theory to establish connections between the properties of the flow and the properties of the magnetic field. Although there are known examples of flows which amplify the magnetic field at a rate independent of R_m, as originally shown by Galloway & Proctor [35], the fields produced by these so-called fast dynamos are highly intermittent in space and fluctuate in sign, confirming at least qualitatively that small scale fields are dominant in kinematic dynamos at high R_m.

The kinematic theories are modified by dynamical effects. It was expected on general grounds that the dynamo would saturate as the magnetic field approached equipartition with the kinetic energy. A saturation process known as "Alfvénization" was first described by (Pouquet et al [88]). Saturation occurs because the small scale fluctuations increasingly resemble Alfvén waves as the mean field grows. In an ideal medium $\boldsymbol{\delta B} \rightarrow \pm\boldsymbol{\delta v}$ in Alfvén units, so their cross product also tends to zero, eliminating the α effect (also shown in the calculation by Moffatt). Alfvénization was first identified by Pouquet et al. [88] using a spectral closure scheme, and later derived using quasilinear theory by Gruzinov & Diamond [44]. In the model of Pouquet et al., kinetic helicity injected at the forcing scale generates magnetic helicity, which undergoes both a direct cascade, to high k, where it is resistively dissipated, and an inverse cascade, to low k. Alfvénization saturates the dynamo once the large scale field is roughly in equipartition with the kinetic energy. In the model of Gruzinov & Diamond, saturation by Alfvénization takes place when the ratio of kinetic to large scale magnetic energy is $\mathcal{O}(R_m)$, confirming the dominance of small scale fields at large R_m. The difference between these calculations may lie in their treatments of diffusion. Gruzinov & Diamond explicitly include λ_Ω, while Pouquet et al. took all the diffusivities to be eddy diffusivities, which are much larger than molecular diffusivities.

The role of kinetic helicity injection has been tested in simulations. Computations by Maron & Cowley [66], without kinetic helicity injection, show a peak in spectral power at the resistive scale, as predicted by Kulsrud & Anderson [55]. Simulations by Maron & Blackman [65] show that a large scale field is generated when kinetic helicity is injected, but that the growth rate is proportional to the

resistive time. Nonlinear simulations with helical forcing by Brummell et al. [17] show that the magnetic lengthscale is proportional to $R_m^{-1/2}$, while *kinematic* investigations testing the sensitivity of fast dynamo action to kinetic helicity have not found an inverse cascade [49].

Thus, we see that much hinges on the resistive time, or, more broadly, the mechanism by which small scale fields are dissipated. Turbulent diffusion of $\langle \boldsymbol{B} \rangle$, parameterized by the β effect, merely stands for spectral transfer out of the large scale. The possibility remains that turbulence could mix the field to the small scale, efficiently destroying it. However, numerical models by Cattaneo et al. [20], and subsequent analytical work by Kim [52] suggest that the mixing is self limiting due to Lorentz forces.

The possibility that the small scale fields might escape through an open boundary rather than being destroyed *in situ*, leading to growth of the large scale field on the convective timescale rather than the resistive timescale, was raised by Parker [86] and Blackman & Field [11] Up to now, neither analytical studies [92] nor numerical models [15] have confirmed this idea.

Astrophysical dynamo theory is at a critical juncture. The fundamental premise of mean field dynamo theory - that kinetic helicity injected at small scales can drive the growth of a magnetic field at large scales - is supported by a variety of analytical treatments and numerical investigations. Mean field theory is a felicitous outcome of turbulence theory in the sense that it can be applied just by calculating a few parameters (or functions): α, β, and the large scale shear. However, neither theory nor simulations have fully come to grips with the large value of R_m and its implications for the small scale fields. From a numerical point of view, computations at large R_m which can capture current sheets as well as accommodating a variety of boundary conditions would seem to be a prerequisite for either validating or superseding mean field theory.

4.2 Magnetic Fluctuations

Steady hydrodynamic turbulence, uninfluenced by boundaries, has been characterized in two regimes. Subsonic turbulence is incompressible, and follows the Kolmogorov scaling ($E(k) \propto k^{-5/3}$). Supersonic hydrodynamic turbulence is compressible, and better described as Burgers turbulence ($E(k) \propto k^{-2}$). (These power law spectra, and all subsequent ones, describe only the so-called inertial range, in which there is no driving or microscopic dissipation, only nonlinear spectral transfer.)

Both these limiting forms of turbulence are spatially intermittent in the sense that the variances of quantities such as the rate of energy dissipation undergo substantial fluctuations from point to point. In Burgers turbulence the basic structures are shock waves, and the large fluctuations and gradients occur principally in the shock fronts. In Kolmogorov turbulence, the basic units are eddies, and intermittency is represented by the concentration of vorticity into small structures. Intermittency does not affect the energy spectrum of Burgers turbulence (the Fourier spectrum of a shock itself is k^{-2}), but is thought to steepen

the Kolmogorov spectrum. This occurs through the local enhancement of the dissipation rate at small scales [57].

Hydrodynamic turbulence in unstratified, nonrotating systems is isotropic. A uniform component of magnetic field causes turbulence to be anisotropic.

An ideal, adiabatic fluid with a uniform component of magnetic field has three linear modes: the shear Alfvén mode, and the fast and slow magnetosonic modes. The velocity field $\boldsymbol{\delta v}$ of the Alfvén mode, is solenoidal, and this mode is driven purely by magnetic tension. The dispersion relation is $\omega = k_\parallel v_A \equiv \omega_A$, where $k_\parallel$ is the component of k parallel to the large scale magnetic field. The fast and slow magnetosonic modes are compressive, and driven by a combination of gas pressure, magnetic pressure, and magnetic tension.

Since the shear Alfvén mode is noncompressive, it is reasonable to ask whether a strong ($\beta < 1$) magnetic field permits a regime of pure shear, supersonic turbulence. The answer is no.

First, suppose $\boldsymbol{k}_\perp \equiv 0$. Then, the nonlinear magnetic pressure gradient associated with the fluctuating transverse magnetic field drives a compressive parallel flow, causing an Alfvén wavetrain to steepen nonlinearly into a train of weak shocks [26], much as a sound wave steepens. The only exception is the infinitely long, circularly polarized wave, which is an exact solution of the ideal MHD equations. The outcome of the evolution of an ensemble of parallel propagating Alfvén waves is a series of shocks, similar to Burgers turbulence, with k^{-2} kinetic and magnetic energy spectra [39]. The steepening time depends on amplitude as $(\delta v/v_A)^{-2}$, while the steepening time of an acoustic wave is proportional to $(\delta v/v_S)^{-1}$. In this sense, shear waves survive longer in a low β plasma than compressive waves of the same amplitude, by a factor of order $\beta^{-1/2}(v_A/\delta v)$, but the outcome in either case is a compressive flow dominated by shocks, and the difference in timescales is only significant for small wave amplitude.

Although Burgers turbulence was originally based on 1D flow, 3D simulations of supersonic MHD turbulence show similar rapid evolution to a shock dominated flow. The distribution of shock strengths in simulations of driven and decaying turbulence, with and without magnetic fields, has been studied by Smith et al. [95,94].

The steepening of Alfvén waves, which can be viewed as a cascade in $k_\parallel$, is suppressed if $\beta \geq 1$. Under the assumption of incompressibility, and in contrast to the highly compressible low β case, interactions occur only between oppositely directed wave packets. In particular there is no self interaction of the kind which leads to steepening.

The incompressible shear Alfvén wave cascade is highly anisotropic, with power transferred much more rapidly in $k_\perp$ than in $k_\parallel$. In physical space, the correlation length transverse to the mean field is much shorter than the correlation length along it. These anisotropic states can be described by so-called reduced MHD [101], which is a quasi-2D approximation describing elongated magnetic structures. It is often applied to (and, indeed, was derived for), low β plasmas, but it neglects the parallel steepening which leads to shocks. This,

as we said above, is a good approximation for long parallel Alfvén transit times and small turbulent amplitudes.

There is not yet complete consensus on the cascade itself. Weak MHD turbulence, in which an individual Alfvén wave packet survives for many periods, can be viewed as weakly perturbative resonant interactions between multiple waves. Sridhar & Goldreich [96] argued that the dominant interaction is a four wave interaction, which results in a spectrum $E(k) \propto k^{-7/3}$. Others [79,8,38] claim that the dominant interactions are three wave interactions, and predict $E(k) \propto k^{-2}$.

In strong MHD turbulence, wave packets survive for only about one wave period. Goldreich & Sridhar [43] developed the concept of a critically balanced cascade in which the wave frequency ω_A is the same as the nonlinear frequency $k_\perp v_\perp$. This, together with the requirement of constant spectral energy flux argument $k_\perp v_\perp^3$, leads to a Kolmogorov spectrum $E(k) \propto k^{-5/3}$ and wavenumber anisotropy $k_\parallel / k_\perp \propto k_\perp^{-1/3}$. Other theories of strong MHD turbulence predict $E(k) \propto k^{-3/2}$, including the original isotropic Iroshnikov-Kraichnan theory [77].

Numerical simulations, if free of confounding computational effects, could play a role in resolving these disagreements. Biskamp & Müller [10] and Cho & Vishniac [25] find $E(k) \propto k^{-5/3}$ over about one decade in k space. Maron & Goldreich [67] slightly extended the inertial range and found $E(k) \propto k^{-3/2}$. These authors argue that the spectrum is flattened because of intermittency. The crucial difference between the HD and MHD cases is that energy cascades in MHD only when oppositely directed wave packets collide; the small filling factor associated with intermittency reduces the collision rate, more than offsetting the enhanced dissipation associated with small structures.

Larger computations, with an extended inertial range, should shortly become available. The differences in the spectra obtained with different codes and under different forms of driving (Biskamp & Müller studied decaying or isotropically forced turbulence, Cho & Vishniac studied isotropically forced turbulence, and Maron & Goldreich studied anisotropically forced turbulence) are at this point comparable to the theoretical disagreements. A joint exercise in which different groups simulated identical problems and implemented identical diagnostics might be quite enlightening.

A variety of processes can terminate the cascade at short wavelengths (see [59] for a recent discussion). These include fluid effects, such as ion-neutral friction, and kinetic effects such as ion gyroresonance absorption and electron Landau damping [91]. The extension of the magnetic spectrum to scales below the viscous cutoff in a high Prandtl number plasma is discussed by [66] and [24].

4.3 Intermittency: Current Sheets and Flux Tubes

Simulations of MHD turbulence show strong intermittency in the distribution of magnetic field gradients, or current. In the weak field, high Prandtl number simulations of [66], the fluctuating field is tightly folded. These results suggest that we look at current sheets.

Current sheets and filaments are sites of intense local heating, and possibly particle acceleration, which makes them interesting in their own right. It is quite likely that in most astrophysical systems, significant departures from flux freezing occur only in these regions, so their existence is important in the evolution of magnetic field topology, and for dynamo action, (see Sect. 4.1). Reconnection at magnetic X-points drives strong, small scale jets. Simulations by Galsgaard & Nordlund [37] show highly time variable dissipation rates associated with the formation and disruption of large scale current sheets.

Current sheets are a prerequisite of, and accompany, magnetic reconnection. Reconnection is a resistive process which occurs on a timescale intermediate between the Ohmic time and the Alfvén time. The reconnection time τ_{rec} is often parameterized by the Lundquist number S; τ_{rec} scales as S^p with $0 \leq p < 1$. If $p = 0$, the reconnection is said to be fast (in correspondence with the definition of a fast dynamo). In Sweet-Parker reconnection, $p = 1/2$ [85], and resistive tearing modes in slabs and cylinders have $p = 3/5$, and $p = 1/3$, respectively [3]. The formation and behavior of current sheets has been studied for many years. Most of this work has focussed on plasmas in or near magnetostatic equilibrium, which can be expected if $\beta \ll 1$ and the plasma is either allowed to relax or is driven at frequencies far below its characteristic Alfvén frequency. Such conditions hold, for example, in stellar coronae, in which the magnetic fieldlines are tied to the underlying photosphere, and evolve quasistatically in response to slow photospheric motion.

Current sheets form as a plasma seeks equilibrium while obeying the topological constraints imposed by flux freezing [75]. Topological constraints can arise from features such as null points [102] or separatrices [62,115,60]. Sweet [103] and Parker [84] suggested that shearing and winding the footpoints of an initially uniform field will create current sheets. Even a simple sinusoidal deformation of the boundaries of a sheared magnetic slab can induce the formation of a current sheet [45].

As a simple example of how current sheets can form, consider the magnetic field of a pair of neighboring, uniformly magnetized superconducting spheres, with antiparallel magnetic dipole moments, surrounded by a perfectly conducting, zero pressure plasma. Some fieldlines have both endpoints on the same sphere, and others have one endpoint on each sphere.

Now suppose that one sphere rotates on its magnetic axis by an angle ϕ. The fieldlines which are not connected to the other sphere also rotate by ϕ, but the fieldlines which are connected to the other sphere cannot rotate uniformly because one endpoint remains fixed. The separatrix surface which divides the two domains of magnetic connectivity becomes a current sheet: the field outside it is sheared while the field inside it is not.

There is still no comprehensive theory for exactly when current sheets form, and how their properties reflect the underlying magnetic topology. Nevertheless, analytical arguments [108,61] and numerical experiments [73,61,36] show that shear, or equivalently current density, grows exponentially rapidly in line tied magnetic fields with randomly moving footpoints. This has some of the same

consequences as an exact singularity, which cannot form in any case in even a slightly resistive plasma.

These results are not fully applicable to MHD turbulence, in which the magnetic field is not in equilibrium. Nevertheless, current sheets form in flows with magnetic null points [87]. The underlying mechanism for the growth of the current is the random stretching of neighboring fieldlines, which leads to large cross-field shear.

What are the consequences of current sheet formation for dissipation of magnetic energy? That is, if energy per unit volume $\dot{e}$ is being added to a system by a driving process, under what circumstances can this energy be dissipated in current sheets?

Consider a current sheet of area L^2 and width δ. Assume the current sheet structure is given by the Sweet-Parker theory, so that $\delta{=}LS^{-1/2}{=}(L\lambda_\Omega/v_A)^{1/2}$. If $N(L)dL$ is the number density of current sheets with L between L and $L+dL$, then the rate at which energy is dissipated by current sheets in this size range is

$$d\dot{w} = \frac{B^2}{4\pi}\frac{v_A}{L}N(L)L^2\delta dL = \frac{B^2}{4\pi}N(L)\left(L^3\lambda_\Omega v_A\right)^{1/2}dL. \tag{26}$$

The rate of energy dissipation by all current sheets is obtained by integrating (26) over L:

$$\dot{w} = \frac{B^2}{4\pi}\left(\lambda_\Omega v_A\right)^{1/2}\int dLN(L)L^{3/2}. \tag{27}$$

If $\dot{e}$ is independent of λ_Ω as $\lambda_\Omega \to 0$, then $\int dLN(L)L^{3/2}$ must scale as $\lambda_\Omega^{-1/2}$ in this limit. This would imply an unbounded number of current sheets as $\lambda_\Omega \to 0$, which is virtually equivalent to a pileup of magnetic energy in fluctuations at the resistive scale. [1]

There are alternative possibilities. One is that the current sheets are not structured according to the Sweet-Parker picture. If δ scaled as λ_Ω, as it does (up to a logarithmic factor) in Petschek's model of fast reconnection (reviewed in [85]), then $\int dLN(L)L^{3/2}$ could be independent of λ_Ω. Note, however, that while in the Sweet-Parker theory, magnetic energy is converted to kinetic energy and to heat at about the same rate, in Petschek's theory magnetic energy is converted predominantly to kinetic energy, and that the realizability of Petschek's model has recently been questioned [9,107].

The second possibility is that the magnetic and velocity fields adjust themselves such that the driving just balances the damping, as they do in mechanical systems such as linear harmonic oscillators. Thus, the energy input rate $\dot{e}$ would depend on λ_Ω, while the applied force itself, presumably would not.

Dedicated simulations of current sheets and reconnection regions themselves allow us to study the small scale aspects of the problem. It is equally important to understand how current sheets are embedded in the overall flow. This has

[1] If we make a similar argument for the efficiency of energy dissipation in shocks, the result is independent of viscosity, because the shock thickness is directly proportional to viscosity. Therefore, the inviscid limit does not require an infinite number of shocks to dissipate the energy input by driving.

already been done to some extent for laboratory experiments [107]. Ultimately, it may be possible to parameterize magnetic reconnection in simulations on larger scales.

Finally, we mention an even more extreme form of intermittency: the segregation of the magnetic field into thin tubes. This is observed in the solar photosphere, where magnetic flux tubes, or sheets, just a few hundred km thick are seen at the borders of convective cells. Flux tubes can form only in high β plasma, so that gas pressure can balance magnetic pressure at the tube walls.

Early attempts to explain photospheric flux tubes relied on a two stage process, proposed by Parker, for concentrating a diffuse field into organized structures. According to this scenario, thermal convection would sweep the field to the borders of convective cells, concentrating the field up to equipartition with the flow. Then reduced convective heat transport would cool the gas, "collapsing" the tube and bringing it to equipartition with the thermal plasma.

An alternative possibility is that intermittent fields are generated *in situ*. It has been shown that small scale turbulent thermal convection operates as a dynamo, and generates a magnetic field with a broad tail of high energy features [19].

It is unclear whether a dynamo operating in a high β plasma should always produce flux tubes. In the solar case, the turbulence is organized by stratification and the thermal gradient, and is subsonic. If the turbulence were supersonic and the field were amplified to equipartition, then flux tubes could not persist.

4.4 Dynamical Effects of Turbulence

The stresses associated with turbulence can exert dynamical forces. In MHD turbulence, these forces include both magnetic stresses and the Reynolds stress. Observations suggest that these stresses are important in the interstellar medium, and especially in molecular clouds.

It was realized early that supersonic motions are nearly ubiquitous in molecular gas, and pose a problem. If the motions are associated with gravitational collapse, the implied star formation rate would be very high [111], but if the motions are turbulent, they should be dissipated very quickly [42]. This conundrum, together with the expectation that reasonably strong magnetic fields should be present, prompted analytical studies of Alfvénic turbulence under molecular cloud conditions [1,114,32]. This work demonstrated that even when $M_A \ll M$ and the waves are purely transverse, ion-neutral friction and nonlinear steepening are strong damping mechanisms, which limit the lifetimes of Alfvén waves to a few 10^6 yr or less under molecular cloud conditions.

Even so, under the assumption that the waves might be replenished by some energy source, various aspects of the theory of self gravitating clouds supported by Alfvén wave stresses were developed and applied to model clouds [90,33,69,68,70]. According to this theory, which is based on the the weak turbulence approach pioneered by Dewar [29], Alfvén waves have an isotropic stress tensor $\boldsymbol{P} = P_w \boldsymbol{I}$. The waves pressure - density relation is $P_w \propto \rho^{1/2}$ in a stratified, static medium, while $P_w \propto \rho^{3/2}$ under slow, spatially uniform changes in

density with time. The negative polytropic index ($n = -1/2$) of the $P-\rho$ relation leads to a large center to surface pressure contrast, in accord with observations (see [70] for a good discussion).

The theory of wave supported clouds is most relevant to objects which are slightly magnetically supercritical, that is, the ratio of mass to magnetic flux is slightly too large for magnetostatic support. If the clouds were subcritical they would be supported by the DC magnetic field. If they were highly supercritical, the amplitude of the turbulence required to sustain them would be so large that the weak turbulence theory would probably not apply, although they could still be turbulently supported. From an observational viewpoint, the most likely candidates for wave support are dense cores with sonic or mildly supersonic linewidths [76].

The picture of strongly supersonic turbulence which emerges from numerical simulations of molecular clouds is quite different from this analytical picture, primarily because, as we discussed in Sect. 4.2, generation of compressive disturbances is entirely unavoidable. These compressive disturbances, or shocks, sweep up most of the mass into thin layers in which the local Jeans length is small. These layers, and the even denser structures formed where layers intersect, collapse if the domain is magnetically supercritical, whether the turbulence is freely decaying or maintained in a steady state by driving [47,81]. The driven models are globally stable in the sense that λ_J computed using the total turbulent energy and mean density is larger than the length of the domain. The small, high density structures collapse because the turbulent power at wavelengths less than the local Jeans length is relatively small (see (17)), and because the intensity of turbulence inside the cores is comparable with the density outside. Nevertheless, the bound structures are small, and the role of numerical diffusion within them, which damps the turbulence and removes the support by the DC magnetic field, has not been quantified. Therefore, while the formation of turbulently supported clumps has not been found in the numerical models, their existence cannot be ruled out based on the present models.

5 Ambipolar Drift

As we have discussed in the preceding sections, both theoretical and computational aspects of astrophysical magnetohydrodynamics are dogged persistently by the enormous value of R_m. There is one situation, however, in which the magnetic diffusivity is large whether or not it is enhanced by turbulence: in weakly ionized gas, the magnetic field and plasma drift with respect to the neutrals. This so-called ambipolar drift is not entirely equivalent to resistive diffusion, because it preserves magnetic topology (the field is frozen to the plasma), but it does change the ratio of magnetic flux to mass, as originally pointed out in [72], and dissipates fluctuations on small scales.

Full treatment of ambipolar drift requires solving the equations of MHD for the plasma and neutrals treated as distinct species coupled by collisions, ionization, and recombination [30]. At the low ionization fractions expected in

molecular clouds, the ion Alfvén speed $v_{Ai} \equiv B/\sqrt{4\pi\rho_i}$ can be several orders of magnitude larger than the other characteristic speeds in the problem, reducing the maximum possible timestep by a similar factor. This and other numerical issues related to implementation of ambipolar drift are discussed by [63,105,64,98].

On timescales longer than the ion - neutral collision time τ_{in}, and at low ionized mass fractions, the ion-neutral drift v_D is determined by balancing the Lorentz force on the ions against the frictional drag by neutrals

$$\boldsymbol{v}_D = \frac{(\boldsymbol{\nabla} \times \boldsymbol{B}) \times \boldsymbol{B}}{4\pi\rho_i\nu_{in}} \equiv \frac{v_{Ai}^2\tau_{in}}{L_B}\hat{f}, \tag{28}$$

where $\hat{f}$ is a unit vector in the direction of the Lorentz force. The second equality in (28) can be taken as a definition of the magnetic lengthscale L_B. The quantity $v_{Ai}^2\tau_{in}$ has units of diffusivity, and from now on we refer to it as λ_{AD}. With this definition, $v_D = \lambda_{AD}/L_B$. Note that $\rho_i/\tau_{in} = \rho_n/\tau_{ni}$ and $\rho_n \approx \rho$, so $v_{Ai}^2\tau_{in} = v_A^2\tau_{ni}$, and λ_{AD} is often expressed in the latter terms.

The ambipolar Reynolds number R_{AD} is defined as the ratio of the bulk velocity v to the drift velocity v_D

$$R_{AD} \equiv \frac{L_B v}{\lambda_{AD}}, \tag{29}$$

and measures how well the magnetic field is frozen to the bulk fluid on scale L_B [112]. Setting $R_{AD} = 1$ correctly predicts the thickness of shocks in which the main dissipation mechanism is ion-neutral friction [31], the wavelength at which Alfvén waves are critically damped [56], and the minimum size of an eddy which can wind up a magnetic field [112].

Simulations of supersonic MHD turbulence [82] with varying strengths of ambipolar drift shows that, on average, the drift smooths out the current, and reduces the rms Lorentz force.

We expect (with one caveat discussed at the end of this section) that the magnetic field should show very little structure below the scale $L_{min} \equiv \lambda_{AD}/v$ at which $R_{AD} = 1$. Consider a volume L^3 of turbulent gas with mean density ρ, turbulent velocity v, and mean magnetic field B. If we introduce the crossing time $\tau_d \equiv L/v$ and the Alfvén Mach number $M_A \equiv v/v_A$, then we can write

$$\frac{L_{min}}{L} = \frac{1}{M_A^2}\frac{\tau_{ni}}{\tau_d}. \tag{30}$$

Expressing v in units of km s^{-1} as v_5, and L in pc as L_{pc}, $\tau_d \sim 3\times10^{13}L_{pc}/v_5$ s. For the particle species expected in molecular clouds, $\tau_{ni} \sim 5\times10^8 n_i^{-1}$s. With the ionization equilibrium relation $n_i \sim 10^{-5}n_n^{1/2}$ (cgs; [71]), (30) becomes

$$\frac{L_{min}}{L} = 1.7\frac{v_5}{n^{1/2}L_{pc}}\frac{1}{M_A^2}. \tag{31}$$

Comparing (31) with the grid spacing $\Delta x/L = 1/N$, we see that a moderately large numerical simulation (say 256^3), which is not highly super-Alfvénic or

extremely dense, should resolve almost all the magnetic structure associated with turbulence.

Equation (30) can also be written in terms of the critical magnetic field $B_c \equiv 2\pi G^{1/2}\rho L$ and the free fall time $\tau_{ff} \equiv (4\pi G\rho)^{-1/2} \sim 5.5 \times 10^{14} n_n^{-1/2}$s as

$$\frac{L_{min}}{L} = \left(\frac{B}{B_c}\right)^2 \frac{\tau_{ni}\tau_d}{4\tau_{ff}^2} \sim \left(\frac{B}{B_c}\right)^2 \frac{\tau_{ni}}{4\tau_{ff}}, \tag{32}$$

where the last step holds for virial equilibrium; $\tau_{ff} \sim \tau_d$. Of course, (30) is entirely independent of G.

The modest value of R_{AD} in molecular clouds has been commented upon and exploited by [54,80].

Ambipolar drift is essentially diffusive when the magnetic field has a well ordered, well combed structure. Because the ambipolar diffusivity is proportional to B^2, the diffusion is nonlinear. Sharp fronts can be generated in the vicinity of magnetic nulls, and minima are steepened [16,63,112,64].

6 Summary and Future Agenda

The theme of this paper is how numerical simulations can best cope with the enormous magnetic Reynolds numbers encountered in most astrophysical problems. We argued that, because R_m (and the Lundquist number S) are so large, resistive effects occur primarily in thin sheets or filaments, which must be treated accurately in order to follow the topological evolution of the magnetic field.

In Sect. 2, we reviewed the magnetic induction equation, giving the solution in the ideal limit, and derived equations for magnetic helicity and magnetic energy. We tested helicity conservation in the ZEUS code by simulating the evolution of the ideal kink mode, and studied fluctuations of helicity in a simulation of a turbulent cloud. The first of these problems is especially suitable for benchmarking numerical codes. That is, it yields a quantitative estimate of the numerical resistivity in the code, from which one can compute the resistive timescale (at least in smooth regions, on large scales) and assess whether it is much longer than other timescales of interest. One can also compare the performance of different codes.

In Sect. 3, we discussed some important dimensionless parameters in astrophysical MHD problems. We verified the large size of the Lundquist number S and magnetic Reynolds number R_m under most conditions, and showed that the magnetic Prandtl number, the ratio of viscous to magnetic diffusivity, is also usually large. This implies that magnetic structure will generally extend to smaller scales than velocity structure. We briefly summarized a variety of numerical techniques from the perspective of the large S limit.

In Sect. 4, we discussed theoretical and numerical results on several basic problems. In Sect. 4.1 we discussed the amplification of magnetic fields by turbulence, and argued that the successful operation of a large scale dynamo depends on fast diffusion of the magnetic field. In Sect. 4.2 we discussed MHD turbulence itself. In Sect. 4.3 we discussed current sheets, which may provide the fast

diffusion necessary for a dynamo, and may be produced in turbulence as a form of intermittency. In Sect. 4.4, we discussed the dynamical effects of turbulence, and its importance in molecular clouds.

In Sect. 5, we discussed ambipolar drift as a mechanism for increasing the magnetic diffusivity in low density, weakly ionized gas. It is now possible to simulate turbulent molecular clouds at a resolution which accounts for all magnetic structure down to the ambipolar diffusion scale. Under certain conditions, however, ambipolar drift can mediate the formation of current sheets, possibly leading to rapid magnetic reconnection.

It is fortunate that a number of different groups are simulating MHD turbulence under astrophysical conditions, using a variety of codes and techniques. There is substantial agreement on a number of issues, such as the short decay time of supersonic turbulence, but disagreement on others, such as the spectrum of strong MHD turbulence. A dedicated effort at benchmarking, in which different groups simulated identical problems and implemented identical diagnostics, would provide some perspective. We suggested in Sect. 4.2 that such an exercise be carried out for strong MHD turbulence.

It is unlikely that it will ever be possible to adequately resolve current sheets and global dynamics in a single simulation. Local studies of current sheets and magnetic reconnection are essential in capturing the physics of these layers. It is equally important to understand how they are embedded in the overall flow, as this constrains their properties, sets the boundary conditions for reconnection, and is necessary for understanding how to parameterize the effects of current sheets in global simulations. The inability to deal adequately with this small scale structure is the greatest present challenge to understanding magnetic field evolution in turbulent flows.

Acknowledgements

We are grateful to E. Falgarone and T. Passot for organizing the conference, and for their hospitality. Our discussions with A. Burkert, M.-M. Mac Low, and J. Maron were especially useful, as were L. Mestel's comments on the manuscript. FH gratefully acknowledges support by the Feodor-Lynen program of the Alexander von Humboldt Foundation. The (U.S.) National Science Foundation provided partial support through grants AST-9800616 and AST-0098701 to the University of Colorado, and through support to the National Center for Atmospheric Research.

References

1. J. Arons, C.E. Max: ApJL **196**, 77 (1975)
2. D.S. Balsara: J. Comp. Phys. **174**, 614 (2001)
3. G. Bateman: *MHD Instabilities*, MIT Press (1978)
4. M.J. Berger, P. Colella: J. Comp. Phys. **82**, 64 (1989)
5. M.J. Berger, J. Oliger: J. Comp. Phys. **53**, 484 (1984)
6. M.J. Berger, R. J. LeVeque: SIAM J. Numer. Anal. **35**, 2298 (1998)

7. P.L. Bhatnagar, E. P. Gross, M. Krook: Phys. Rev. **94**, 511 (1954)
8. A. Bhattacharjee, C.S. Ng: ApJ **548**, 318 (2001)
9. D. Biskamp: *Nonlinear Magnetohydrodynamics*, Cambridge Univ. Press (1993)
10. D. Biskamp, W.-C. Müller: Physics of Plasmas **7**, 4889 (2000)
11. E. Blackman, G.B. Field: ApJ **534**, 984 (2000)
12. S. Bonazzola, E. Falgarone, J. Heyvaerts, M. Perault, J.L. Puget: A&A **172**, 293 (1987)
13. S.I. Braginski: Rev. Plasma Phys. **1** 205 (1965)
14. A. Brandenburg: 'Computational aspects of astrophysical MHD and turbulence' in *The Fluid Mechanics of Astrophysics and Geophysics*, vol 8. ed. A. Ferriz-Mas, astroph/0109497
15. A. Brandenburg, W. Dobler: A&A **369**, 329 (2001)
16. A. Brandenburg, E.G. Zweibel: ApJL **427**, 91 (1994)
17. N.H. Brummell, F. Cattaneo, S.M. Tobias: Fl. Dyn. Res. **28**, 237 (2001)
18. V.M. Canuto: ApJ **428**, 729 (1994)
19. F. Cattaneo: ApJL **515**, 39 (1999)
20. F. Cattaneo, D.W. Hughes, E.-J. Kim: Phys. Rev. Lett. **76**, 2057 (1996)
21. S.E. Caunt, M.J. Korpi: A&A **369**, 706 (2001)
22. S. Chandrasekhar: Proc. R. Soc. London A **210**, 26 (1951)
23. S. Childress, A.D. Gilbert: *Stretch, Twist, Fold: The Fast Dynamo*, (Springer Heidelberg 1995)
24. J. Cho, A. Lazarian, E. Vishniac: ApJL submitted (2002)
25. J. Cho, E. Vishniac: ApJ **539**, 273 (2000)
26. R.H. Cohen, R.M. Kulsrud: 'Nonlinear hydromagnetic wave evolution in the solar wind' in *Solar wind three; Proceedings of the Third Conference, Pacific Grove, Calif., March 25-29, 1974*, p. 382
27. A.W. Cook: J. Comp. Phys. **154**, 117 (1999)
28. R. Courant, K.O. Friedrichs, H. Lewy: Über die partiellen Differenzengleichungen der mathematischen Physik. Math. Ann. **100**, 32 (1928)
29. R.L. Dewar: Phys. Fluids **13**, 2710 (1970)
30. B.T. Draine: MNRAS **220**, 133 (1986)
31. B.T. Draine, C.F. McKee: ARA&A **31**, 373 (1993)
32. B.G. Elmegreen: ApJ **299**, 196 (1985)
33. M. Fatuzzo, F.C. Adams: ApJ **412**, 146 (1993)
34. H. Friedel, R. Grauer, C. Marliani: J. Comp. Phys. **134**, 190 (1997)
35. D.J. Galloway, M.R.E. Proctor: Nature **356**, 691 (1992)
36. K. Galsgaard, Å. Nordlund: unpublished Technical Report of the University of Copenhagen (1995)
37. K. Galsgaard, Å. Nordlund: J. Geophys. Res. **101** 13445 (1996)
38. S. Galtier, S.V. Nazarenko, A.C. Newell, A. Pouquet: ApJL **564**, 49 (2002)
39. C.F. Gammie, E.C. Ostriker: ApJ **466**, 814 (1996)
40. N.Y. Gnedin, A. Ferrara, E.G. Zweibel: ApJ **539**, 505 (2000)
41. S. K. Godunov: Mat. Sb. **47**, 271 (1959)
42. P. Goldreich, J. Kwan: ApJ **189**, 441 (1974)
43. P. Goldreich, S. Sridhar: ApJ **438**, 763 (1995)
44. A.V. Gruzinov, P.H. Diamond: Phys. Rev. Lett. **72**, 1651 (1994)
45. T.S. Hahm, R. Kulsrud: Phys. Fluids **28**, 2412
46. J.F. Hawley, J.M. Stone: Comp. Phys. Comm. **89**, 1 (1995)
47. F. Heitsch, M.-M. Mac Low, R.S. Klessen: ApJ **547**, 280 (2001)
48. F. Heitsch, E.G. Zweibel, M.-M. Mac Low, P.S. Li, M.L. Norman: ApJ **561**, 800 (2001)

49. D.W. Hughes, F. Cattaneo, E. Kim: Phys. Lett. A **223**, 167 (1996)
50. A. Hujeirat, R. Rannacher: New Astr. Rev. **45**, 425 (2001)
51. O.S. Jones, U. Shumlak, D.S. Eberhardt: J. Comp. Phys. **130**, 231 (1997)
52. E. Kim: Phys. Plasmas **7**, 1746 (2000)
53. R.M. Kinney, B. Chandran, S. Cowley, J.C. McWilliams: ApJ **545**, 907 (2000)
54. R.S. Klessen, F. Heitsch, M.-M. Mac Low: ApJ **535**, 887 (2000)
55. R.M. Kulsrud, S. W. Anderson: ApJ **396**, 606 (1992)
56. R.M. Kulsrud, W.P. Pearce: ApJ **156**, 445 (1969)
57. M. Lesieur: *Turbulence in Fluids* (Kluver, Dordrecht 1990)
58. R.J. LeVeque, D. Mihalas, E. Dorfi, E. Müller: "Computational Methods for Astrophysical Fluid Flow" in *27th Saas-Fee Advanced Course Lecture Notes*, eds. O. Steiner and A. Gautschy (Springer 1998)
59. Y. Lithwick, P. Goldreich: ApJ **562**, 279 (2001)
60. D.W. Longcope, S.C. Cowley: Phys. Plasmas **3**, 2885 (1996)
61. D.W. Longcope, H.R. Strauss: ApJ **437**, 851
62. B.C. Low, R. Wolfson: ApJ **324**, 574 (1988)
63. M.-M. Mac Low, M.L. Norman, A. Königl, M. Wardl: ApJ **442**, 726 (1995)
64. M.-M. Mac Low, M.D. Smith: ApJ **491**, 596 (1997)
65. J. Maron, E.G. Blackman: ApJ **566**, L41 (2002)
66. J. Maron, S. Cowley: (astro-ph/0111008)
67. J. Maron, P. Goldreich: ApJ **554**, 1175 (2001)
68. C.E. Martin, J. Heyvaerts, E.R. Priest: A&A **326**, 1176 (1997)
69. C.F. McKee, E.G. Zweibel: ApJ **440**, 686 (1995)
70. C.F. McKee, J.H. Holliman II: ApJ **522**, 313 (1999)
71. C.F. McKee, E.G. Zweibel, A.A. Goodman, C. Heiles: 'Magnetic Fields in Star-Forming Regions – Theory' in *Protostars and Planets III*, eds: E.H. Levy, J.I. Lunine (University of Arizona Press 1993), p. 327
72. L. Mestel & L. Spitzer: MNRAS **116**, 503 (1956)
73. Z. Mikić, D.D. Schnack, G. van Hoven: ApJ **338**, 1148 (1989)
74. H.K Moffatt: *Magnetic field generation in electrically conducting fluids*, Cambridge University Press (1978)
75. H.K. Moffatt: J. Fl. Mech **159**, 359 (1985)
76. P.C. Myers, G.A. Fuller: ApJ **396**, 631 (1992)
77. K. Nakayama: ApJ **556**, 1027 (2001)
78. J. von Neumann, R.D. Richtmyer: J. Appl. Phys. **23**, 232 (1950)
79. C.S. Ng, A. Bhattacharjee: ApJ **465**, 845 (1996)
80. V. Ossenkopf, M.-M. Mac Low: *in preparation*
81. E.C. Ostriker, J.M. Stone, C.F. Gammie: ApJ **546**, 980 (2001)
82. P. Padoan, E.G. Zweibel, Å. Nordlund: ApJ **540**, 332 (2000)
83. E.N. Parker: ApJ **122**, 293 (1955)
84. E.N. Parker: ApJ **174**, 499 (1972)
85. E.N. Parker: *Cosmical Magnetic Fields*, Oxford Univ. Press (1979)
86. E.N. Parker: ApJ **401**, 137 (1992)
87. A. Pouquet: in *Astrophysical Fluid Dynamics*,eds. J.-P. Zahn & J. Zinn-Justin, Elsevier (1993)
88. A. Pouquet, U. Frisch, J Leorat: J. Fluid Mech. **77**, 321 (1976)
89. K.H. Prendergast, K. Xu: J. Comp. Phys. **109**, 53 (1993)
90. R.E. Pudritz: ApJ **350**, 195 (1990)
91. E. Quataert, A. Gruzinov: ApJ **520**, 248 (1999)
92. R.R. Rafikov, R.M. Kulsrud: MNRAS **3**14, 839 (2000)

93. F. J. Robinson, K.L. Chan: MNRAS **321**, 723
94. M.D. Smith, M.-M. Mac Low, F. Heitsch: A&A **362**, 333 (2000)
95. M.D. Smith, M.-M. Mac Low, J.M. Zuev: A&A **356**, 287 (2000)
96. S. Sridhar, P. Goldreich: ApJ **432**, 612 (1994)
97. M. Steenbeck, F. Krause, K.H. Rädler: Z. Naturforsch. A **21**, 369 (1966)
98. J.M. Stone: ApJ **487** 271 (1997)
99. J.M. Stone, M.L. Norman: ApJS **80**, 753 (1992a)
100. J.M. Stone, M.L. Norman: ApJS **80**, 791 (1992b)
101. H.R. Strauss: Phys. Fluids **19**, 134 (1976)
102. S.I. Syrovatskii: Sov. Phys. JETP **33**, 933 (1971)
103. P.A. Sweet: in *Electromagnetic Phenomena in Cosmical Physics*, p. 123, Cambridge Univ. Press (1958)
104. H.-Z. Tang, K. Xu: J. Comp. Phys. **165**, 69 (2000)
105. G. Toth: MNRAS **274** 1002 (1995)
106. J.K. Truelove, R.I. Klein, C.F. McKee et al.: ApJL **489**, 179 (1997)
107. D.A. Uzdensky, R.M. Kulsrud: Phys. Plasmas **7**, 4018 (2000)
108. A.A. van Ballegooijen: ApJ **298**, 421 (1985)
109. H.C. Yee: J. Comp. Phys. **131**, 216 (1997)
110. U. Ziegler: A&A **367**, 170 (2001)
111. B. Zuckerman, N.J. Evans II: ApJ **192**, 149 (1974)
112. E.G. Zweibel, A. Brandenburg: ApJ **478** 563 (1997)
113. E.G. Zweibel, C. Heiles: Nature **385**, 131 (1997)
114. E.G. Zweibel, K. Josafatsson: ApJ **270** 511 (1983)
115. E.G. Zweibel, M.R.E. Proctor: in *Topological Fluid Mechanics*, ed. H.K. Moffatt & A. Tsinober, Cambridge Univ. Press, 187 (1990)

Adaptive Mesh Refinement in MHD Modeling. Realization, Tests and Application

Udo Ziegler

Astrophysikalisches Institut Potsdam, An der Sternwarte 16, D-14482 Potsdam, Germany

Abstract. The technique of adaptive mesh refinement (AMR) as it is realized in the magnetohydrodynamics code NIRVANA is described. Basic principles and algorithms of the operator-split solver of NIRVANA are discussed. Two test problems – the Orszag-Tang vortex problem and the problem of a magnetic flux torus in a stagnation flow – are presented which have been performed to check the implementation of AMR. As a first astrophysical application of the AMR code the three-dimensional dynamical evolution of a thin magnetic flux tube embedded in a stratified, rotating medium is studied. The AMR result is compared with highly simplified calculations using the thin flux tube approximation.

1 Introduction

Many interesting phenomena in plasma fluid dynamics can be treated within magnetohydrodynamic theory (MHD). Applications can be found in thermonuclear fusion, magnetospheric- and solar physics, geophysics, and in astrophysical research. By virtue of its non-linear character the MHD equations are too complex to be solved by pencil in general. To avoid too restrictive assumptions typical for analytical investigations one relies on numerical simulations. The computational modeling of MHD systems, however, can be a great challenge, too, especially for astrophysical fluid flows. This has several reasons. Many physical processes with different time scales may be involved requiring different numerical techniques. The diversity in time scales may be accompanied by separate spatial scales to be captured in computer models. For instance, the nonlinearities in the equations permit the existence of steep gradients such as shocks, contact discontinuities, shear layers, and current sheets which have to be represented on the grid in an accurate manner.

This paper is devoted to the numerical code NIRVANA, a code for the simulation of astrophysical, non-relativistic fluid flows. NIRVANA has been designed with the claim to model magnetohydrodynamical processes for a wide variety of astrophysical systems from planets to the intergalactic medium. It is a general-purpose code for simulations in two- or three space dimensions including a description of the effects of heat conduction and selfgravity. Some typical applications of the code are thermal convection in stars, accretion flows around compact objects, the generation and evolution of jets or the simulation of dynamical processes in the interstellar medium like the formation of protostellar objects in dense molecular clouds. One important idea behind the philosophy

E. Falgarone and T. Passot (Eds.): LNP 614, pp. 127–151, 2003.
© Springer-Verlag Berlin Heidelberg 2003

of NIRVANA is to render easy upgrades of the numerical method by the need to consider new physical effects. One further aspect dropped from the beginning in the development of the code is flexibility in order to allow, for example, simulations in the most important orthogonal (Cartesian, cylindrical, spherical) coordinate systems or the use of different gas equations of state. To meet that requirements NIRVANA is based on a fractional-step approach and uses a mixed finite-difference/finite-volume scheme on a staggered mesh.

One essential feature of NIRVANA which renders it prominent from many other codes is the possibility of adaptive mesh refinement (AMR). The need for AMR is indispensable in multi-scale problems since, due to limited computer resources, it is impossible to span the total integration domain with just one fine grid. In an AMR procedure grid points are clustered only in spatial regions which need a higher resolution while grid points are less compactly distributed elsewhere. AMR thus tries to achieve a given accuracy at a minimum of computational costs. In time-dependent problems mesh refinement must be flexible enough to react quickly upon the appearance of new structures in the solution and to capture all features of interest. There exists different approaches to do so. AMR methods of Lagrangian type are characterized by the spatial redistribution of a fixed number of grid points moving with the flow ([9], [10], [16]). Lagrangian methods are best for 1D problems but less qualified in 2D and 3D because of grid distortion effects which require frequent mesh regularizations, thereby, introducing an uncertain amount of artificial diffusion. A completely different ansatz makes use of block-structured finer subgrids overlaid on a basic grid. This approach has been worked out in detail in [5]. The subgrids can recursively be nested up to a prescribed refinement depth in order to reach the desired accuracy. NIRVANA implements that type of mesh refinement originally designed for the Euler equations of gas dynamics and later extended to similar hyperbolic PDE's as described in [6], [7], [36], [4].

The presentation of NIRVANA begins in Sect. 2 with a brief description of the numerical principles it is based upon. In Sect. 3 the idea behind block-structure AMR is illustrated and its actual implementation in NIRVANA is discussed. Section 4 presents some code tests in conjunction with AMR and studies the problem of the rise of thin magnetic flux tubes in vertically stratified disks as an astrophysical application of relevance.

2 Numerical Principles

In the MHD regime a viscous, resistive and thermal conducting plasma is described by the equations

$$\partial_t \varrho = -\nabla \cdot (\varrho \boldsymbol{v}) \,, \tag{1}$$

$$\partial_t (\varrho \boldsymbol{v}) = -\nabla \cdot (\varrho \boldsymbol{v}\boldsymbol{v}) - \nabla p + \nabla \cdot \tau + \frac{1}{\mu}(\nabla \times \boldsymbol{B}) \times \boldsymbol{B} - \varrho \nabla \Phi \,, \tag{2}$$

$$\partial_t \boldsymbol{B} = \nabla \times (\boldsymbol{v} \times \boldsymbol{B} - \eta \nabla \times \boldsymbol{B}) \,, \tag{3}$$

$$\partial_t e = -\nabla \cdot (e\boldsymbol{v}) - p \nabla \cdot \boldsymbol{v} + Q_{\rm vis} + \nabla \cdot (\kappa \nabla T) + \frac{\eta}{\mu} |\nabla \times \boldsymbol{B}|^2 \,. \tag{4}$$

The dependent variables are the mass density ϱ, pressure p, thermal energy density e, temperature T, velocity $\boldsymbol{v}$, magnetic field $\boldsymbol{B}$ and gravitational potential Φ. The remaining quantities in the above equations are the viscous stress tensor τ given by

$$\tau_{ij} = \nu\varrho\left((\nabla\boldsymbol{v})_{ij} + (\nabla\boldsymbol{v})_{ji} - \frac{2}{3}\nabla\cdot\boldsymbol{v}\delta_{ij}\right) \tag{5}$$

where ν denotes the kinematic viscosity coefficient, the viscous heating term

$$Q_{\mathrm{vis}} = \sum_{ij}\tau_{ij}(\nabla\boldsymbol{v})_{ij}\ , \tag{6}$$

the magnetic permeability μ, the thermal conductivity coefficient κ, and the magnetic diffusivity $\eta = 1/\sigma\mu$ where σ is the electrical conductivity. The thermodynamic variables are related through the ideal gas equation of state

$$p = \frac{k}{\overline{\mu}m_{\mathrm{u}}}\varrho T = (\gamma-1)e \tag{7}$$

where k is Boltzmann's constant, m_{u} the atomic mass unit, $\overline{\mu}$ the mean molecular weight and $\gamma = c_p/c_V$ the ratio of specific heats, c_p and c_V. The gravitational potential is determined from the Poisson equation

$$\triangle\Phi = 4\pi G\varrho \tag{8}$$

where G is the gravitational constant.

NIRVANA is a grid-based computer code which solves the above PDE's on a logically-rectangular domain. The numerical algorithms are implemented in a covariant fashion and allow the use of Cartesian, cylindrical or spherical coordinates. Several properties of NIRVANA make it a well suited code for the simulation of astrophysical fluid flows in three space dimensions. Most parts of the solver are time-explicit and are based on robust and simple algorithms which make the code comparatively fast. Although more specialized and, hence, more accurate methods exist (e.g. Godunov-type schemes like PPM, [44]), such schemes are more difficult to code and may lack in flexibility when new pieces of physics are to be added to the equations.

NIRVANA applies the technique of operator-splitting which breaks the numerical solution of the PDE's into parts. Hereby the full solution after a time-step $\delta t = t_{n+1} - t_n$ is found step by step where n denotes the time level index. Whenever possible the result of a preceding step is used in the following step. The solution procedure for the time-dependent equations (1)–(4) takes the form

$$\begin{aligned}
(U^{(1)} - U^{(0)})/\delta t &= L_1[U^{(0)}, U^{(1)}]\ , \\
(U^{(2)} - U^{(1)})/\delta t &= L_2[U^{(1)}, U^{(2)}]\ , \\
&\vdots \\
(U^{(M)} - U^{(M-1)})/\delta t &= L_M[U^{(M-1)}, U^{(M)}]
\end{aligned} \tag{9}$$

where $U^{(m)}$ denotes the numerical approximation to the true solution $u = (\varrho, e, \boldsymbol{v}, \boldsymbol{B})$ after the m-th substep with $U^{(0)} \equiv U^n$, $U^{(M)} \equiv U^{n+1}$. L_m, $m = 1, ..., M$, are discretized representations of the operators $\mathcal{L}_m$ which make up the full nonlinear operator $\mathcal{L} = \sum_m \mathcal{L}_m$ of the system written compactly as $\partial_t u = \mathcal{L}[u]$. If substep m is time-explicit, the r.h.s. depends only on the state variable $U^{(m-1)}$ of the preceding step. Except for the heat conduction part in the energy balance equation, $\partial_t e = \nabla \cdot (\kappa \nabla T)$ which is solved with an ADI method [2], all substeps are time-explicit. The same ADI method used to solve the heat transport equation is, applied in an iterative manner, used to solve the Poisson equation for the gravitational potential. In the present stage of development the ADI solver does not work in conjunction with AMR. However, work is on the way to do so. First results in this direction are promising. Multigrid methods would be an alternative to solve the Poisson equation on block-structured meshes. However, traversing the entire multigrid hierachy on a deeply nested mesh configuration can significantly degrade processing efficiency and, therefore, may become prohibitive for complex applications ([43]).

The time-explicit parts of the code limit the time-step for evolving the dynamical equations to meet the stability criterion of Courant–Friedrichs–Lewy. It takes the form

$$\delta t < C \cdot \min_{\text{all cells}} \min \left\{ \frac{\delta s}{|\boldsymbol{v}| + c_f}, \frac{(\delta s)^2}{4\nu}, \frac{(\delta s)^2}{4\eta} \right\} \tag{10}$$

with $\delta s = \min\{\delta x, \delta y, \delta z\}$, δx, δy and δz the grid spacings in x-,y-,z-direction (Cartesian notation is used here and in the rest of this paragraph), and $c_f = (c_S^2 + c_A^2)^{1/2}$ the fast magnetosonic speed given by the sound speed $c_S = (\gamma p/\varrho)^{1/2}$ and Alfvén speed $c_A = |\boldsymbol{B}|/(\mu\varrho)^{1/2}$. Condition (10) states that information cannot travel a distance more than the dimension of a numerical cell within one time-step. The quantity C is a safety factor smaller than unity (typically $C = 0.5$).

The above multistep procedure can mainly be grouped into two steps, the advection step to which belongs the advection of hydrodynamic quantities as well as the transport of magnetic field by the fluid flow, and the source step. Latter consists of several individual steps based on the method of finite-differencing. For the advection step a second-order finite-volume approach is used. The hydrodynamic advection part can be written in the compact form

$$\partial_t u + \nabla \cdot \boldsymbol{f}(u) = 0 \tag{11}$$

where u now stands for one of the advected quantities $\varrho, e, \varrho v_x, \varrho v_y, \varrho v_z$ and $\boldsymbol{f}(u) = u\boldsymbol{v}$ is the flux density function. Integrating this equation over a cell volume $\mathcal{V}$ and over time from t_n to t_{n+1} gives

$$\int_{\mathcal{V}} u(x,y,z,t_{n+1})\, dV - \int_{\mathcal{V}} u(x,y,z,t_n)\, dV + \int_{t_n}^{t_{n+1}} \left[\oint_{\partial\mathcal{V}} \boldsymbol{f}(u) \cdot d\boldsymbol{A} \right] dt = 0\,. \tag{12}$$

This integral identity is the starting point for the derivation of a conservative scheme. It simply states that the change of the cell-integrated quantity u during

the time-step is given by the flux of that quantity through the cell surface. The accuracy of the scheme is determined by the order of approximation of the flux function. In NIRVANA van Leer's monotonic upwind scheme for conservation laws is used as described in [42]. It is stable, second-order accurate and has the property of total variation diminishing (in short TVD) which means that no spurious oscillations are numerically generated during the advection process. This does not mean, of course, that the full code is TVD and, indeed, it is not.

Similarly, the integral form of (3) is given by

$$\int\limits_{\mathcal{A}} \boldsymbol{B}(x,y,z,t_{n+1})\,d\boldsymbol{A} - \int\limits_{\mathcal{A}} \boldsymbol{B}(x,y,z,t_n)\,d\boldsymbol{A} - \int\limits_{t_n}^{t_{n+1}} \left[\oint\limits_{\partial\mathcal{A}} \boldsymbol{E}\cdot d\boldsymbol{s} \right] dt = 0 \qquad (13)$$

where $\mathcal{A}$ stands for any of the three coordinate faces of a cell and $\boldsymbol{E} = \boldsymbol{v} \times \boldsymbol{B} - \eta\nabla \times \boldsymbol{B}$ is the usual electromotive force (EMF) modified by the diffusivity term. The EMF takes on the role of the flux density function in fluid advection. On stability grounds the procedure of time averaging in the last term in (13) must ensure an accurate representation of Alfvén waves which are often a source of numerical instability because of their incompressible nature. The idea, originally described in [37] for the 2D situation, is to estimate the EMF (without diffusion term which is treated separately) from a set of simplified equations describing Alfvén wave propagation. Assuming one-dimensionality and incompressibility such shear waves obey

$$\begin{aligned} \partial_t v_i &= \frac{B_j}{\mu\rho}\partial_j B_i - v_j \partial_j v_i \,, \\ \partial_t B_i &= B_j \partial_j v_i - v_j \partial_j B_i \,, \quad i,j = x,y,z \quad i \neq j \,. \end{aligned} \qquad (14)$$

A total of six sets of such coupled equations exist to cover the various combinations of polarization direction and propagation direction both of which are orthogonal to each other. These sets of equations are used to evaluate $\boldsymbol{v} \times \boldsymbol{B}$ at intermediate time level $t_n + \delta t/2$ to be used then in the transport scheme. A more detailed description of the method can be found in [47].

3 Adaptive Mesh Refinement

AMR in NIRVANA is based on a block-structured approach. The principle is illustrated in Fig. 1. Assume a main (Cartesian, cylindrical or spherical) grid with equidistant mesh spacing in each coordinate direction. In the following the main grid is referred to also as base level. Refinements on the base level are carried out according to a specific criterion described in detail later. If the criterion for refinement is fulfilled, the base level is overlaid by finer grid cells grouped into rectangular patches with dimension of 4 cells per coordinate direction i.e. 4^2 cells in 2D and 4^3 cells in 3D. The mesh spacing of this finer subgrid is half that of the main grid corresponding to a refinement ratio of 2. In principle, any integer number would be conceivable. However, as pointed out in [22] a

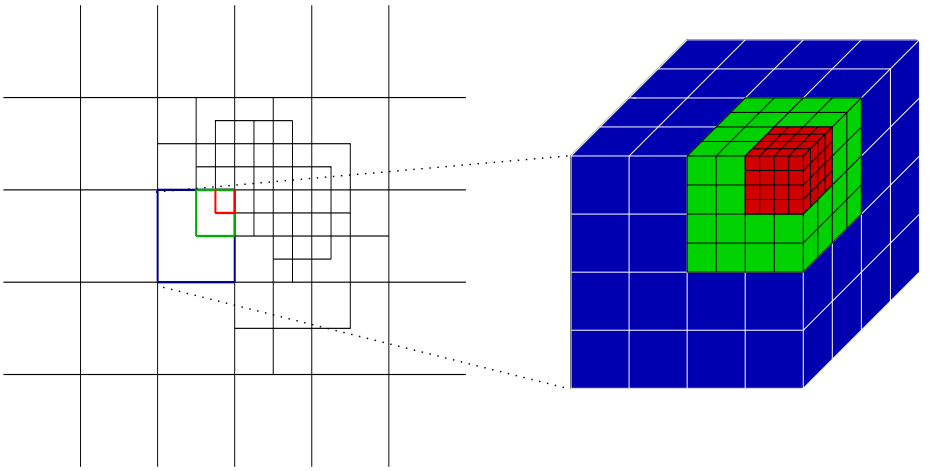


Fig. 1. Example of a hierarchical grid system made up of the main grid and a set of subgrids (patches) with finer mesh spacing. Patches can recursively be nested to achieve a prescribed resolution

smaller refinement ratio is in favor of a slighter variation of numerical errors at refinement interfaces. Mesh refinement is multi-directional, i.e. is performed simultaneously in each coordinate direction, to avoid unphysical artifacts due to an unequal discretization error in the different directions. Patches can recursively be nested until a prescribed resolution is achieved or a maximum level of nesting is reached. One finally ends up with a hierarchical grid system comprised out of the main grid and a set of patches with different resolution. All patches sharing the same resolution make up a refinement level. Starting with zero for the base level we will denote refinement levels by $\ell = 1, ..., L$ where L is the maximum refinement level at a time. Note that L in general varies with time limited by some maximum value prescribed by the user.

The small size of the patches allows a flexible mesh adaptation. This is at the expense of an increased amount of computational overhead compared to the use of larger patches. However, the increase in overhead is partly balanced or even overbalanced by two effects. First, larger patches may cover regions which actually need no refinement but must be taken along. Second, there is no need to solve the equations on patches which are completely covered by finer-grid patches, thus, speeding up the code. Unfortunately, too small patches are not well suited for supercomputers with vector facility. This problem will be addressed in near future by a dynamic clustering algorithm: neighboring patches are temporarily sticked together forming larger units which speeds up the integration process without loss in flexibility of grid adaptation.

When applied to explicit numerical schemes, the presented AMR technique implies a refinement in both space and time. This is because the CFL condition requires that the time-step for level $\ell + 1$ is at least half the time-step for the coarser level ℓ. To advance the solution on the hierarchical grid system from time

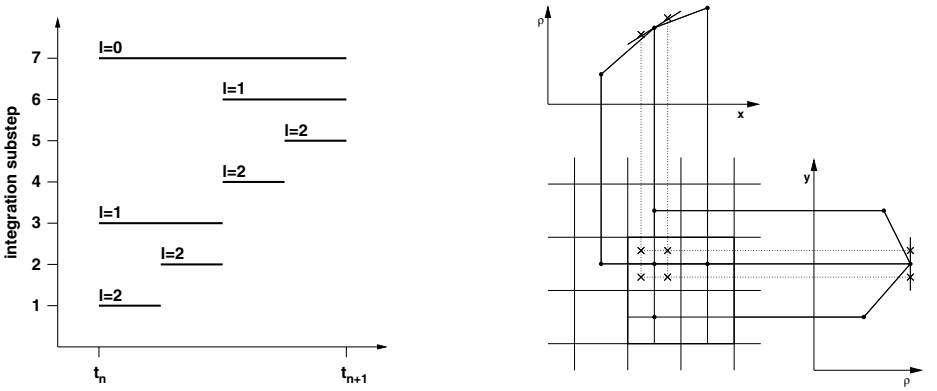


Fig. 2. Left: integration sequence for a hierarchical grid system consisting of 3 levels. Right: the principle of MPLR. Coarse cell values are denoted by filled circles, fine cell values to be determined through MPLR by crosses. If the coarse mesh density violates monotonicity in some coordinate direction the corresponding slope is set to zero

t_n to time t_{n+1} a series of integrations of the equations on the different levels is necessary. There are two principle possibilities: one could start integrating the finest level first and continue at coarser levels to end up finally on the base level, or alternatively, the base level could be integrated first advancing the finer levels step by step. The first method is used in NIRVANA. The resulting integration cycle for a 3-level system consisting of the base level and two refined levels is illustrated in Fig. 2 (left panel). The sequence of integrations for this example is $\ell = 2, 2, 1, 2, 2, 1, 0$.

Integrating the equations on a certain level means integrating the equations on each individual patch which belongs to that level. In contrast to the main grid integration, solving the MHD equations on a patch implies the specification of special boundary conditions. Generally, the solver requires two ghost zones on each side per coordinate direction. Three different possibilities exist where boundary values can come from: (i) If the patch lies on a physical boundary, corresponding boundary conditions as prescribed for the main grid are used; (ii) if the patch lies in the interior of the domain and there is no neighboring patch of the same level, boundary values are interpolated from the coarser level. Note that the data available from the coarser level can lag in time by one fine level time-step (e.g. in Fig. 2 substep 5 may need data from substep 3). Better results may be obtained by time-averaged boundary values. However, this requires the inverse integration cycle $0 \longrightarrow L$ (not realized in NIRVANA) and the simultaneous storage of two time levels per refinement level doubling the required memory; (iii) if there is a neighboring patch of the same refinement level boundary values are copied from this patch. Since patches of a refinement level are integrated in an arbitrary sequential order some care is necessary to avoid assigning boundary values from an already integrated patch. This would be inconsistent with the advection scheme violating the conservation property and generating numerical noise. To avoid that, before integration of a refinement level starts, a copy of this

refinement level is created which serves then as a reservoir for boundary data. After integration the refinement level clone is destroyed.

Further modifications are necessary to adjust the solutions of any two adjacent refinement levels. Communication between levels occurs in two steps. To retain the properties of mass conservation and solenoidality of the magnetic field, fulfilled by the single-grid solver, corrections of the corresponding fluxes, namely the mass flux and EMF, are necessary at refinement interfaces. Without such corrections the interface fluxes obtained in the fine-level integration and coarse-level integration are inconsistent. Since the solution on the finer level is the more accurate one and this level is integrated before the coarser level, it is appropriate to replace the coarse-level fluxes by the known fine-level fluxes. Note once more that this modification takes place only at refinement interfaces. Moreover, after the coarser level has been integrated to the same point in time as the finer level, the coarse level solution is replaced by the solution of the overlaying finer level. In case of the gas density, thermal energy density and magnetic field this updating process is performed under the constraint of mass-, thermal energy- and magnetic flux conservation. The whole procedure applies recursively on the hierarchical grid system which means that mass is conserved and that the magnetic field is evolved divergence-free. This is not true for the total energy and for the momenta since the numerical scheme is a priori not conservative in these quantities.

If a patch is added to the grid system, each numerical cell in the patch has to be assigned by initial values for the basic variables. Variable assignments must be done in a fashion, of course, not to violate mass conservation and the divergence-free constraint of the magnetic field. As a rule, initial values are computed through interpolation from the underlying coarser level. The density (and the thermal energy density as well) is obtained by a procedure called monotonized piecewise linear reconstruction (MPLR). MPLR in each underlaying coarse grid cell reconstructs the density field to first order out of the nearest cell neighbors. Then, the density on the finer mesh of the new patch can be computed from this piecewise linearized function. The situation is illustrated in Fig. 2 (right panel). As can be shown easily MPLR conserves mass. However, a complication occurs in case of non-Cartesian coordinates as pointed out in [28]. In this case a modified version of MPLR is used. MPLR is also applied for certain vector components whereas other vector components are obtained by simple linear interpolation. To guarantee that no magnetic divergence is introduced, in addition, the discretized version of $\nabla \cdot \boldsymbol{B} = 0$ is used directly to calculate magnetic field components.

One of the key parts of a functioning AMR algorithm is the choice of the refinement criterion. In finite difference/finite volume methods usually one of the two following procedures are used. The first one is based on a Richardson-type 'error' estimation by comparing numerical solutions on adjacent refinement levels. This criterion works well in smooth regions of the flow but is less suited near shocks. This is because shocks attract a roughly fixed number of grid points independent of the level of refinement i.e. the detected error would approximately

be the same irrespective of the level. Thus, even weak shocks would be maximally resolved. Alternatively, mesh refinement can be regulated by identifying steep gradients in the variables or functionals of them. The refinement criterion in NIRVANA is based on the latter approach. For refinement on level $\ell < L$ the criterion takes the form

$$\frac{|\delta f|}{|f| + f_{\mathrm{ref}}} \frac{\delta s^{(\ell)}}{\delta s^{(0)}} > \varepsilon, \tag{15}$$

where $f = \{\varrho, e, v_i, B_i\}$, $f_{\mathrm{ref}} \geq 0$ is a reference value chosen zero for the scalars and positive for the vector quantities, and ε is a threshold parameter. Generally, the value for ε depends somewhat on the problem to be solved and must be determined experimentally. Examples are given in the test calculations. A new patch is inserted at a certain position whenever (15) is fulfilled for *any* f in a buffer zone around that position. A patch is released if (15) with ε replaced by 0.7ε is violated for *all* f. The factor 0.7 softens mesh reorganization in that it prohibits that patches just created are being immediately destroyed again. The refinement criterion is checked within an integration cycle several times. This permits the grid system to react quickly upon temporal changes in the solution and ensures that structures with steep gradients are captured at all times. In future applications of the code to star formation problems the numerical Jeans condition, $\lambda_{\mathrm{Jeans}}/4 < \delta s^{(\ell)}$ where λ_{Jeans} is the local Jeans length ([39]), has to be incorporated in addition to avoid artificial fragmentation in self-gravitating flows ([40], [23]).

4 Code Tests and Application

4.1 AMR-MHD Test Problems

A variety of test problems with different degree of complexity, geometry, and spatial dimension have been performed to check AMR in NIRVANA. Most of these test problems are described in [46], [47] and [48]. Here, two MHD problems are presented one in 2D, the so-called Orszag-Tang vortex problem, and one in 3D, the evolution of a thin magnetic flux ring in a stagnation flow.

Orszag-Tang Vortex. The first test example describes a 2D MHD vortex problem which is often quoted as a simple model to study the evolution of MHD turbulence ([29], [31], [45]). The ideal MHD equations with a ratio of specific heats of $\gamma = 5/3$ are solved. The initial configuration is defined on a rectangular domain given by $(x, y) = [0, 2\pi] \times [0, 2\pi]$ with periodic boundary conditions in both coordinate directions. The domain is spanned by a main grid of 50×50 mesh points. The choice of AMR parameters is such that at most three levels of refinement ($L_{\max} = 3$) are allowed which gives a maximal improvement in resolution of a factor of 8 relative to the main grid. The threshold parameter in the refinement criterion (15) is set to $\varepsilon = 0.06$. The initial vortex structure takes

the form

$$\boldsymbol{v}(x,y) = (-\sin y, \sin x, 0) \ , \tag{16}$$

$$\boldsymbol{B}(x,y) = (-\sin y, \sin 2x, 0) \tag{17}$$

which is a pure planar configuration with no z-component. As illustrated in Fig. 3a the flow and the magnetic field contain X-points. Note in this context the different wavelength of B_x and B_y. The initial physical state is completed by specifying the density $\varrho = 25/9$ and thermal energy density $e = 5/2$. The system is unstable and develops into turbulence. The transition towards turbulence is supported by random disturbances of amplitude 10^{-6} in the velocity field.

As expected the flow becomes complex as time proceeds. The main features of the solution are steep gradients in the thermodynamic variables and a current sheet in the center of the domain. The current sheet results from the encounter of the oppositely oriented magnetic field vortices. The solution is depicted in Fig. 3b,d showing a contour plot of the density (plot (b)) and the resulting magnetic field topology (plot (d)) after 300 time-steps corresponding to the dimensionless time $t = 3$. The current sheet is indirectly represented in the density plot by the central, horizontally aligned filament indicating a peak in the density. Figure 3e illustrates the distribution of grid points at that time. Grid points are strongly clustered in the dark regions of the plot where they can no longer be distinguished from each other. As expected, the grid point distribution reflects to large extent the solution of the problem indicating the position of sharp features.

The AMR result is checked by comparing it against a non-AMR calculation using a high-resolution grid of size 400×400 (Fig. 3c). To demonstrate the importance of having a sufficient resolution for this problem a low-resolution run of 50×50 grid points has also been performed (Fig. 3f). The AMR simulation and high-resolution simulation yield comparable results. There are only minor differences between both solutions demonstrating that AMR works well. The gain in performance obtained with AMR is approximately 5.3 which is quite convincing bearing in mind that AMR provides results of comparable quality. Obviously, the AMR result is much better than the low-resolution counterpart. In the latter, gradients in the density are much weaker and, in particular, the current sheet in the center is totally missing. It should be stressed that no analytical solution is known to this problem. Thus, strictly speaking, correctness of the results is not proven. However, work by other researches using different numerical methods confirm our findings (see e.g. [38]).

Magnetic Flux Ring in a Stagnation Flow. The next test problem is a highly simplified model to mimic the process of magnetic reconnection. We consider a thin magnetic flux ring placed in a prescribed two-dimensional stagnation flow. Note that, although the flow is 2D, the problem is 3D. The time evolution of the magnetic field is investigated solving the ideal induction equation i.e. solving (3) with the diffusivity coefficient η set to zero. The neglect of the diffusion term implies that the resulting reconnection effect has no physical origin but is purely numerical: magnetic field lines with opposite orientation but equal field

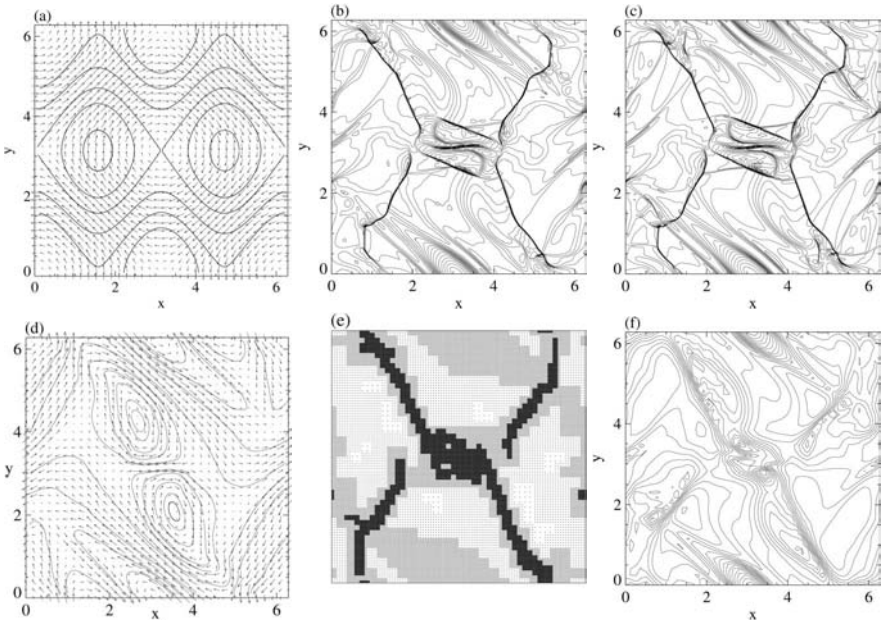


Fig. 3. Initial magnetic field lines and velocity vectors (**a**), contour plot of the density at $t = 3$ (**b**), density contours from a 1-grid high-resolution run (**c**), magnetic field lines and $\boldsymbol{B}$-vectors at $t = 3$ (**d**), grid point distribution at $t = 3$ (**e**), and density contours from a 1-grid low-resolution run (**f**)

strength cancel each other if advected into a grid cell. This mechanism more or less mimics a diffusive process with an unphysical and unknown diffusivity. Nevertheless, it turns out a fascinating problem to check AMR.

The flux ring is assumed to have circular cross-section with radius r, ring radius R and its center is placed at position $(0, 0, z_0)$ in a rectangular Cartesian domain $(x, y, z) \in [-50, 50] \times [-60, 60] \times [0, 60]$. The ring axis lies parallel to the z-axis. The numerical method expects the initial magnetic field to be divergence-free. This can be achieved by constructing an appropriate vector potential. For a flux ring with uniform field embedded in a field-free environment the vector potential is given by

$$\boldsymbol{A} = \begin{cases} \tilde{\boldsymbol{A}} & T(x,y,z) < r \\ \tilde{\boldsymbol{A}} \frac{r^2}{T^2} & \text{otherwise} \end{cases} \tag{18}$$

where

$$\tilde{\boldsymbol{A}} = \left(\frac{(z - z_0) \cdot x}{2\sqrt{x^2 + y^2}}, \frac{(z - z_0) \cdot y}{2\sqrt{x^2 + y^2}}, -\frac{\sqrt{x^2 + y^2} - R}{2} \right) \tag{19}$$

and

$$T(x,y,z) = \left[\left(\sqrt{x^2 + y^2} - R \right)^2 + (z - z_0)^2 \right]^{1/2} . \tag{20}$$

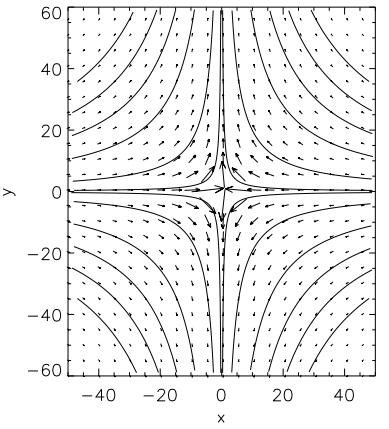


Fig. 4. Structure of the 2D stagnation flow represented by streamlines and velocity arrows

The meaning of the function T is that the level surface $T(x, y, z) = r$ is the surface of the flux ring. The magnetic field, $\boldsymbol{B} = \nabla \times \boldsymbol{A}$, takes the form

$$\boldsymbol{B} = \begin{cases} \left(\frac{-y}{\sqrt{x^2+y^2}}, \frac{x}{\sqrt{x^2+y^2}}, 0 \right) & T(x, y, z) < r \\ \mathbf{0} & \text{otherwise} \end{cases} \tag{21}$$

i.e. a toroidal ring with constant magnetic flux $|\boldsymbol{B}| = 1$ inside the ring embedded in a field-free outer medium as desired. For the following simulation parameters of the ring are chosen as $R = 25$, $r = 1$ and $z_0 = 30$, respectively. The initial configuration is shown in Fig. 5.

The stagnation flow is given by

$$\boldsymbol{v} = \left(\frac{-x}{x^2 + y^2}, \frac{y}{x^2 + y^2}, 0 \right) \tag{22}$$

which is z-independent (see Fig. 4). The effect of this flow on the magnetic flux ring is to squeeze it along the x-axis and to stretch it along the y-axis. Analytically, the velocity goes to infinity as $(x, y) \longrightarrow (0, 0)$. In the numerical realization, however, this is not the case because velocity is not discretized in the coordinate origin but nearby. Hence, in the vicinity of the origin the velocity is very large but finite. This fact implies, however, a rather small CFL time-step to be used. The structure of the velocity profile (22) ensures that 'reconnection' takes place in a finite time. For a hypothetical flowfield with $|\mathbf{v}| \longrightarrow 0$ as $(x, y) \longrightarrow (0, 0)$ there would be no reconnection event in finite time in the limit of vanishing diffusivity because magnetic field lines meet only at $t \longrightarrow \infty$.

The numerical setup of the problem is as follows. The computational domain is covered by a main grid of size $100 \times 120 \times 60$ mesh points. An equal grid spacing is assumed in each coordinate direction. Initially, the flux ring is captured

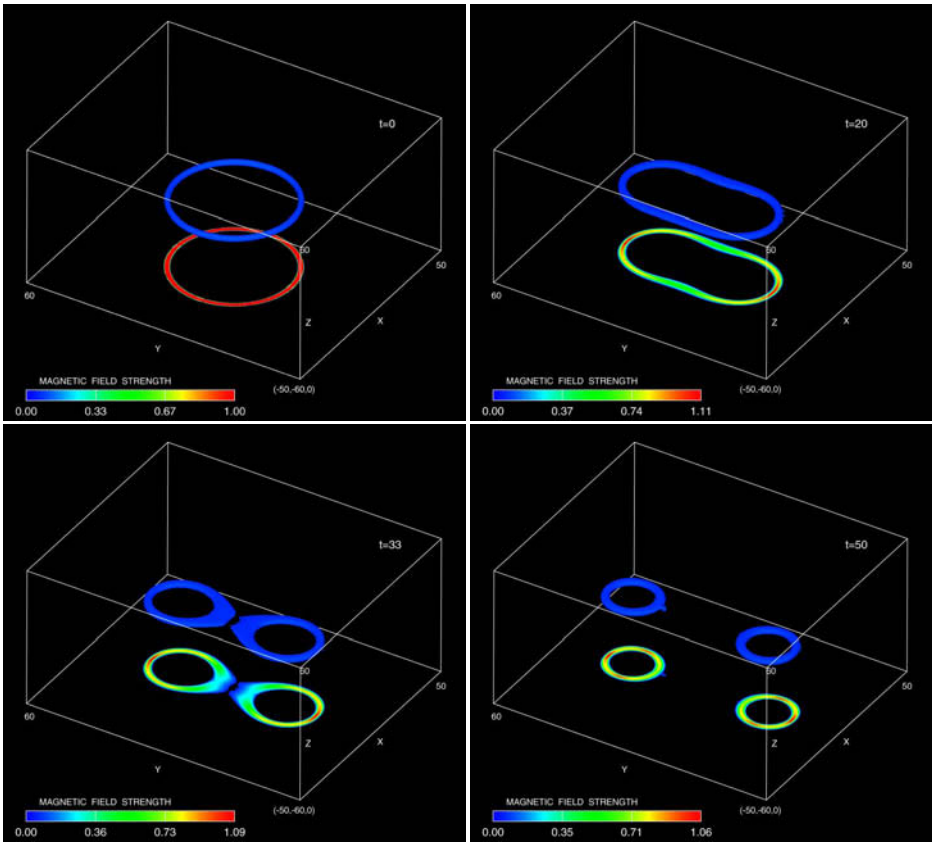


Fig. 5. Isosurface $|\mathbf{B}| = 0.05$ and contour plot (parallel through the fluxtube at $z = z_0$ and projected onto the xy-plane) of the magnetic field strength at times $t = 0$ (initial configuration) and $t = 20, 33, 50$

by refinement level $\ell = L_{\max} = 2$. Thus, the flux ring is four times better resolved than the surrounding field-free medium. On the adaptive mesh the cross-sectional diameter of the flux ring is represented by roughly 8 grid points i.e. only 2 grid points on the main grid. AMR parameters are chosen such that the magnetic field structure is fully resolved in the course of evolution. Because of this one expects no significant differences to the solution of a hypothetical high-resolution simulation with $400 \times 480 \times 240$ grid points. Unfortunately, due to limited resources, such a comparison computation could not be carried out. Since, furthermore, no analytical solution for this problem exists the reliability of AMR results has to be founded on plausibility arguments. The evolution of the flux ring is followed up to a dimensionless time $t = 50$ which corresponds to a CPU execution time of 43.3 hours on a HP workstation. The maximum computer memory accessed during the run was 212 Mbyte. These values were

81.8 days respective 4.7 GBytes for the hypothetical $400 \times 480 \times 240$ case. Hence, the speedup with AMR is ≈ 45.

The result is illustrated in Fig. 5. A sequence of snapshots at different times $t = 0, 20, 33, 50$ is shown. At $t = 20$, the initially toroidal flux ring is already significantly distorted and has transformed into an elongated structure in y-direction. The magnetic field strength has decreased in the stretched parts and increased in the squeezed parts. As time proceeds, tube sections of oppositely oriented field lines come closer. If they have approached to within one grid cell, numerical reconnection sets in and continues up to the breakup of incoming magnetic flux. The final state consists of two separate flux rings moving apart from each other along the y-axis. The developed new flux rings have smaller major radius compared to the initial flux ring and they are slightly thicker in diameter. The latter effect is likely of numerical nature caused by the inadequacy of the advection scheme to perfectly transport the magnetic discontinuity at the surface of the flux tube but smoothing it out over the cross-section.

4.2 Application: Rise of Isolated Flux Tubes in Disks

Model Description. As an example of application of the AMR code with relevance in astrophysical research the dynamical evolution of a thin magnetic flux tube embedded in a vertically stratified, non-selfgravitating disk is considered. Here, thin means that the diameter of the flux tube is small compared to the scale height of the stratification. The motivation for such a study is as follows. There is evidence from both analytical work ([24]) and numerical simulations ([20], [49]) that turbulent plasmas have the tendency to organize the magnetic field into ropes or similar structures. Given this intermittent nature of the magnetic field the problem of magnetic field generation in disks by a dynamo process can be approximated in its simplest form by assuming idealized magnetic flux concentrations in form of thin flux tubes. To let such a flux tube dynamo work efficiently, however, some mechanism must be present to prevent large magnetic flux loss due to buoyancy effects as predicted by 1D flux tube models. It is therefore of very interest to study the rising behavior of a 3D flux tube using AMR techniques.

The problem is simplified by employing the local box approximation. One considers a rectangular patch of a disk located at a fiducial radius R_0. The ideal MHD equations are solved in a Cartesian coordinate system attached to the box center and corotating with the local angular velocity $\boldsymbol{\Omega}_0 = \Omega(R_0)\hat{\boldsymbol{z}}$ of the disk. The coordinate axes are oriented in a sense that, locally, the unit vector $\hat{\boldsymbol{x}}$ points in radial direction, $\hat{\boldsymbol{y}}$ in azimuthal direction and $\hat{\boldsymbol{z}}$ in vertical direction. Figure 6 schematically illustrates the local model in which the flux tube is a straight cylinder.

An isothermal equation of state, $p = c_S^2 \varrho$, is assumed where c_S is the sound speed. Equation (4) is then redundant. The momentum equation has to be supplemented by the centrifugal force, $\varrho\boldsymbol{\Omega}_0 \times (\boldsymbol{\Omega}_0 \times \boldsymbol{x})$, and Coriolis force, $-2\varrho\boldsymbol{\Omega}_0 \times \boldsymbol{v}$, which is due to the corotating frame of reference, and supplemented by an explicit acceleration $\varrho\boldsymbol{g}$ (e.g. through a massive central object) to force rotation

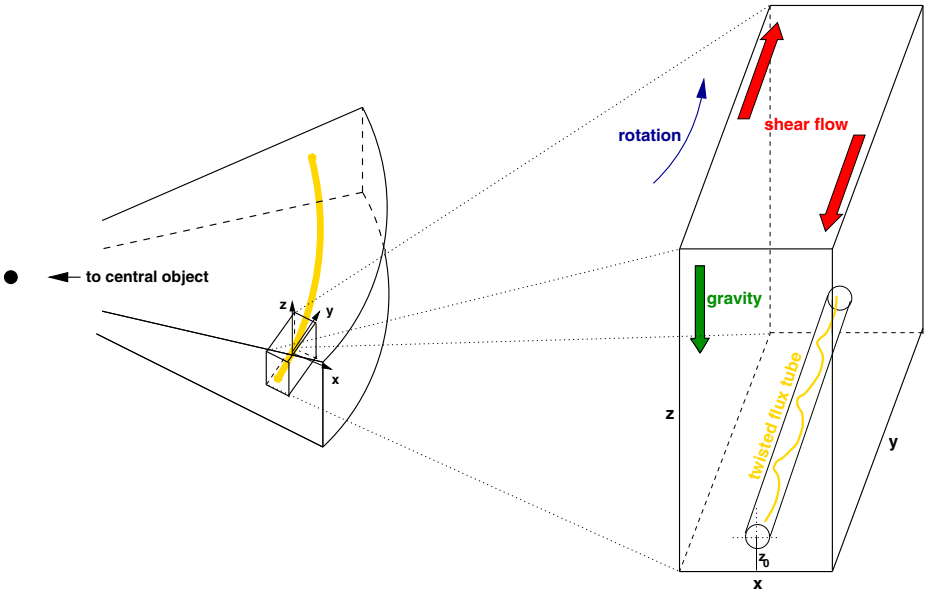


Fig. 6. Schematical illustration of the flux tube/disk model

and to compensate the vertical pressure gradient. In the local box approach (with $|x|, |y|, |z| \ll R_0$) $\varrho \boldsymbol{\Omega}_0 \times (\boldsymbol{\Omega}_0 \times \boldsymbol{x}) + \varrho \boldsymbol{g} = 2\varrho\Omega_0^2 q x \hat{\boldsymbol{x}} - \varrho(2c_s^2/H_0^2) z \hat{\boldsymbol{z}}$. The first term on the r.h.s. describes tidal forces where $q = -d \log \Omega / d \log R|R_0$ is a measure for the local shear rate derived from the rotation profile $\Omega(R)$. For a Kepler disk with $\Omega \propto R^{-1.5}$ it follows $q = 1.5$. The second term on the r.h.s. is just the vertical component of $\varrho \boldsymbol{g}$ where H_0 is the scale height of the isothermal stratification given by the Gauss profile

$$\varrho = \varrho_0 \exp\left(-z^2/H_0^2\right) \tag{23}$$

with ϱ_0 the density in the disk midplane. In case the disk is differentially rotating ($q \neq 0$) the background velocity in the steady state is $\boldsymbol{v}(t=0) = -q\Omega_0 x \hat{\mathbf{y}}$ which describes a uniform shear flow in y-direction.

The magnetic flux tube of radius $\tilde{R}_t$ is placed a distance z_0 above the disk plane with the tube axis oriented parallel to the y-axis as shown in Fig. 6. In a cylindrical coordinate system $\{\tilde{R}, \tilde{\phi}, \tilde{z}\}$ anchored in the tube with the tube axis as $\tilde{z}$-axis the adopted magnetic field reads

$$\begin{pmatrix} B_{\tilde{R}} \\ B_{\tilde{\phi}} \\ B_{\tilde{z}} \end{pmatrix} = \begin{cases} \begin{pmatrix} 0 \\ \xi \tilde{R}/\tilde{R}_t B_{\tilde{z}} \\ B_0 \left(1 - \tilde{R}^2/\tilde{R}_t^2\right)^p \end{pmatrix} & \text{for } \tilde{R} \leq \tilde{R}_t \\ \mathbf{0} & \text{for } \tilde{R} > \tilde{R}_t \end{cases} . \tag{24}$$

There are two components: an axial one, $B_{\tilde{z}}$, and a toroidal one, $B_{\tilde{\phi}}$, i.e. the flux tube may be twisted. The field strength drops off with radius at a rate

depending on the exponent p. The rate of twist is determined by the parameter ξ which is connected to the pitch angle as $\theta = \tan^{-1}(B_{\tilde{\phi}}/B_{\tilde{z}}) = \tan^{-1}(\xi\tilde{R}/\tilde{R}_t)$. In the Cartesian simulation box, the y-component of the magnetic field is given by $B_y(x,z) = B_{\tilde{z}}(\tilde{R}(x, z-z_0))$ with $\tilde{R} = (x^2 + (z-z_0)^2)^{1/2}$. To guarantee solenoidality of the magnetic field numerically B_x and B_z are computed from the vector potential component

$$A_y = \frac{1}{2}\frac{\xi}{p+1}\tilde{R}_t B_0 \left(1 - \frac{\tilde{R}^2}{\tilde{R}_t^2}\right)^{p+1} \tag{25}$$

i.e. $B_x = \partial A_y/\partial z$ and $B_z = -\partial A_y/\partial x$. Of course, the resulting transverse field in the Cartesian coordinate system is identical to the toroidal field in the tube cylindrical coordinate system, $\mathbf{B}_\perp = B_x\hat{\mathbf{x}} + B_z\hat{\mathbf{z}} = B_{\tilde{\phi}}\mathbf{e}_{\tilde{\phi}}$, where $\mathbf{e}_{\tilde{\phi}}$ is the azimuthal unit vector in the tube coordinate system. Note that the magnetic field does not depend on y, hence, is slab symmetric. The field configuration is not force-free and a pressure gradient within the tube is necessary to compensate the Lorentz force. Equilibrium is achieved by calculating the pressure p_i inside the tube from the relation

$$p_i = p - \frac{1}{2\mu}\left(B_\perp^2 + B_y^2\right) \ . \tag{26}$$

The internal density ϱ_i then follows from the assumption of an isothermal state i.e. $p_i/\varrho_i = p/\varrho$. Consequently, $\varrho_i < \varrho$ and the flux tube rises upwards.

The 3D main grid is represented by $64 \times 128 \times 128$ grid points and spans a volume of $(x,y,z) \in [-0.5, 0.5] \times [0,2] \times [0,2]$ measured in units of H_0. In the cases IV and V (see below) the x-interval has been extended to $[-0.75, 0.75]$ to reduce possible boundary effects. A maximum of two refinement levels has been allowed, hence, on the finest level one ends up with an effective resolution of $256 \times 512 \times 512$. Reflecting boundary conditions are used in x-direction and periodic boundary conditions in y-direction. At the bottom z-boundary reflecting boundary conditions are adopted and outflow conditions are used at the top.

Fixed parameters of the problem are the tube radius $\tilde{R}_t = 0.04H_0$, the tube position $z_0 = 2.5\tilde{R}_t$, exponent $p = 0.5$ and strength of the axial field B_0 expressed by $\beta = 2\mu p(\tilde{R}=0)/B_0^2 = 0.5$. Variable parameters of the problem are Ω_0, q and ξ. Five different cases are considered:

case I: $\Omega_0 = q = 0, \xi = 0.17$ (twisted tube in a non-rotating disk)

case II: $\Omega_0 = q = \xi = 0$ (untwisted tube in a non-rotating disk)

case III: $\Omega_0 \neq 0, q = 0, \xi = 0.17$ (twisted tube in a uniformly rotating disk)

case IV: $\Omega_0 \neq 0, q = 1.5, \xi = 0$ (untwisted tube in a Kepler disk)

case V: $\Omega_0 \neq 0, q = 1.5, \xi = 0.17$ (twisted tube in a Kepler disk)

As noted in [25] flux tubes with a field profile (24) are unstable due to the kink instability if $\xi > p^{1/2}$. In the present simulations of twisted flux tubes this criterion is not fulfilled and the tubes are, at least initially, kink-stable.

Before going on it should not be concealed that the following results probably depend on resolution. This is because the solution is sensitive to magnetic diffusion. Although resistivity is not included in the model, the results suffer from numerical diffusion due to truncation errors in the code. The level of numerical diffusion, in turn, depends on the resolution. In effect, the solution does not necessarily converge to the true ideal MHD solution. Future models therefore should include a physical resistivity term and should be validated by convergence tests not possible at present because of a lack in available computer power.

Case I. Case I considers the evolution of a weakly twisted flux tube in a non-rotating, vertically-stratified environment. Under such conditions the flux tube rise is characterized by two major findings, namely, i) the maintenance of the slab symmetry of the magnetic structure during the ascend and ii) the formation of two magnetic sidelobes lagging behind the flux tube. Preserving slab symmetry is not as trivial as it might look at first place. Several magnetic instabilities are known which could break symmetry leading to a three-dimensional configuration. One of these instabilities is the kink instability. The effect of the kink instability is to convert twist into writhe thereby deforming a straight twisted flux tube into a helical untwisted flux tube ([14]). Here, the applied helical magnetic field is subcritical and the flux tube is kink-stable. However, as pointed out in [21], due to the expansion of the flux tube during its rise in lower-density regions the helical field at some height may become supercritical i.e. the toroidal magnetic field component becomes relatively stronger to the axial field component. The simulations show that, except in a narrow layer around the surface of the flux tube, the helical magnetic field never becomes supercritical. This locally enhanced twist turns out not sufficient to initiate a full-blown kink instability. Another potential instability which lead to symmetry breaking is the Parker instability. As first shown in [30], a lower density magnetic layer in a gravitational field becomes unstable under certain conditions. Loosely speaking, the restoring tension force of a vertically-perturbed horizontal magnetic field may be too small to balance the destabilizing effect of buoyancy. In case of a thin flux tube the critical perturbation wavelength above which the Parker instability occurs is $\lambda_{\mathrm{crit}} = 2\pi H_0$ ([33]). The effect of the Parker instability would be to arc the flux tube. Here, the Parker instability plays no role because the dimensions of the computational domain are smaller than the critical wavelength.

Since the solution basically remains two-dimensional (no y-dependence) it is sufficient to visualize results in the (x, z)-plane at some y-coordinate, say in the middle of the y-interval. Figure 7a presents a grey-scale image of the B_y-distribution overlaid by the velocity field after dimensionless time $t = 3.66$ (corresponding to the number of 'orbits' in cases III–V). The formation of sidelobes is clearly seen. Each sidelobe at this time contains roughly 30 percent of the total magnetic flux and is embedded in a swirling flow. The vortex on the left is rotating counterclockwise whereas the right vortex rotates clockwise. The magnetic flux accumulated in the sidelobes originates from the flux tube where it is peeled off from the surface. At the front of the flux tube a vorticity boundary

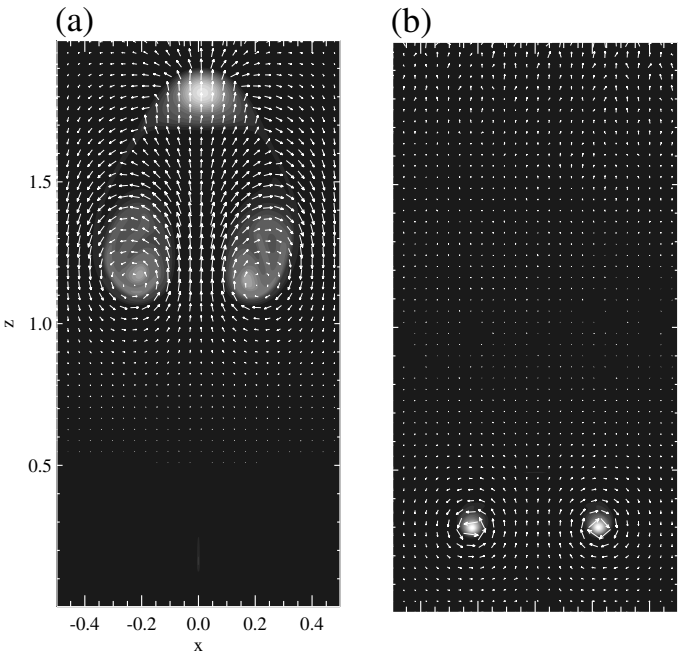


Fig. 7. Snapshot at $t = 3.66$ of the B_y-distribution for a weakly twisted flux tube (**a**) and untwisted flux tube (**b**). In each plot B_y is scaled from its minimum (darker tones) to its maximum (lighter tones)

layer of thickness $\approx 1/15$ of the tube diameter has formed (note that this thickness ratio may depend on resolution). In this small region the kinetic energy of the flow is comparable to or exceeds the magnetic energy of the non-axial field component, thus, favoring the break off of magnetic flux. The (anti-)clockwise rotating vortex has a small (positive) negative velocity component in y-direction. y-momentum is not generated locally, i.e. in the wake, but in the vorticity layer of the flux tube due to magnetic tension forces resulting from a distortion of the tube's cross-section. Then, y-momentum is advected downwards along with the magnetic flux and mixed with the ambient medium by the surrounding flow.

Case II. The rising behavior of a flux tube possessing a purely axial (y-) magnetic field is completely different. First, like in case I, the developing solution is to a very good approximation two-dimensional with no significant variations of the variables along the y-axis. Figure 7b shows the resulting structure after the same time as in the previous twisted tube case. The rise of the flux tube quickly ceases and the tube splits into two pieces. The vertical motion is substantially decelerated and the two fragments, enclosed by vortices of opposite circulation, are deflected in horizontal direction. The fragments are still buoyant but the buoyancy force is partly compensated by a downward-directed lift force of hydrodynamic nature. This lift force is exerted on an object moving translationally

while carrying around it a net circulation ([32]). The observed fragmentation is due to a torque on the flux tube. This torque, on the one hand, is produced by the buoyancy gradient within the flux tube and, on the other hand, is due to the asymmetric pressure distribution around the tube surface exerted by the external medium. Note that a small but sufficient toroidal magnetic field component, missing here, would have a stabilizing effect as demonstrated in case I. In the twisted tube case the magnetic tension force which arises when the cross-section of the tube is deformed counteracts the disruptive impact of the torque. The spatial coherence of a flux tube rising in a Gauss-stratified atmosphere therefore crucially depends on the presence or absence of magnetic field line twist. This fact is relevant also for flux tubes in the solar convection zone ([34], [26], [27], [12], [13], [11]).

Case III. Case III considers the evolution of a weakly twisted flux tube in a uniformly rotating disk. The influence of rotation on the rise of the flux tube can be parameterized in a fair manner by the dimensionless number $Z = \tau_{\rm buoy}/\tau_{\rm rot}$, the ratio of the buoyancy time scale, $\tau_{\rm buoy} = (H_0/a_{\rm buoy})^{1/2}$ where $a_{\rm buoy}$ is the buoyancy acceleration, to the rotation period $\tau_{\rm rot} = 2\pi/\Omega_0$. The effect of rotation is negligible for $Z \ll 1$ because then the flux tube rises the distance H_0 in a fraction of orbital time. Here, the buoyancy acceleration can be estimated to $a_{\rm buoy} = (\varrho/\varrho_i - 1)|g_z| \approx \Omega_0^2 H_0/\beta$ with $\beta^{-1} \approx (\varrho/\varrho_i - 1)$ and $|g_z| \approx \Omega_0^2 H_0$. Hence, $Z \approx \beta^{1/2}/2\pi \approx 0.1$ which lets expect moderate modifications in the flux tube's rising behavior. Differences in the evolution arise due to the action of the Coriolis force affecting the flow structure which, in turn, changes pressure forces on the flux tube.

Like in the cases I and II the slab symmetry of the problem is not broken and no 3D features occur. Figure 8ab presents a close-up view of the magnetic field structure and flow structure at time $t = 2.67$. There are a few major differences compared to the non-rotating but otherwise identical flux tube model of case I. One apparent effect of rotation is the reduction of the rise speed: at earlier time the flux tube yet has not reach the height of the flux tube in case I shown at $t = 3.66$ (see Fig. 7a). Although there is no direct influence on the vertical tube motion (since $\boldsymbol{\Omega} \parallel \hat{\boldsymbol{z}}$), the Coriolis force acts on the surrounding flow in a sense to enhance the pressure difference between the front and rear of the flux tube. The resulting downward-directed force slows down the vertical motion. Somewhat surprisingly, the flux tube during its ascend does not expand as much as found without rotation. Its the Coriolis force again which counteracts expansion. As can be seen from Fig. 8b a transverse flow in y-direction has developed which, in the vicinity of the flux tube, is positive on the left side ($x < 0$) and negative on the right side ($x > 0$). The effect of the Coriolis force resulting from this flow component is to squeeze the flux tube, hence, counteracting expansion due to the decrease in external density. As a consequence, the peak magnetic field strength measured within the flux tube has only decreased to $\approx 0.6 B_0$ which is a factor of ≈ 6 larger than without rotation.

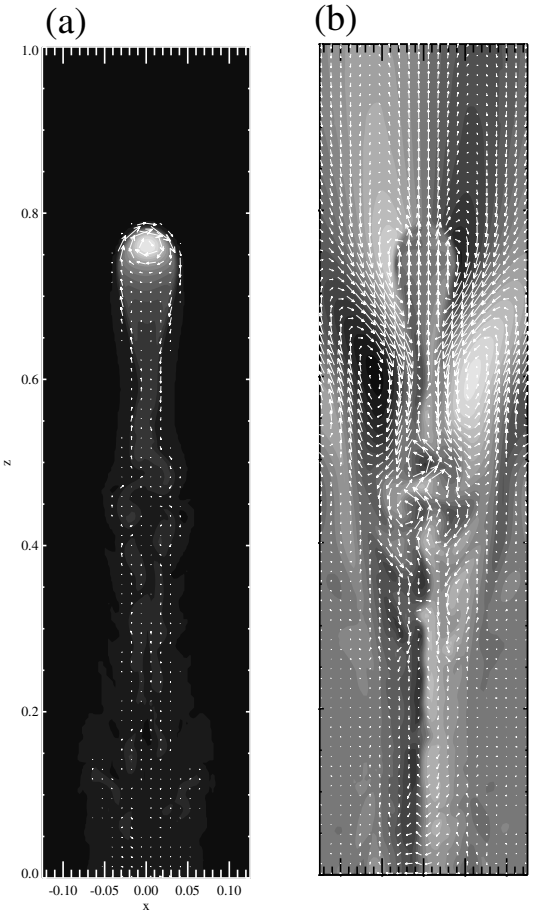


Fig. 8. Close-up view, $(x, z) = [-0.12, 0.12] \times [0, 1]$, of the magnetic field structure (**a**) and velocity field (**b**). The y-component of the fields is represented as grey-scale image where in case of the magnetic field lighter tones mean stronger fields and in case of the velocity lighter (darker) tones denote positive (negative) values. The perpendicular components are shown in terms of arrows

One further major difference compared to case I is the absence of magnetic sidelobes. Magnetic flux peeled off from the tube surface does not accumulate in oppositely spinning sidelobes but is channeled downwards which makes the wake to appear in a comet-like shape. The collimation of magnetic flux is due to nearby large-scale vortices which rise in accord with the flux tube. This vortex pair is a product of the interaction of the flow induced by the tube's motion and rotation. Its origin is therefore of different nature than the vortices accompanying the sidelobes in case I. This can be seen by noting the different sense of circulation. The loss of magnetic flux from the flux tube to the wake is less severe than without rotation. Most of the total magnetic flux is still concentrated in the flux

tube at the given time. The wake is hydrodynamically unstable developing a vortex street which is characterized by a few smaller eddies having an alternating sense of circulation.

Case IV. Case IV deals with the scenario of an untwisted flux tube embedded in an accretion disk. The presence of differential rotation adds new complexity to the problem because of possible implications resulting from the magnetorotational instability (MRI). In fact, the MRI is a natural source for breaking the slab symmetry of the problem. As demonstrated in [3] a differentially rotating plasma with outward decreasing angular velocity, $d\Omega/dR < 0$ ($\Omega \propto R^{-1.5}$ in a Kepler disk), is dynamically unstable if a weak magnetic field is present. Linear growth rates of the instability depend on the initial magnetic field configuration. Growth rates are large (of the order of Ω_0) for a poloidal magnetic field in the disk but small (fractions of Ω_0) for a pure azimuthal disk field. In the latter case non-axisymmetric perturbations are required. Three-dimensional numerical simulations of the non-linear stage of the MRI have revealed that such systems become turbulent independent from the details of the initial magnetic field configuration ([17], [8], [1], [18], [19], [49], [50]). The dynamics of an untwisted flux tube is expected to be not much influenced by the MRI since growth times for the purely axial magnetic field of the tube are one order of magnitude larger than the total simulation time. The situation is, as we will see, different for a twisted tube discussed as case V.

The very initial phase of evolution is similar to the previously discussed cases and is driven by the start-up buoyancy force. The later evolution is determined by the tendency of the flux tube to split due to differential buoyancy within the tube and by stabilizing effects due to rotation. The background shear flow plays only a minor role in the evolution. The end result of the simulation is shown in Fig. 9a corresponding to a time $t = 3.17$. Clearly, the flux tube shows significant deformations of its cross-section but at this time no separation into individual fragments has occurred. We recall that without rotation an untwisted flux tube is quickly disrupted and breaks into two halves which form a vortex filament pair. The subsequent fate of the structure is unclear. It is conceivable that during the later course of evolution the disintegration continues which, ultimately, may lead to fragmentation into several pieces. On the other hand, the flux tube has only a small vertical speed which is in favor of a scenario of a slowly rising flux tube retaining a high degree of coherence.

Case V. Addition of a weak twist field has two obvious consequences: (i) the x-component of the magnetic field is stretched by the background shear flow which strengthens the y-component in the lower half of the tube and weakens it in the upper tube half. The resulting magnetic force has two effects. First, built-up gradients of magnetic pressure inside the flux tube tries to disrupt the flux tube. Second, magnetic tension forces generate an axial flow along the tube which tries to compensate shear inside the tube. The axial flow causes a Coriolis force which rip on the flux tube and tries to separate it horizontally; (ii) the presence

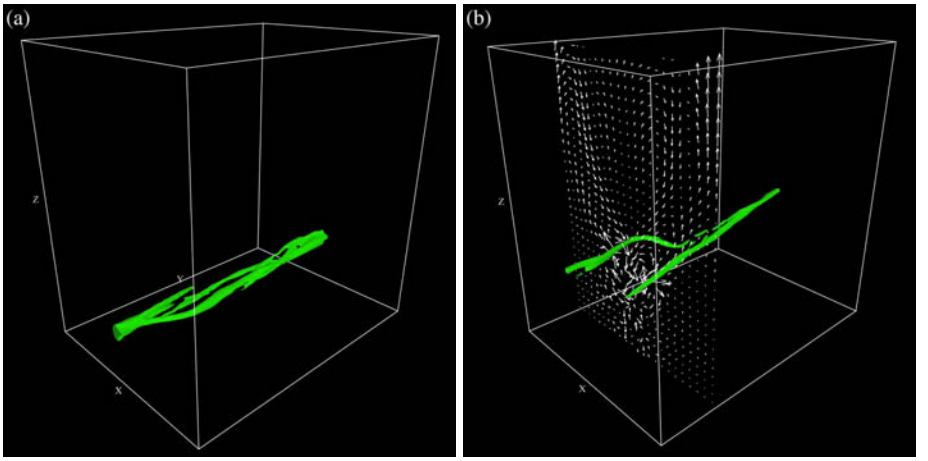


Fig. 9. Isovolume of the magnetic field strength, $|\boldsymbol{B}| > |\boldsymbol{B}|_{\max}/2$, for an untwisted tube at $t = 3.17$ (**a**) and a weakly twisted tube at $t = 3.51$ (**b**) in a Kepler disk. In the twisted tube case the transverse component of velocity in the plane $y = -0.6$ is overplotted

of twist, i.e. the presence of a poloidal magnetic field component, makes the flux tube more prone to the MRI with growth rates of the order of Ω_0. These different effects work in conjunction and let expect a complex dynamical behavior.

Figure 9b illustrates the latest stage of evolution. The snapshot is taken at $t = 3.51$ and shows the spatial location of magnetic flux where $|\boldsymbol{B}| > B_{\max}/2$. $B_{\max} > 0$ is the maximum field strength in the box at the given time. The structure of the projected flow in a slice through the computational domain at $y = -0.6$ is overplotted. Compared with an untwisted tube a more spectacular picture emerges. The resulting magnetic field topology is rather complicated. The initially straight flux tube has become significantly deformed and at one end splits into two ropes giving it a fork-like appearance. The wake in the vicinity of the magnetic structure is characterized by a chaotic flow rather than an ordered one like in the cases I-III. Further away in the upper parts of stratification a more regular flow pattern exists.

AMR Versus 1D Thin Flux Tube Approach. Major aim of this study was to compare the (more realistic) 3D simulations with the 1D thin flux tube approximation (TFA). In the TFA the flux tube is parameterized as a curve in space where cross-sectional variations of the variables are neglected ([35], [15]). One considers this justified if the diameter of the flux tube is small compared with the scale-height of stratification as is the case in this model. Moreover, highly simplified assumptions are usually made to describe the interaction of the flux tube with the ambient medium. In the TFA the rise of a straight, horizontal tube in a stratified medium with external density distribution $\varrho(z)$ is described

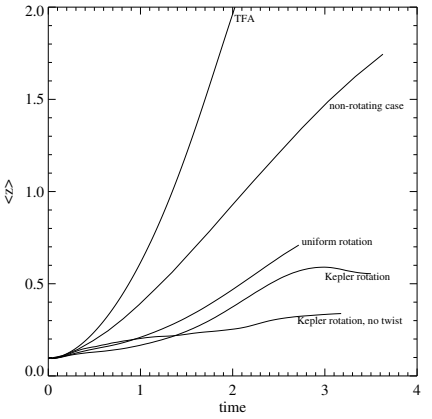


Fig. 10. Time dependence of the vertical position of the flux tube for the various models except case II

by the equations

$$\begin{aligned} \frac{dz}{dt} &= v_z \ , \\ 2\varrho\frac{dv_z}{dt} &= (\varrho - \varrho_i)|g_z| - C_{\rm D}\frac{\varrho|v_z|v_z}{\pi R} \end{aligned} \tag{27}$$

where $C_{\rm D}$ is the aerodynamic drag coefficient and R is the (z-dependent) radius of the flux tube. The factor 2ϱ in front of the acceleration term in the force equation is due to the enhanced inertia effect. Using the requirements of pressure balance, mass conservation, magnetic flux conservation, and an isothermal equation of state, one can determine the flux tube height z as a function of time. Note that rotation of the disk with $\boldsymbol{\Omega}_0 \parallel \hat{\boldsymbol{z}}$ does not alter the vertical motion of the flux tube in this approximation. Therefore, the above ansatz can be used equally well for the non-rotating case and rotating cases.

For a comparison with the TFA, the vertical flux tube position in the 3D models is defined by the ($\boldsymbol{B}^2$-weighted) mean z-coordinate

$$\langle z\rangle(t) = \frac{\int z\boldsymbol{B}^2 dV}{\int \boldsymbol{B}^2 dV} \ . \tag{28}$$

Based on this definition the rising behavior for the various cases is monitored in Fig. 10. The curve labeled with letters 'TFA' is computed from (27) using $C_{\rm D} = 2$ and an initial tube radius $R_0 = R(z = z_0)$ equal to the $\boldsymbol{B}^2$-weighted RMS radius i.e.

$$R_0 = \left(\frac{\int \left[x^2 + (z - z_0)^2\right] \boldsymbol{B}^2 dV}{\int \boldsymbol{B}^2 dV}\right)^{1/2} \ . \tag{29}$$

In the TFA the tube rises exponentially as expected under isothermal conditions and leaves the numerical domain after ≈ 2 orbits. Note that this result depends

on the choice of $C_{\rm D}$. For subsonic flow and in the here relevant Reynolds number regime from ≈ 1 to $\approx 10^5$ $C_{\rm D}$ is of the order of unity.

The 3D twisted tube model in a non-rotating disk (case I: curve labeled 'non-rotating case') is most comparable to the TFA. However, the speed of rise is significantly lower than in the TFA. The discrepancy arises because of the loss of magnetic flux under realistic conditions. This magnetic flux loss reduces the buoyancy and, hence, slows down the flux tube. The reduction in speed is even more pronounced if the disk is uniformly rotating (case III: curve labeled 'uniform rotation') which is due to the impact of the Coriolis force. The largest deviations from the TFA are found for flux tubes ascending in a Kepler disk (case IV: curve labeled 'Kepler rotation, no twist' and case V: curve labeled 'Kepler rotation'). Actually, the presence of shear plays only a minor role in case IV and the rising behavior is dominated by the action of differential buoyancy and rotation. The result is a very slow flux tube rise. On the other hand, shear has a major impact on the evolution of a weakly twisted flux tube. In this case the flux tube stops its rise at $t \approx 3$ and turns around.

The study shows that 3D numerical models are an absolute necessity to simulate the correct dynamical behavior of a thin magnetic flux tube ascending in a stratified atmosphere. The large difference in spatial scales – the dimension of the flux tube is small compared to the density scale height of the stratification – require a high resolution. To reduce computational costs for such problems the use of adaptive mesh techniques as implemented in the code [1]NIRVANA are advantageous.

References

1. M. Abramowicz, A. Brandenburg, J-P. Lasota: Mon. Not. Roy. Astron. Soc. 281, L21, (1996)
2. W.F. Ames: In: *Numerical methods for partial differential equations, 1977*, (Academic Press, Inc.)
3. S.A. Balbus, J.F. Hawley: ApJ 376, 214, (1991)
4. J. Bell, M.J. Berger, J. Saltzman, and M. Welcome: SIAM J. Sci. Stat., 15, 127, (1994)
5. M.J. Berger.: PhD thesis, Stanford University, (1982)
6. M.J. Berger, J. Oliger: J. Comp. Phys. 53, 484, (1984)
7. M.J. Berger, P. Colella: J. Comp. Phys. 82, 64, (1989)
8. A. Brandenburg, Å. Nordlund, R.F. Stein, U. Torkelsson: ApJ 446, 741, (1995)
9. G.S. Dietachmayer: Monthly Weath. Rev. 120, 1707, (1992)
10. G.S. Dietachmayer, K.K. Droegemeyer: Monthly Weath. Rev. 120, 1675, (1992)
11. S.B.F. Dorch, Å. Nordlund: Astron. Astrophys. 338, 329, (1998)
12. T. Emonet, F. Moreno-Insertis: ApJ 458, 783, (1996)
13. T. Emonet, F. Moreno-Insertis: ApJ 492, 804, (1998)
14. Y. Fan, E.G. Zweibel, M.G. Linton, G.H. Fisher: ApJ 505, L59, (1998)
15. A. Ferriz-Mas, M. Schüssler: Geophys. Astrophys. Fluid Dyn. 72, 209, (1993)
16. B. Fiedler: CRM Proc. Lecture Notes 16, AMS, Providence, RI, 99, (1995)

[1] NIRVANA is freely available at URI http://www.aip.de/groups/MHD

17. J.F. Hawley, F. Gammie, S.A. Balbus: ApJ 440, 742, (1995)
18. J.F. Hawley, F. Gammie, S.A. Balbus: ApJ 464, 690, (1996)
19. J.F. Hawley, J.M. Stone: ApJ 501, 758, (1998)
20. T. Horiuchi, S. Salo: PASJ 42, 661, (1993)
21. V. Kutznetsov, A. Hood: Solar Phys. 171, 61, (1997)
22. Y. Kallinderis, J.R. Baron: Comput. Fluids 21, 77, (1992)
23. R.I. Klein: J. Comp. Appl. Math. 109, 123, (1999)
24. N.I. Kleeorin, I.V. Rogachevskii: Proceedings of ESA workshop SP-311 on Plasma Astrophysics, Telavi, Georgia, (1990)
25. M.G. Linton, R.B. Dahlburg, G.H. Fisher, D.W. Longcope: Astrophys. Journal 507, 404, (1998)
26. D.W. Longcope, G.H. Fisher, S. Arendt: ApJ 464, 999, (1996)
27. F. Moreno-Insertis, T. Emonet: ApJ 472, L53, (1996)
28. R. Mönchmeyer, E. Müller: Astron. Astrophys. 217, 351, (1989)
29. A. Orszag, C.M. Tang: J. Fluid Mech. 90, 129, (1979)
30. E.N. Parker: ApJ 191, 245, (1974)
31. J.M. Picone, R.B. Dahlburg, L.O. Drury: Phys. Fluids B 3, 29, (1991)
32. P.G. Saffman: In: *Vortex Dynamics, 1992*, (Cambridge Univ. Press)
33. G.P. Schramkowski, A. Achterberg: Space Sci. Rev. 68, 331, (1994)
34. M. Schüssler: Astron. Astrophys. 71, 79, (1979)
35. H.C. Spruit: Astron. Astrophys. 98, 155, (1981)
36. O. Steiner, M. Knölker, M. Schüssler: In: *Solar Surface Magnetism, Proc. NATO Advanced Research Workshop, 1993*, ed. by R.J. Rutten, C.J. Schrijver (Kluwer Academic Publishers, Dordrecht)
37. J.M. Stone, M.L. Norman: ApJS 80, 791, (1992)
38. G. Toth: J. Comp. Phys. 161, 2, (2000)
39. K.J. Truelove, R.I. Klein, C.F. McKee, J.H. Holliman, L.H. Howell, J.A. Greenhough: ApJ 489, L170, (1997)
40. K.J. Truelove, R.I. Klein, C.F. McKee, J.H. Holliman, L.H. Howell, J.A. Greenhough, T.D. Woods: ApJ 495, 821, (1998)
41. W.-M. Tscharnuter, K.-H. Winkler: Comp. Phys. Comm. 18, 171, (1979)
42. B. van Leer: J. Comput. Phys. 23, 276, (1977)
43. G. Wang, D.K. Tafti: Int. J. Supercomp. Appl. and High Performance Comp., 13, 1, (1999)
44. P.R. Woodward, P. Collela: J. Comp. Phys. 54, 115, (1984)
45. A.L. Zachary, A. Malagoli, P. Collela: Siam J. Sci. Comp. 15, 263, (1994)
46. U. Ziegler, H.W. Yorke: Comp. Phys. Comm. 101, 54, (1997)
47. U. Ziegler: Comp. Phys. Comm. 109, 111, (1998)
48. U. Ziegler: Comp. Phys. Comm. 116, 65, (1999)
49. U. Ziegler, G. Rüdiger: Astron. Astrophys. 346, 1141, (2000)
50. U. Ziegler, G. Rüdiger: Astron. Astrophys. 378, 668, (2001)

Observations of Interstellar Magnetic Fields

Richard Crutcher[1], Carl Heiles[2], and Thomas Troland[3]

[1] University of Illinois, Urbana IL 61801, USA
[2] University of California, Berkeley CA 91820, USA
[3] University of Kentucky, Lexington KY 40502, USA

Abstract. This article describes how interstellar magnetic fields are detected, measured, and mapped, the results of such observations, and the role played by interstellar magnetic fields in the physics of the interstellar medium. A goal of the observations is the measurement of the morphology and strengths of the uniform (B_u) and random (B_r) components of magnetic fields. Observational techniques probe either the component of **B** parallel to the line of sight ($B_\parallel$) or in the plane of the sky ($B_\perp$). Tracers of $B_\parallel$ are Faraday rotation of the position angle of linearly polarized radiation and Zeeman splitting of spectral lines. Tracers of $B_\perp$ are the strength of synchrotron radiation and linear polarization of synchrotron radiation and of emission or absorption from dust and spectral lines.

Starlight polarization shows that on large spatial scales the Galactic magnetic field is not heavily tangled ($B_u/B_r \approx 0.7 - 1.0$), that the field is generally parallel to the Galactic plane near the plane, that the local field points approximately along the local spiral arm (pitch angle 9.4°, center of curvature 7.8 kpc distant towards $\ell \approx -15.4°$), and that off the Galactic plane there is considerable small-scale structure to the field. Galactic synchrotron emission shows magnetic spiral arms with a total strength $B_t \approx 6$ μG and $B_u \approx 4$ μG. Pulsar data show evidence for reversals of the field direction with Galactic radius and yield $B_r \approx 5$ μG and $B_u \approx 1.5$ μG; the morphology of the large-scale mean field is consistent with dynamo generation. H I Zeeman detections for diffuse clouds yield $B_\parallel \sim 5 - 20$ μG with many limits $B_\parallel < 5$ μG. A recent survey of Galactic H I in absorption against extragalactic sources confirms the result that the fields in diffuse clouds are often quite weak.

The critical parameter for evaluating the importance of magnetic fields in star formation is the ratio of the mass to the magnetic flux, M/Φ_B; observations focus on measuring both this quantity and the morphology of fields in dense regions. Zeeman observations of molecular lines are consistent with $B \propto \sigma_v\sqrt{n}$, which is the theoretical prediction for flattened cores supported by a combination of a uniform magnetic field pressure and turbulence. In cores, motions are approximately Alfvénic, and M/Φ_B has a critical to slightly supercritical value. The ratio of B_r/B_u appears to decline with density. In some molecular cores there is evidence for the "hourglass" pinch that would be produced by cloud contraction with the magnetic field frozen into the matter.

1 Introduction

The interstellar medium is permeated by magnetic fields that have important astrophysical consequences. The focus of this review is on describing new observational results, particularly in the dense interstellar medium, where the role of magnetic fields in star formation remains uncertain, and on reviewing the

E. Falgarone and T. Passot (Eds.): LNP 614, pp. 155–181, 2003.
© Springer-Verlag Berlin Heidelberg 2003

observational evidence for the relative importance of the uniform and random (turbulent) fields.

The morphology and strengths of the uniform (B_u) and random (B_r) components of magnetic fields on Galactic scales has been an area of active research for the past 50 years, since the discovery of interstellar polarization of starlight. Beck [2] has recently written an excellent and comprehensive review of Galactic and extragalactic magnetic fields that we will summarize here. Our discussion will focus on several recent results on the large-scale mean Galactic field.

The role of magnetic fields in the star formation process has been the subject of many theoretical and observational investigations. Recent reviews include those of McKee [38], Shu et al. [47], and Mouschovias and Ciolek [41]. An important question is whether or not the magnetic energy density is comparable to other energy densities, such as gravitational, thermal, and turbulent. If so, then magnetic fields must play a significant role in cloud evolution and star formation. If not, then their role may be secondary or unimportant. There are two ways in which this question may be framed so that it can be addressed by observations.

What is the magnetic field morphology? It may be heavily tangled, which would suggest that the magnetic energy density is dominated by turbulence, or it may be fairly uniform. If the field is relatively uniform, its geometry may have the hourglass shape predicted by theories in which magnetic fields dominate cloud support, with the axis of the hourglass along the minor axis of the flattened disk structures (oblate spheroids) that would form due to there being less support against gravity along field lines. If $B_u < B_r$, turbulence dominates, but the magnetic field may still be an important component of the physics of star formation. Finally, recent theoretical work [15], [16] (see also Fiege in this book) has led to the proposal that molecular clouds may be mainly filaments (prolate spheroids) with magnetic fields having a helical morphology – a combination of a toroidal and a poloidal field.

What is the magnetic field strength? This is the crucial issue, which is addressed indirectly by the question of whether the field is tangled or uniform. If we can measure the strengths of the random and uniform components of the total magnetic field, we can quantitatively answer the question about the relative importance of turbulent and magnetic energies. And we can answer the question whether magnetic fields support clouds against gravitational collapse. The crucial parameter is M/Φ_B, where M is a cloud's mass and Φ_B is its magnetic flux. If B is sufficiently strong (M/Φ_B critical or subcritical), the static magnetic pressure can support a cloud core against gravity; collapse and subsequent star formation then depend on ambipolar diffusion to increase M/Φ_B until the core is supercritical and magnetic pressure is no longer sufficient for support. Supercritical cores cannot be supported by static magnetic pressure.

Our discussion of magnetic fields in dense regions will focus on these two questions.

2 Observational Techniques

All methods of observing magnetic fields in the interstellar medium are sensitive only either to the component of $\boldsymbol{B}$ parallel to the line of sight, $B_{\parallel}$, or to the component in the plane of the sky, $B_{\perp}$. Direct tracers involve the detection of polarized radiation emitted by or passing through the interstellar medium.

2.1 Polarization

The polarization of electromagnetic radiation may be conveniently described by the Stokes parameters I, Q, U, and V, where I describes the total intensity, Q and U the linear polarization, and V the circular polarization of radiation. The total polarized intensity is $I_p = \sqrt{Q^2 + U^2 + V^2}$. What is often stated for observations of linear polarization is the fractional polarization I_p/I and the position angle ψ of the major axis of the polarization ellipse, $\psi = \frac{1}{2} \arctan(U/Q)$.

If radiation is observed with a detector sensitive to linear polarization, the maximum signal will be detected when the detector is oriented at the position angle ψ. If a detector sensitive to circular polarization is used to measure the intensities I_R and I_L of the right and left circularly polarized components, Stokes V may be inferred directly from the difference. In radio astronomy, when the amplitude and phase of the electromagnetic wave received by each antenna feed are available, the full set of Stokes parameters may be inferred from the cross-correlation of orthogonal (linear or circular) polarization feeds.

2.2 Tracers of $\boldsymbol{B}_{\parallel}$

Faraday Rotation

A linearly polarized electromagnetic wave may be decomposed into two oppositely circularly polarized waves. These two waves propagate through an interstellar medium that contains ions and magnetic fields with different phase velocities. The result is that the position angle of the linear polarization will rotate. This Faraday rotation produced by propagation through a path length L is given by $\Delta\psi \propto \lambda^2 \int_0^L B_{\parallel}(z) n_e(z) dz$. Hence, observations of Faraday rotation of linearly polarized radiation (usually synchrotron radiation) may be used to infer the strength of the line-of-sight component of $\boldsymbol{B}$ if information about the electron density is known or assumed. Faraday rotation is generally quoted as a rotation measure, $RM = -\int_0^L B_{\parallel} n_e dz$. Note the convention that a negative RM means $\boldsymbol{B}$ is pointed away from the observer, $B > 0$.

In the special case of observations of pulsars, information about the electron density may be obtained from the dispersion with radio frequency in the arrival time of radio pulses. The dispersion measure is given by $DM = \int_0^L n_e dz$. Then the mean line-of-sight component of $\boldsymbol{B}$ is $\overline{B}_{\parallel} = RM/DM$. This $B_{\parallel}$ is a vector average along the line of sight and may be much smaller than the localized fields if there are reversals in field direction.

Zeeman Splitting

If a spectral line forming region is permeated by a field $\boldsymbol{B}$, the radiation is split by the normal Zeeman effect into three separate frequencies, $\nu_{\sigma-} = \nu_0 - \nu_Z$, $\nu_\pi = \nu_0$, and $\nu_{\sigma+} = \nu_0 + \nu_Z$, where ν_0 is the line frequency, $\nu_Z = B \times Z$, and Z is the Zeeman sensitivity in Hz/μG. For a magnetic field in the plane of the sky, the three Zeeman components are linearly polarized with the π component parallel to and the σ components perpendicular to the magnetic field direction. For a magnetic field along the line of sight, $I_\pi = 0$ and the I_σ are oppositely circularly polarized. If the magnetic field is parallel (antiparallel) to the direction of propagation of the radiation, $I_\sigma+$ is right (left) circularly polarized. In the general case of arbitrary angle θ between the line of sight and the magnetic field, all three Zeeman components will be elliptically polarized.

If B is sufficiently large, then $\nu_Z \sim \Delta\nu$ where $\Delta\nu$ is the line width. Then the Zeeman components may appear separated, and the full magnetic field strength can be derived directly from the measured ν_Z. Generally $\nu_Z \ll \Delta\nu$, and it is not possible to infer complete information about $\boldsymbol{B}$ from Zeeman observations. Although the Stokes parameters V, Q, and U for the Zeeman components provide in principle full information about magnetic field strength and direction, in practice full information on $\boldsymbol{B}$ cannot be obtained owing to the extreme weakness of Q and U (by a factor $\sim \nu_Z/\Delta\nu$ compared with V). The Stokes V spectra reveal the sign (i.e., direction) and magnitude of the line-of-sight component $B_\parallel$. That is, $V = (dI/d\nu)\nu_Z \cos\theta$. The sign convention is that $B_\parallel$ positive means that the line-of-sight component of B points away from the observer. By fitting the frequency derivative of the Stokes parameter I spectrum $dI/d\nu$ to the observed V spectrum, $B_\parallel = \nu_Z \cos\theta/Z$ may be inferred. For a large number of clouds whose magnetic fields are randomly oriented with respect to the observed line of sight, the statistical relationship between $B_\parallel$ and $|\boldsymbol{B}|$ is $\overline{B}_\parallel = \int_0^{\pi/2} |\boldsymbol{B}| \cos\theta \sin\theta d\theta / \int_0^{\pi/2} \sin\theta d\theta = \frac{1}{2}|\boldsymbol{B}|$.

In general, only those species with an unpaired outer electron will have a magnetic moment that scales with the Bohr magneton, $M_B = eh/4\pi m_H c = 1.4$ Hz/μG. For the 1420 MHz line of H I, $Z = M_B$. Molecular lines that are sensitive to the Zeeman effect have $Z \approx M_B$; for example, $Z = 0.98$ Hz/μG for the 1667 MHz line of OH. Otherwise, the Zeeman effect will scale with the nuclear magneton, 1840 times smaller than M_B. Unfortunately, most molecules do not have large Zeeman splitting factors Z, so the possibilities for Zeeman observations are limited. Lines for which the Zeeman effect has been detected that have been used for measurements of interstellar magnetic fields are the 21-cm line of H I, 18-cm and 2-cm lines of OH, the 1.3-cm H_2O maser line (a nuclear magneton case), and the 3-mm lines of CN. Other possible Zeeman molecules that might become useful probes of interstellar magnetic fields include CH, CCS, SO, and O_2; O_2 has strong Zeeman splitting and may probe high-density interiors of clouds, but transitions would have to be observed from space.

2.3 Tracers of $B_\perp$

Synchrotron Radiation

An electron moving with velocity $\boldsymbol{v}$ in a magnetic field $\boldsymbol{B}$ will move in a helical path around the field lines due to the Lorentz force $\boldsymbol{v} \times \boldsymbol{B}$ and will therefore emit linearly polarized electromagnetic radiation due to this acceleration. An ensemble of electrons is generally treated by assuming that the distribution function of the energies of the electrons is a power law with index p: $N(E)dE = N_0 E^{-p} dE$. For an assumed uniform magnetic field in a region of extent L along the line of sight, the intensity of radiation is $I(\nu) \propto N_0 L B_\perp^{n+1} \nu^{-n}$, where $n = (p-1)/2$. For a uniform magnetic field radiation is linearly polarized perpendicular to $\boldsymbol{B}$, with the fractional linear polarization $I_p/I = \frac{n+1}{n+5/3}$. Typical values of p are $2 < p < 4$.

Hence, by measuring the position angle of the linearly polarized synchrotron radiation emitted by relativistic electrons, the direction of the magnetic field in the plane of the sky may be mapped. However, it is necessary to account for Faraday rotation to do this, which is a significant limitation. An estimate of the ratio of strengths of the random and uniform components of the magnetic field may be obtained from the degree of polarization, since for a completely random field the synchrotron radiation will be unpolarized.

By making various assumptions, the most important being equilibrium between the energy densities of the relativistic electrons and of the magnetic field, an estimate of the strength of the magnetic field may be obtained from the flux density S_ν of the synchrotron emission: $B_{eq} \propto \left(S_\nu \nu^n\right)^{2/7} D^2/V$, where D is the distance to the emitting gas and V is its volume.

These techniques are most useful for mapping the large-scale morphologies of interstellar magnetic fields in external galaxies and in the Milky Way.

Dust Polarization

Linear polarization at optical, infrared, or millimeter wavelengths arises from elongated grains with their short axes aligned with the magnetic field; Lazarian et al. [36] have discussed the various mechanisms that may lead to polarization. The Davis-Greenstein [13], [14] mechanism assumes that charged, spinning interstellar grains are aligned via paramagnetic relaxation, which damps the components of a grain's angular momentum perpendicular to $\boldsymbol{B}$, gradually aligning the spin axis with the magnetic field direction. Although the original Davis-Greenstein theory considered only thermal spins, much weaker magnetic fields can also align grains. Two mechanisms are superthermal spin energies produced by ejection of H_2 molecules from their formation sites on grains and radiative torquing. Internal damping due to Barnett relaxation normally causes the grain to spin about its axis of greatest moment of inertia, so the long axis of the grain is oriented perpendicular to the field.

However, there are other alignment mechanisms. One which may dominate paramagnetic relaxation in protostellar jet or wind regions is Gold alignment

[17], [35] in which gas streaming past the grains excites their angular momenta perpendicular to the flow direction, causing the grains to tend to rotate with their long axes parallel to the flow. Unless there is reason to believe that some other process dominates, Davis-Greenstein type alignment is assumed.

Grain properties and the grain alignment mechanism are not understood in sufficient detail for field strengths to be reliably inferred, but observations of linear polarization trace the position angle of the component of $\boldsymbol{B}$ in the plane of the sky.

Starlight passing through the interstellar medium will be linearly polarized due to extinction by aligned interstellar grains along the line of sight, so long as the magnetic field is not completely random. Because Davis-Greenstein alignment produces the maximum extinction cross-section perpendicular to the magnetic field, extinction will be greater perpendicular to field than parallel to it. Hence, the position angle of the polarization ellipse of background starlight will be parallel to the magnetic field direction. On the other hand, when *emission* from dust grains is observed, the position angle of maximum emission will be perpendicular to the magnetic field. Linearly polarized thermal emission from dust grains at far infrared and millimeter wavelengths may be used to map magnetic field morphologies.

Polarization of radiation due to extinction or emission from dust grains has been used for 50 years to study interstellar magnetic fields. The main application of this technique has been to map the morphology of $\boldsymbol{B}$. But in the early days of starlight polarization studies, Chandrasekhar and Fermi [6] proposed an indirect, statistical method of inferring field strengths from these data. The method depends on the fact that motions within interstellar plasmas are nonthermal; hence, magnetic fields will not be uniform, but will have a perturbed or MHD wave component. If we start with the expression for the Alfvén velocity $V_A = B/\sqrt{4\pi\varrho}$, we note that velocity perturbations δV and magnetic field perturbations δB will be related by $\delta V = \delta B/\sqrt{4\pi\varrho}$ (for incompressible turbulence). If $B_\perp$ is the magnitude of the average field in the plane of the sky, δB will produce a perturbation in the uniform direction of the position angles of polarization. If $\delta\phi = \delta B/B_\perp$ is the dispersion in the position angles of the polarized emission, $B_\perp = \mathcal{Q}\sqrt{4\pi\varrho}(\delta V/\delta\phi)$, where $\mathcal{Q}$ is a factor of order 1. A logical inconsistency in applying this expression to observations is that what is needed is the velocity perturbation in the plane of the sky (δV), while what is available from observations is the radial velocity dispersion of spectral lines, σ_{turb}. However, it may be that at least statistically δV may be inferred from σ_{turb}. Ostriker et al. [43] have studied the applicability of the Chandrasekhar-Fermi method of estimating $B_\perp$ and have calibrated $\mathcal{Q}$ by "observing" the polarized dust emission from their simulations of interstellar plasmas; they found that $\mathcal{Q} \approx 0.5$ for the conditions that prevail in molecular clouds. Heitsch et al. [30] have extended this work. Although there appears to be a large uncertainty in $\mathcal{Q}$ at low densities or where $B_u < B_r$, in dense regions $\mathcal{Q}$ appears to be well constrained, and the method yields useful results. In addition, algebra gives us an expression for the ratio of turbulent to magnetic energies: $\beta_{turb} = (\sigma_{turb}/V_A)^2 \approx 2.7\delta\phi^2$. Hence,

linear polarization observations allow us to map the morphology of magnetic fields, infer the average strength of the field in the plane of the sky, and infer the ratio of turbulent to magnetic energies.

Hildebrand et al. [31] have reviewed the principles of observations and analysis of far-infrared polarimetry of thermal dust emission.

Spectral-Line Linear Polarization

Linear polarization may also arise in radio-frequency spectral lines formed in the interstellar medium [18],[33]. For a molecule with (say) levels with total angular momenta F = 1 and F = 0, in the presence of a magnetic field the degeneracy of magnetic sublevels will be removed and the level with $F = 1$ will be split into three sublevels. The three transitions will be polarized as usual for this situation, but so long as the three sublevels of the $F = 1$ state are populated equally, the net polarization is zero. If the line optical depth to the source is anisotropic, the radiation field will be anisotropic and the radiative rates between level $F = 0$ and the three sublevels of $F = 1$ will be unequal. Thus, unequal populations of the three sublevels can be sustained, and net linear polarization will result. Since astrophysical sources are always anisotropic at some level, linear polarization of line emission is expected. The degree of polarization depends on the degree of anisotropy, the optical depth of the spectral line, the angles between the line of sight, the magnetic field, and the axis of symmetry of the velocity field, and the ratio of the collisional excitation rate to the radiative rate. The Zeeman splitting of the magnetic sublevels need only be larger than the natural width of the spectral line for this Goldreich-Kylafis effect to be possible; since the observed molecules need not have a large Zeeman effect, it is applicable to abundant molecules like CO. The direction of the polarization can be either parallel or perpendicular to the magnetic field; therefore, care must be taken in inferring the direction of $\boldsymbol{B}$. As for dust polarization, the degree of line polarization cannot be used to infer the field strength. Maps of the position angles of the polarization yield the field morphology and open the prospect of exploring this morphology as a function of radial velocity. In addition, the Chandrasekhar-Fermi relationship may be applied to infer the mean value of B_u and B_r/B_u.

3 Large-Scale Magnetic Fields

In this section we discuss results on the large-scale magnetic field of the Milky Way and of other galaxies inferred from optical polarization, H I Zeeman, pulsar, and synchrotron emission studies. One should keep in mind that these observational techniques sample different phases of the interstellar medium, which will lead to apparently conflicting results. Pulsars sample the warm ionized medium, optical polarization and H I Zeeman primarily the neutral medium, and synchrotron emission the entire volume. We have already referred the reader to the excellent review by Beck [2], which notes that the mean Galactic magnetic field seems to follow the spiral arms and the random component exceeds the mean

component in strength. We first summarize these results, then discuss several new results.

3.1 Starlight Polarization

Figure 1 shows a plot of starlight polarizations for nearly 10,000 stars. Each line represents the polarization measurement for a star; the length is proportional to the fractional polarization and the orientation is in the direction of maximum polarization, which is parallel to the magnetic field. These data show several important characteristics of the local interstellar magnetic field [25]. First, for Galactic latitudes b within about 10° of the Galactic plane, the magnetic field is generally parallel to the plane of the Galaxy. Second, the polarization directions are consistent with parallel local magnetic field lies that appear to converge as you look along the direction of the field. The convergence points to Galactic longitudes of $\ell \sim 80°$ and $\ell \sim 260°$, which suggests that this is the orientation of the local spiral arm. In principle it is possible to derive the ratio B_u/B_r from an analysis of the uniformity of the optical polarization data. The results are somewhat dependent on models that must be adopted for the analysis, but are reasonably self-consistent: $B_u/B_r \approx 0.7 - 1.0$ [25].

At positions more than $\sim 10°$ off the Galactic plane there is considerable small-scale structure. Positions in Taurus, Perseus, Ophiuchus, and Orion correspond to massive molecular clouds undergoing active star formation; the magnetic field lines are obviously perturbed from the regular, large-scale pattern. Finally, at high Galactic latitude several local interstellar bubbles each produced

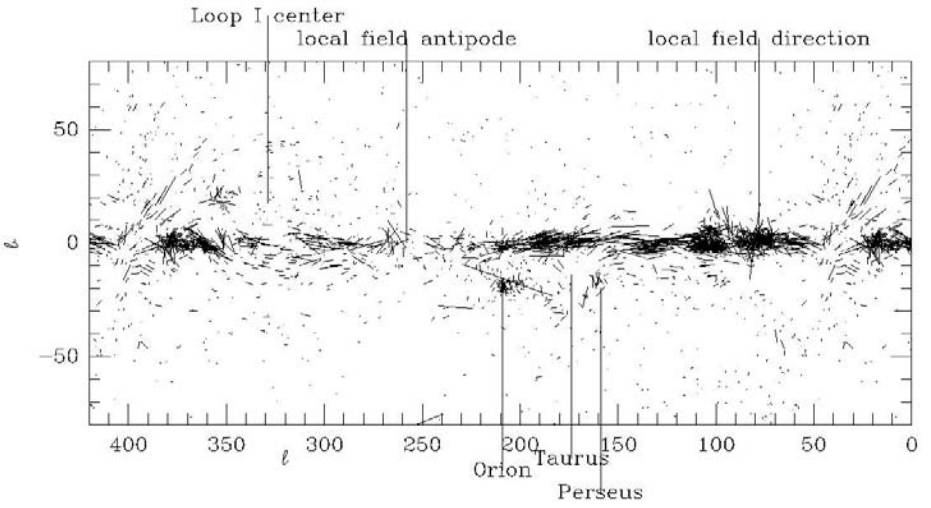


Fig. 1. Optical starlight polarization. Each line segment represents the polarization direction and percentage of a single star. Polarization percentages range up to about 3%. Since the starlight polarization is produced by absorption, the line segments are parallel to the magnetic field.

by multiple supernovae may be seen. The most prominent is Loop I, whose walls are outlined by a loop of magnetic field lines with a radius of $\sim 60^\circ$.

3.2 Synchrotron Emission

Synchrotron emission may be used to infer the total (uniform and random) strength B_t of the Galactic magnetic field. Unfortunately, Faraday rotation depolarizes emission from distant regions of the Galaxy, so the polarization that you see is dominated by nearby emission. Maps of the Galaxy in radio synchrotron emission show spiral arms, suggesting that $\boldsymbol{B}$ is stronger in arm than in interarm regions (assuming that the emission structure is dominated by variation in $\boldsymbol{B}$ and not in the density of relativistic electrons). However, these "magnetic" spiral arms may not be coincident with "matter" spiral arms traced by H I and young stars. Derivation of the field strengths involves various assumptions (especially energy equipartition) and careful analysis – the best recent results (summarized by Beck [2]) are the following. The strength of the total magnetic field is $B_t \approx 6 \pm 2$ μG locally, and $B_t \approx 10 \pm 3$ μG at $R = 3$ kpc. Near the solar circle the ratio of the uniform to the total magnetic field strength $B_u/B_t \approx 0.6$; thus, $B_u \approx 4$ μG. Since polarized emission does not distinguish between coherent and incoherent fields, these estimates are likely to be upper limits.

Synchrotron emission decreases with height above the Galactic plane. When modeled as a thin and thick disk with equal emission at the Galactic plane, the scale heights are about 0.2 and 1.8 kpc at the sun. Also, the scale height increases with Galactic radius. Because synchrotron emission depends on B^{n+1}, the scale height of $\boldsymbol{B}$ is significantly larger than that of the synchrotron emission.

3.3 Pulsars

Faraday rotation of the linearly polarized radiation from pulsars and extragalactic radio sources provides information about the strength of the interstellar magnetic field along the line of sight. Pulsar data give the mean field $\overline{B}_\parallel$ directly from the ratio RM/DM. However, $\overline{B}_\parallel$ is a weighted mean, weighted by the electron density; this method therefore samples mainly the warm ionized component of the interstellar medium, which is believed to occupy a small fraction of the volume of interstellar space [23].

The pulsar data show clear evidence for reversals in the magnetic field direction at various Galactocentric radii. Although these reversals are generally assigned to interarm regions, recent work has suggested that the field reverses within spiral arms themselves [32], [21]. The field in the local spiral arm points in the direction of Galactic rotation. $\overline{B}_\parallel \approx 1.4 \pm 0.2$ μG near the Sun, rising to ~ 3.5 μG in the Sagittarius arm and ~ 6.5 μG inside the second reversal radius (5.5 kpc). Note that these values are for the uniform component, B_u, only; field reversals and possible anticorrelaton between n_e and B_u lead to the inferred value being an underestimate. Statistical analysis of the deviations of $\overline{B}_\parallel$ from a uniform field or from comparisons of $\overline{B}_\parallel$ values from pulsars close together in

the sky yield estimates of the strength of the random component of the magnetic field; the result is $B_r \approx 5$ μG. Therefore, $B_u/B_r \approx 0.3$.

3.4 H I Zeeman

Magnetic fields in the cold neutral medium (CNM) of the Galaxy can be studied via the Zeeman effect in H I absorption lines [25]. One study of a large region in which RM's are consistently large and negative yielded many upper limits $B_\parallel \approx 7$ μG, with detections at lower values when greater sensitivity was achieved. One position with $B_\parallel \approx 12$ μG came from an H I filament probably produced by an interstellar shock which compressed both the gas and the field. Observations of the H I Zeeman effect toward well-known interstellar bubbles (Loop I, Loop II, the Eridanus bubble) yield $B_\parallel \sim 5 - 10$ μG. The magnetic field has been enhanced by compression in the walls of the bubbles. When the H I Zeeman effect is detected in diffuse clouds, the observations yield typical values $B_\parallel \sim 5 - 20$ μG (although there are many upper limits $B_\parallel < 5$ μG), with clouds of densities up to $n(H\ I) \sim$ several hundred being sampled. Hence, there is no clear increase in $\boldsymbol{B}$ with gas density for $n < 10^3$ cm^{-3}.

3.5 Galactic Center

The central region of the Milky Way has very strong magnetic fields, with strengths of several milligauss. Evidence for strong fields comes from large-scale, linear filaments of radio emission, highly polarized radio continuum emission, polarized infrared emission, Zeeman measurements of H I and OH absorption, and anisotropic scattering of radio emission from small sources. Models of the Galactic center have a circumnuclear disk threaded by a poloidal field, with the rotation of the disk producing an azimuthal component.

3.6 Other Galaxies

A problem in studying the interstellar magnetic field in the Milky Way Galaxy is that we are inside it, which allows detailed studies of the local interstellar medium but makes global study difficult. Observations of synchrotron emission from external galaxies provide a picture of the large-scale interstellar magnetic fields in spiral galaxies. Maps of the linear polarization of synchrotron radiation are maps of the ordered component of $B_\perp$. In general magnetic field lines show a strong spiral structure. Sometimes the polarization maps define spiral structure even more clearly than tracers of material spiral arms, although often the magnetic spiral arms are found offset from the material arms [2]. This result that B_u is stronger in the interarm regions than in the arms is actually due to B_u/B_r being higher in the interarm regions. A possible reason for this is that magnetic fields in the densest parts of spiral arms are somewhat tangled or random due to star formation or supernova activities, reducing the fractional polarization. Downstream from the spiral density-wave shock, gas will be accelerated away

from the shock in streaming motions, which will act to untangle magnetic field lines and increase the polarization. Hence, the direction of $B_\perp$ in spiral galaxies may be the best technique for study of these large-scale streaming motions in spiral galaxies.

Although the magnetic fields in spiral galaxies are symmetric, there does not seem to be a single symmetry mode. In some an axisymmetric spiral arm pattern dominates with the magnetic field direction everywhere directed inwards along the spiral pattern. Others have a bisymmetric mode, with reversals in the direction of magnetic fields from one arm to the next, as seen in the Milky Way Galaxy. However, only a few galaxies are classifiable in either way. The strong spiral magnetic patterns are probably best explained by dynamo action in spiral galaxies. In external spiral galaxies, $B_u \approx 1-5$ μG. A typical value for the total magnetic field strength in galaxies is $B_t \approx 10$ μG, with large uncertainties.

3.7 Recent Developments

We now describe preliminary results of a large H I Zeeman survey currently underway at Arecibo Observatory, and three recent developments concerning the large-scale diffuse Galactic mean magnetic field in the plane of the Galaxy ($z = 0$) and at large z, and the diffuse Galactic random magnetic field.

H I Zeeman

Heiles and Troland [29] are undertaking an extensive study of properties of the Galactic H I gas using the H I Zeeman technique. Such studies, involving single dish radio telescopes that sample both on- and off-source line profiles, can reveal magnetic field strengths and also other physical parameters of interest. These include the line excitation temperature T_s, the H I column density, and the non-thermal contribution to the line width. From these parameters, estimates can be made of the relative importance of thermal, non-thermal (i.e. turbulent), and magnetic energies in the cold neutral medium (CNM).

Observations toward extragalactic continuum sources reveal H I absorption and emission along random lines of sight, that is, along lines of sight not associated with morphologically identified H I structures in the Galaxy. Therefore, the ensemble of such observations provides a statistical measure of the physical properties of the CNM. In order to achieve the highest possible sensitivity and include the greatest number of sources, Heiles and Troland used the 300-m Arecibo telescope. They derived all Stokes parameters for on- and off-source profiles using a correlation technique [27].

Analysis of these data assumes that individual velocity components in the H I absorption profiles can be fit by Gaussians, each of which represents an isothermal and spatially distinct region of the CNM. Under this assumption, a least squares fitting process involving on- and off-source profiles reveals the parameters of each Gaussian component, including T_s and $N(H\ I)$. Moreover, the Stokes parameter V absorption profiles reveal the line-of-sight magnetic field

strength (or upper limit) in each Gaussian component via the H I Zeeman effect. A description of the fitting technique is given in Heiles [28], who discusses the statistics of T_s and other parameters, excluding the magnetic field. These studies confirm the highly supersonic nature of the CNM, revealing sonic mach numbers in the gas of 1.2 to 10, with most values between 2 and 6. This range is comparable to that found by Crutcher [11] in self-gravitating clouds previously observed for molecular Zeeman effects.

The magnetic field in the CNM appears quite weak. Analysis of the magnetic field data included 58 Gaussian components toward about half as many sources. For each of these components, the error in derived line-of-sight magnetic field strength $\sigma(B_\parallel) < 10$ μG. The average fitted value for $|B_\parallel|$ is 4.7 μG, and in most cases the field is not detected to a significance of 3-σ. These data imply that magnetic fields in the CNM are often < 5 μG. Curiously, the CNM magnetic field strength seems no higher than that inferred for the general Galactic ISM for which $\overline{n} \approx 0.5$ cm^{-3} [2]. It is possible that reversals in field direction along the lines of sight reduce the measured values of $B_\parallel$. This effect cannot be excluded; it must be present at some level. However, magnetic fields in the Galaxy are well ordered on many length scales, as indicated by linear polarization studies of the emission and extinction produced by dust grains. Therefore, it seems unlikely that Zeeman-effect studies of magnetic fields are severely affected by magnetic field reversals.

A key question about Galactic H I gas is the role of the magnetic field in its energetics. If the field energy exceeds that of the non-thermal motions (turbulence), then turbulence is sub-Alfvénic. Conversely, if the field is weak, then turbulence is super-Alfvénic, and it should dissipate rapidly. The intermediate case, where the two energies are comparable, is equipartition. Crutcher [11] found evidence for approximate equipartition in the relatively small sample of self-gravitating clouds with Zeeman effect measurements. His equipartition criterion is that the magnetic Alfvén number $m_A \approx 1$. To investigate this question in the CNM, Heiles and Troland compare $|B_\parallel|$ from the Zeeman effect measurements with the field strength B_{eq} expected for equipartition. This value in μG is approximately given by the relation $B_{eq} = 0.4\sqrt{n(H\ I)}\Delta v_{NT}$ [42]. Here Δv_{NT} is the non-thermal contribution to the line width (FWHM, km s^{-1}), obtained after the thermal contribution is removed.

Figure 2 shows $2B_\parallel$ vs. B_{eq}. (The factor of 2 is the statistical correction accounting for the fact that we measure only one component of the field.) B_{eq} was computed assuming $n(H\ I) = 100$ cm^{-3}. As expected from the generally low measured magnetic field values, the majority of cases have $2B_\parallel < B_{eq}$. For such CNM components, the turbulence is apparently super-Alfvénic, equivalent to $m_A > 1$. However, there are a significant number of CNM components for which the opposite conclusion holds, and the formal average of all derived values for the ratio $2B_\parallel/B_{eq} = m_A$ is close to unity. Although there may be a wide variation in the ratio of turbulent to magnetic energy in the CNM, this medium appears to be statistically close to equipartition.

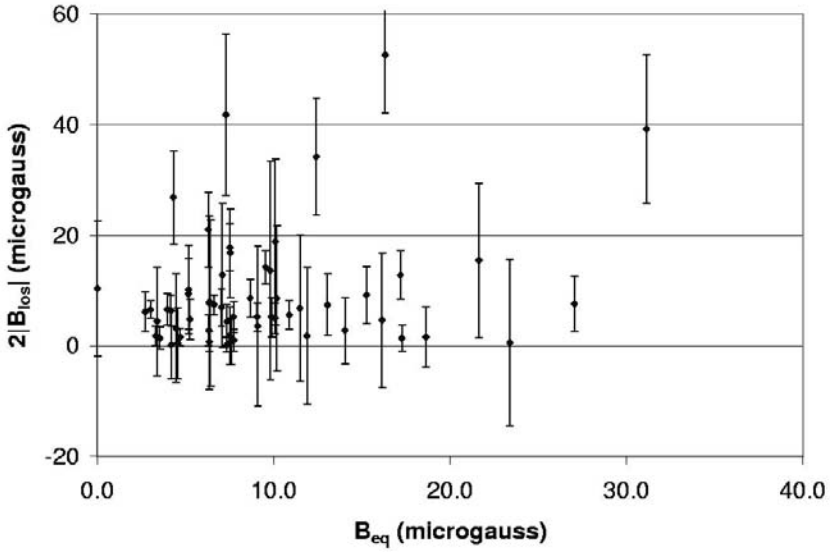


Fig. 2. Arecibo H I Zeeman results; here $B_{los} \equiv B_{\|}$

The Large-Scale Mean Field at $z = 0$

The large-scale mean field is obtained from two methods, Faraday rotation and starlight polarization. Faraday rotation gives a field inclination angle $\sim 8°$ (in the same sense as the spiral arms) [21], [32]. Analysis of optical starlight polarization yields a similar pitch angle and, also, the curvature of the field, with the center of curvature located ~ 9 kpc distant [24]. Whether the magnetic pitch angle is equal to that of the spiral arms is uncertain, both because the spiral arms are not very well defined and because the *real* errors on magnetic pitch angles are probably underestimated by standard error analysis procedures.

The new results [50] are from an analysis of a much expanded set of starlight polarizations (compendium by Heiles [26]). This analysis corrects a conceptual and procedural error in the earlier work and also derives important information about the random field, which we summarize below in our discussion of the random field. The improved analysis finds a larger pitch angle, $9.4° \pm 0.3°$. The larger sample of stars allows a more accurate determination of field curvature, which is consistent with the older one, but the new errors are much smaller. The center of curvature is 7.8 ± 0.5 kpc distant towards $\ell = -15.4° \pm 0.6°$.

The Large-Scale Mean Field at High z

Away from the Galactic plane, the Galactic field has been a confusing patchwork of seeming randomness. This is partly because extragalactic radio sources contain their own Faraday rotation, leading to randomness, and partly because the Galaxy itself has structure that differs from the simple $z = 0$ case. The

structure has quadrilateral symmetry in Galactic coordinates; the direction of the azimuthal field reverses from the southern to the northern hemispheres of the Galaxy. In addition, the vertical component of the Galactic magnetic field near the Sun points from the South to the North Galactic pole. This is shown dramatically and convincingly in the recent review by Han [22] (see his Fig. 3 and associated discussion), based on the original data fits and the same suggestion by Han et al. [20]. This field morphology (A0-mode), with a large-scale dipole field, is that which can be generated by a dynamo process. This result is clearly very important for understanding the origin and maintenance of the Galactic field.

The Random Field

It is well known that the random field exceeds the mean field by a considerable factor. Two recent works show that the random field is primarily longitudinal, i.e. that the component of random field that lies *parallel* to the mean field dominates the *perpendicular* component.

Brown and Taylor [4] use the Canadian Galactic Plane Survey data to measure Faraday rotations for a dense net of sources in the Galactic plane over a wide range of longitude, $\ell \sim 70^\circ$ to 150°. They find that the Rotation Measures (RMs) decrease as $\cos[\ell - (84^\circ \pm 4^\circ)]$, which is their result for the mean field direction. They also find that the *dispersions* in RM exhibit the same dependence. In other words, the dispersions in RM are proportional to RM. The dispersions are produced by the random component of the field. Hence, one concludes that the field fluctuations are primarily oriented along the mean field direction.

Walawender et al. [50] find a similar result from their analysis of the randomness of starlight polarization. The details are too much for the present discussion, so we give a very brief description. When looking along the mean field direction, there would be zero starlight polarization except for the existence of the random field. But, of course, the polarization is nonzero in this direction, and this provides an estimate of $\langle b^2_{r,\perp}\rangle \equiv \langle B^2_{r,\perp}\rangle/\langle B\rangle^2$, i.e. the ratio of the square of the perpendicular random component to the square of the mean field. Similarly, when looking perpendicular to the mean field one sees fluctuations in the polarization angle, which provides an estimate of $\langle b^2_{r,||}\rangle \equiv \langle B^2_{r,||}\rangle/\langle B\rangle^2$. In this way one obtains an estimate of the ratio of strengths of the two random field components. They find $\langle b^2_{r,||}\rangle/\langle b^2_{r,\perp}\rangle \approx 4.0$ (this number is subject to change with further analysis).

This result, that the random field component is primarily parallel to the mean field, is confirmed by two completely independent methods and, to our knowledge, has never been seen before. One can think of it as the mean field having lots of reversals with the field lines folding back on themselves. This has an important ramification for starlight polarization, which is sensitive only to the field *orientation* and not the direction: Random fields of this nature do not reduce the apparent starlight polarization! They do, however, reduce the apparent Faraday rotation, which *is* sensitive to direction. The large-scale Galactic

field is dominated by the random component, so clearly this result has important ramifications for theory.

4 Magnetic Fields and Star Formation

4.1 Theoretical Framework

Magnetic fields may play two crucial roles in the star formation process – they can provide support against gravitational collapse of dense interstellar clouds and they can solve the problem of how angular momentum is transported away from collapsing protostellar cores.

The crucial parameter for evaluating the importance of magnetic fields in star formation is the ratio of the mass to the magnetic flux, M/Φ_B. Theoretical values for M/Φ_B come from detailed computer simulations of interstellar clouds threaded by magnetic fields. The critical value of $M/\Phi_B \approx 1/\sqrt{63G}$ [40]. It is convenient to discuss observed M/Φ_B values in units of the critical value, such that $\lambda_B \equiv (M/\Phi_B)_{observed}/(M/\Phi_B)_{critical}$. Clouds that are initially magnetically supercritical ($\lambda_B > 1$) will collapse on relatively short time scales unless there is some other support mechanism. If the magnetic field was insufficient to stop the initial collapse, its compression during collapse cannot bring the cloud into equilibrium and halt the collapse. On the other hand, an initially magnetically subcritical ($\lambda_B < 1$) cloud will become unstable in its core due to ambipolar diffusion. Only ions are frozen into the magnetic field, while the neutral material (by far the majority of the mass) can contract gravitationally unaffected directly by the magnetic field. However, neutrals will collide with ions in this process. This frictional force will provide support against gravity to the neutrals, but there will be a drift of neutrals into the core without a significant increase in the magnetic flux; this is ambipolar diffusion. Eventually the core becomes supercritical, and dynamical collapse and star formation can proceed.

The mass to magnetic flux ratio can be determined empirically from the magnetic field strength and the column density. Since $M = 2.8N(H_2)m_H \times A$ and $\Phi_B = B \times A$, $\lambda_B = 2.8N(H_2)m_H\sqrt{63G}/B \approx 1.0 \times 10^{-20}N(H_2)/B$ cm$^2\mu$G, where A is the surface area of a cloud over which $N(H_2)$ and B are determined.

The role of magnetic fields may also be investigated with the virial theorem [39] (ignoring possibly important terms such as surface pressure): $\mathcal{W}+2\mathcal{T}+\mathcal{M} = 0$, where the gravitational, kinetic, and magnetic virial terms respectively are: $\mathcal{W} = \frac{3}{5}aGM^2/R$, $\mathcal{T} = \frac{3}{2}M\sigma^2$, and $\mathcal{M} = \frac{1}{3}b|\boldsymbol{B}|^2R^3$. For the simple case of a uniform density sphere, $a = b = 1$. A flattened, centrally condensed geometry will cause $a > 1$. For magnetically critical simulation models, $a \approx 1.2$ and $b \approx 0.3$.

4.2 Zeeman Results

Measurements of magnetic field strengths in molecular clouds depend on detection (and sometimes mapping) of the Zeeman effect in molecular lines (although when atomic hydrogen comes from photodissociation of H_2 near H II regions,

the H I line may trace molecular gas). Most molecular Zeeman detections and sensitive limits have been in the 18-cm ground-state Λ-doublet lines of OH. However, the OH radical probably does not trace densities above $n(H_2) \sim 10^4$ cm^{-3} in most cases [8], [48]. An exception is Λ-doublet OH lines that arise in excited rotational states, which require high gas densities for collisional excitation. Recently, detections of the Zeeman effect have been achieved toward the OMC1 and DR21OH cores in the 3-mm lines of CN [10], which sample densities $n(H_2) \sim 10^{5-6}$ cm^{-3} [49].

Crutcher et al. [9] carried out a sensitive survey of dark cloud cores. Only two of 12 cores were detected, but limits on $B_\parallel$ for the others allowed a meaningful test of whether M/Φ_B was magnetically critical. The statistical analysis found $\overline{B}_t \approx 16$ μG, which implied that the central regions were approximately magnetically critical. The data were found to be consistent with the hypothesis that dark clouds are in approximate virial equilibrium between magnetic and gravitational energy and that the supersonic line widths observed in dark clouds are the result of MHD motions such as Alfvén waves.

Crutcher [11] collected all information available by 1999 on Zeeman measurements in self-gravitating clouds and analyzed these data. There were 15 measurements of $B_\parallel$ in self-gravitating molecular clouds and an additional 12 sensitive upper limits, mainly the dark clouds discussed above. A statistical analysis of these Zeeman results yielded $\lambda_B \approx 1 - 2$. The result is dependent on the cloud geometry that is assumed. For spherical clouds, the observed column density can be used to compute λ_B without correction for geometry; the result found was $\lambda_B \approx 2$; that is, supercritical by ~ 2. However, if clouds are highly flattened (as theory of magnetically supported cores predicts), $N_{observed} = N_0/\cos\theta$, where N_0 (the column density along the minor axis of the core) should be used to evaluate λ_B. This extra $\cos\theta$ factor leads to $\lambda_B = 1$ being statistically consistent with the data. The analysis in terms of the virial theorem yields the result that kinetic and magnetic energies are approximately the same, and clouds are in approximate virial equilibrium.

The upper frame of Fig. 3 shows the data for Zeeman detections as plotted by Crutcher [11], $\log B_\parallel$ vs. $\log n(H_2)$. The data show a clear correlation (correlation coefficient = 0.84) of $B_\parallel$ with $n(H_2)$; the large scatter is due partially to the fact that only one component of $\boldsymbol{B}$ is observable. The actual fit is $B \propto n^\kappa$, $\kappa = 0.47 \pm 0.08$. For $n(H_2) > 10^3$ cm^{-3}, the data are consistent with $B \propto \sqrt{n}$, a result in agreement with the theoretical prediction $B \propto \sqrt{nT}$ for magnetically and thermally supported cores, since dense cores do not have a wide range of temperatures. All but one of the clouds with non-detection of the Zeeman effect, which are not plotted in Fig. 3, come from the dark cloud survey [9]. This survey had poor angular resolution ($18'$) and sampled predominantly regions with $n(H_2) < 10^3$ cm^{-3}. These dark cloud data are consistent with the $B \propto \sqrt{n}$ found for the Zeeman-detected cores at mainly higher density, but the scarcity of Zeeman detections at the lower densities does not permit a firm conclusion. There may be little or no increase in $|\mathbf{B}|$ with n for $n > 10^3$ cm^{-3}.

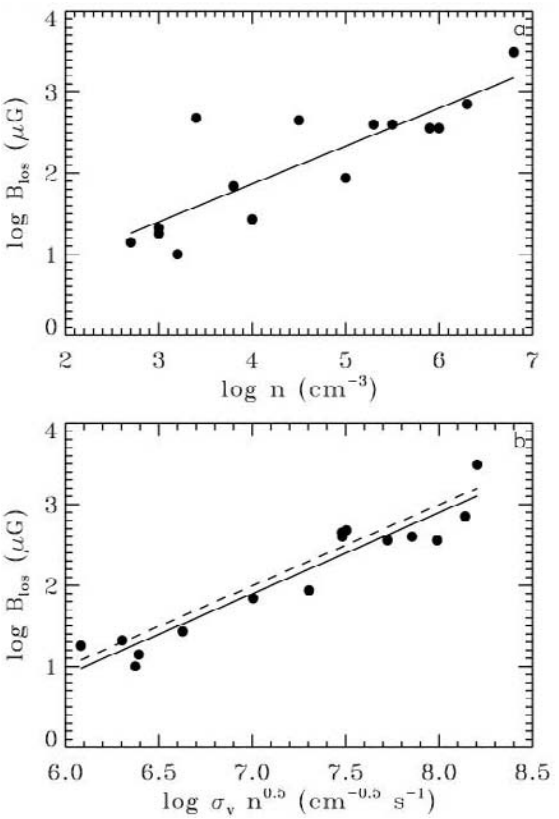


Fig. 3. Strength of $B_{\|}$ versus the gas density; the solid lines are least squares fits to the data, the dashed line is a model prediction

Basu [1] pointed out that the $\sqrt{T}$ term in the predicted $B \propto \sqrt{nT}$ expression comes from models with static magnetic and thermal support. If the observed supersonic motions are due to turbulence that provides a thermal-pressure like support, the prediction is $B \propto \sigma_v \sqrt{n}$, where σ_v is the turbulent velocity dispersion. The lower figure in Fig. 3 shows the data plotted this way. The solid line is the best fit, while the dotted line is for a theoretical prediction from the detailed models that assume static magnetic and thermal support [1]. Clearly, the scatter is much reduced; the correlation coefficient is 0.95. This agreement with the prediction is strong support for the theory of cloud support by a combination of static magnetic fields and MHD turbulence.

The Crutcher analysis [11] has been criticized since it used only the clouds with *detected* Zeeman signals. This was due to the desire to investigate the B vs. n relationship, which could not be done for upper limits to $B_{\|}$, and to the fact that almost all of the sensitive upper limits had already been analyzed separately [9]. For the density range which was the focus of the 1999 analysis, $n(H_2) > 10^3$ cm^{-3}, only one sensitive non-detection was excluded. This excluded result and

other insensitive Zeeman upper limits are consistent with the inferred results, so the Crutcher approach did *not* bias the conclusions.

Bourke et al. [3] carried out a new survey of 23 molecular clouds for the Zeeman effect in OH. The Zeeman effect was clearly detected in the cloud associated with the H II region RCW 38, with a field strength of $B_{\|} = 38 \pm 3$ μG, and possibly detected in a cloud associated with the H II region RCW 57, with a field strength of -203 ± 24 μG. The remaining 21 measurements give formal upper limits to the magnetic field strength, with typical 1-σ sensitivities less than 20 μG. For 22 of the molecular clouds they also determined the column density of the gas in which they made a sensitive search for the Zeeman effect. They carried out an analysis of the new results combined with previous Zeeman data for a comparison of theoretical models with the data. This study highlighted the importance of the large number of non-detections. The conclusion was that the typical cloud is approximately magnetically critical if clouds are modeled as highly flattened structures and if the true values of the field for the non-detections are close to the 3-σ upper limits. If instead these values are significantly lower (for example, similar to the 1-σ limits), then the typical cloud is generally magnetically supercritical. They noted that possible significant field structure within relatively large single-dish beams makes it difficult to distinguish magnetically subcritical and supercritical clouds. The results and conclusions of the Bourke et al. study were similar to those of Crutcher [11].

Clearly, the question of whether λ_B is subcritical or supercritical in pre-star formation cores has not yet been answered satisfactorily. In order to improve the situation, Troland and Crutcher have started a project to provide as definitive as possible a test of the proposition that molecular cloud evolution is controlled primarily by ambipolar diffusion, that is, $\lambda_B < 1$. The project is using the Arecibo telescope (for high angular resolution) to obtain sensitive OH Zeeman observations towards a statistically significant number (~ 30) of dark cloud cores. Because this survey requires several hundred hours of telescope time, it will take several years. One of the first results of this project was the detection of the Zeeman effect toward the L1544 dark cloud core [12] (Fig. 4). The signal-to-noise ratio in the Stokes V profiles is not high, but is typical of Zeeman observations. The fact that the same magnetic field strength is inferred from the Zeeman effect in two lines (the 1665 and 1667 MHz lines of OH) with Zeeman splitting factors Z differing by nearly a factor of two authenticates the reality of the magnetic field detection. The result is $B_{\|} = +10.8 \pm 1.7$ μG. The L1544 starless core has been observed to have infall motions and may be close to forming a star. Ciolek and Basu [7] computed a magnetically supported model for the L1544 core and showed that the data are consistent with the model. However, in order to fit the observations, the model does require that the magnetic field be nearly in the plane of the sky. The L1544 and other new Arecibo data have not yet been included in a general analysis of the type discussed above. When the Arecibo OH Zeeman survey is complete, a re-analysis of all the Zeeman data should provide a more definite answer to the question about the value of λ_B and hence about the role of magnetic fields in star formation.

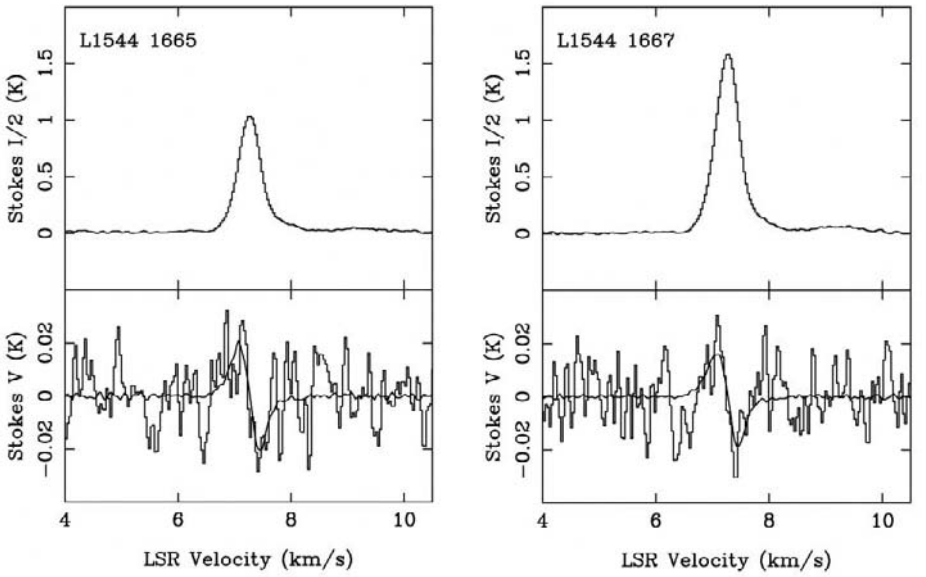


Fig. 4. Arecibo OH Zeeman profiles toward L1544. Histogram plots are the data; $dI/d\nu$ is superposed on the observed V as a line plot

Finally, Watson et al. [51] developed statistical relationships between the rms values of the irregular component of the magnetic field and spatial variations in the circular polarization of the spectral line radiation. They then analyzed an H I Zeeman map toward Orion A in order to infer the strength of the irregular component of the magnetic field. They found that the rms of the irregular component is comparable in magnitude to the mean magnetic field within the cloud. Hence, the turbulent and Alfvén velocities should also be comparable.

4.3 Field Morphology

Observations of the field morphology provide significant empirical input to and tests of theories. Background starlight polarization observations show that large-scale fields in clouds are smooth, with little change in the vicinity of clumps or cores. Near-infrared observations make it possible in principle to probe up to $A_V \sim 40$ magnitudes. However, the polarization does not seem to increase much beyond about $A_V \approx 1 - 2$ magnitudes, which suggests that the grains in dense clouds do not produce polarization and hence cannot be used to map magnetic field morphologies at large extinctions [19]. However, Wiebe and Watson [52] propose a different explanation. They calculate the linear polarization of starlight due to extinction by aligned dust grains when the starlight traverses a medium with irregular magnetic fields. This medium is intended to represent the optically thick components of interstellar clouds that are observed to make little, if any, contribution to the polarization of starlight. In agreement with the observations,

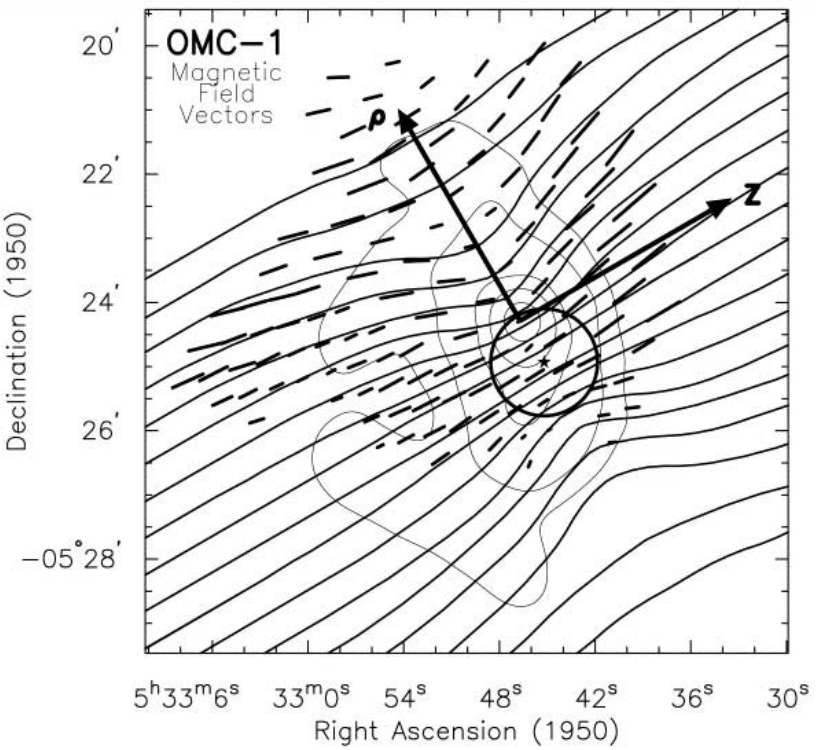


Fig. 5. Magnetic field map of OMC1 from 100 μm dust polarization mapping. Polarization vectors have been rotated 90°, so the short line segments represent the magnetic field direction at each point. Contours show the total intensity of the dust emission. A model (hourglass) magnetic field morphology is shown; z is the symmetry axis of the hourglass and ρ is the radial axis. The symmetry axis of the model field is at the superposed star, 0.1 pc southwest of KL, which is at the emission intensity peak. The small circle shows the radius of the OMC1 molecular ridge

they found that the polarization properties of the starlight can be essentially unchanged. For this, the rms of the irregular component must be greater than the average magnetic field, which in turn tends to imply that the turbulent velocities in these interstellar clouds are super-Alfvénic.

In any case, significant polarization of thermal dust emission from dense cores is seen at far infrared and millimeter wavelengths observations, which have proven to be the best technique for probing magnetic field morphology in dense interstellar clouds. Since the emission is generally optically thin, this technique samples material throughout dense molecular gas. However, sensitivity limits this technique to $A_V > 20$ magnitudes. The maps show that the polarization directions in dense molecular clouds are generally fairly uniform, at least on large spatial scales. Figure 5 [46] shows far infrared polarization observations of $B_\perp$ in Orion Molecular Cloud 1 (OMC1). The typical percentage polarization is $\sim 2\%$,

although values up to $\sim$ 10% are observed. Analysis of the OMC1 data [86] suggests that there is an "hourglass pinch" to the field lines, as expected toward dense cores on theoretical grounds. At the higher angular resolution available with mm-wave arrays, more structure is seen [45]. In particular, near the dense central core there is a change in position angle of almost 90°, which may be due to a twisted magnetic field or to Gold alignment in the outflow from the core. But the main geometry of the field is perpendicular to the minor axis of the cloud, which is expected if the magnetic field is providing significant support against gravity.

Maps of polarized dust emission toward filamentary clouds [37] show evidence for a helical magnetic field structure. Toward OMC3 the polarization pattern does not appear to be consistent with a uniform field local to the OMC3 filament or with the field orientation in the dense OMC1 core to the south. Depolarization is observed along the length of the filament and is not limited to the dense cores within the filament, which suggests that the cause of the depolarization is not simply a lack of polariz ing grains in high-density regions. Matthews et al. [37] successfully model the observed polarization pattern in terms of a helical magnetic field geometry.

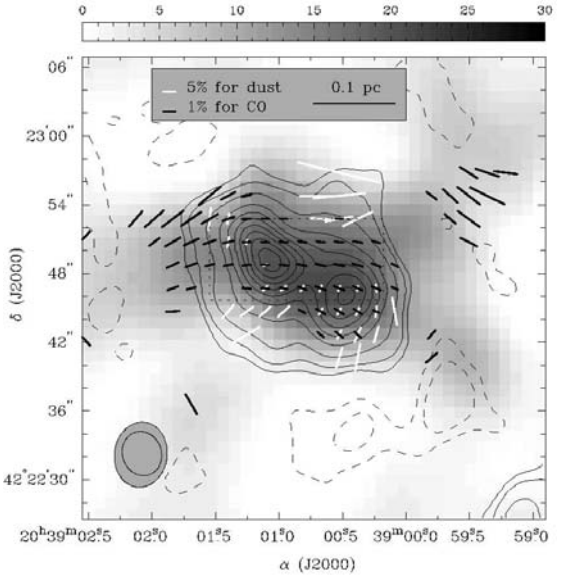


Fig. 6. BIMA 1.3 mm dust and CO maps of DR21OH. Contours (-30, -18, 18, 30, 60, 120, 180, 240, 300, 360, 420, 480 mJy beam^{-1}) show total dust emission. The grey scale shows the integrated intensity of CO (Jy beam^{-1} km s^{-1})

The BIMA millimeter-wave array has recently been used to map polarized dust emission at high angular resolution. One result [34] is for DR21OH (Fig. 6), which is located 3′ north of the H II region DR21 in the Cygnus X molecular

cloud/H II region complex at a distance of about 3 kpc. Its association with masers of OH, H_2O, and CH_3OH suggests the presence of high-mass, young stellar objects. Unlike DR21, DR21OH does not show extended outflows and shocked molecular hydrogen, which further indicates that DR21OH is in an earlier stage of evolution than DR21. The main DR21OH component has been resolved into two compact cores, MM1 and MM2 [54]; their masses are about 100 $M_\odot$. The line-of-sight magnetic field strength, $B_\parallel$, was determined from CN Zeeman detections to be -0.4 mG for MM1 and -0.7 mG for MM2 [10]. CN traces high density molecular gas ($n(H_2) \sim 10^5 - 10^6$ cm^{-3}).

The Goldreich-Kylafis effect predicts that the CO polarization direction is either parallel or perpendicular to the magnetic field direction. The observations are consistent with the theoretical predictions: near the northern edge, ϕ_{CO} is approximately parallel to ϕ_{dust}; near the MM1 and MM2 cores, the CO polarization vectors are close to perpendicular to the dust polarization vectors (average $\phi_{CO} - \phi_{dust} = 95°$).

Taking the position angle dispersion from the CO observations, $\delta\phi \sim 22°$, and adopting the density (10^6 cm^{-3} for MM1 and 2×10^6 cm^{-3} for MM2) and linewidth (2.3 km s^{-1}) from CN [11], Lai [34] derived $B_\perp \sim 0.9$ mG for MM1 and ~ 1.3 mG for MM2. Hence, $B_\perp \approx 2B_\parallel$, which taken literally would imply that **B** is at an angle of $\sim 30°$ to the plane of the sky. These fields yield $\lambda_B \approx 2.8$ and 4.7 respectively, which imply supercritical cores, and $\beta_{turb} \approx 0.4$.

One apparently ubiquitous feature of polarized dust emission maps is that the percentage polarization is anticorrelated with the total intensity of the emission. As an example, the data for DR21OH [34] are shown in Fig. 7.

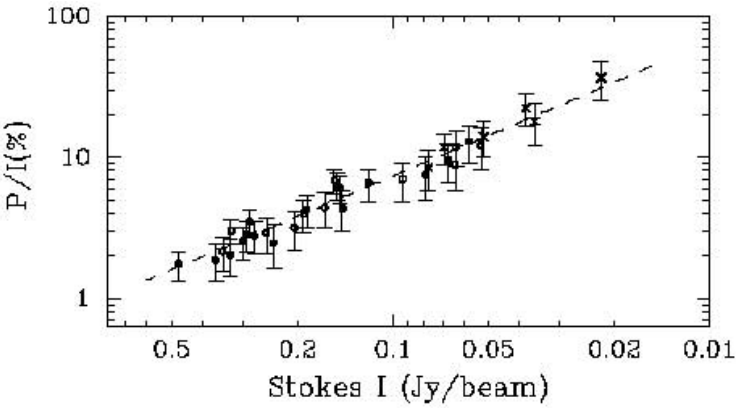


Fig. 7. Depolarization with intensity toward DR21OH

The W51 e1 and W51 e2 molecular cores have also been observed with BIMA [34]. The cores have masses of ~ 150 and 110 $M_\odot$ and are located on the eastern edge of the W51 molecular cloud/H II region complex. The average directions of the magnetic fields are approximately parallel to the minor axis of the two cores,

suggesting that matter collapsed along the field lines to form the parent cloud of the double cores. The observed polarization angle dispersions in W51 e1 and e2 are equal to the measurement uncertainty. The upper limits for $\delta\phi$ yield lower limits for the average strengths of the magnetic field, $B_\perp(e1) > 0.3$ mG and $B_\perp(e2) > 1.3$ mG. These fields imply upper limits for the mass-to-flux ratios of $\lambda_B(e1) < 3.1$ and $\lambda_B(e2) < 4.9$; that is, possibly supercritical. The turbulent to magnetic energy ratio $\beta_{turb} < 0.06$, implying that the magnetic field dominates turbulence.

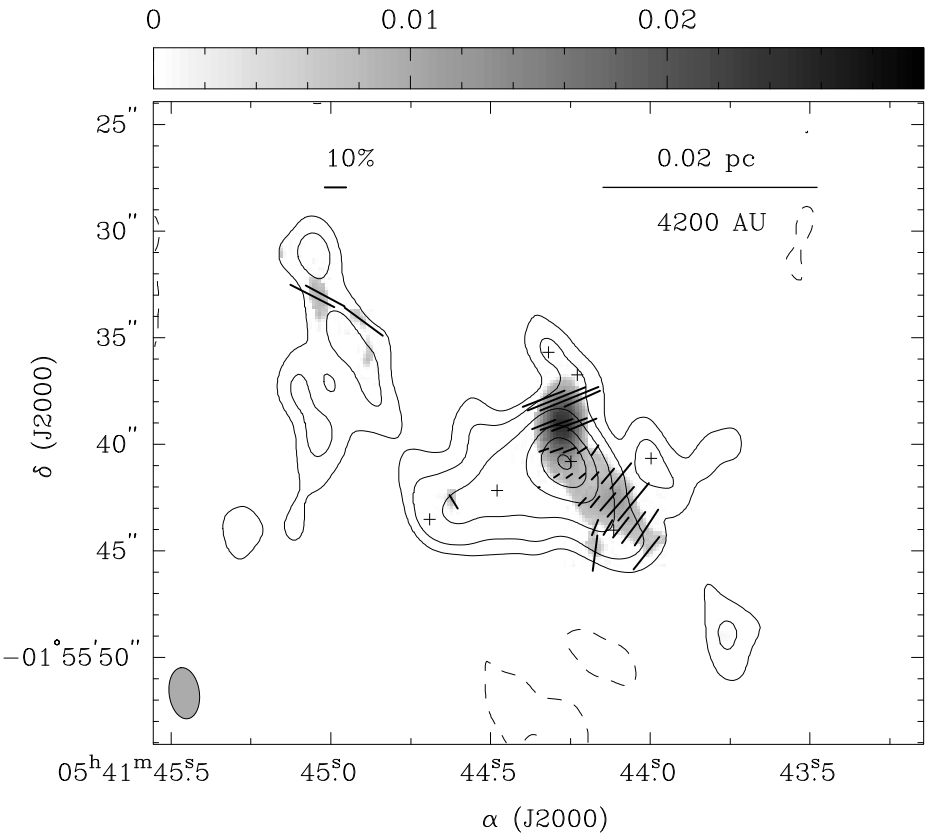


Fig. 8. BIMA polarization results for NGC 2024 FIR 5. The 1.3-mm Stokes I dust emission is shown as contours (at -5, -3, 3, 5, 10, 20, 30, and 40 σ levels, with the 1 σ noise level = 7.3 mJy beam^{-1}. The polarized emission is show as grey scale; the calibration scale at the top is in Jy beam^{-1}. Percentage polarization vectors are superposed

A third recent BIMA observation is for NGC 2024 FIR 5 [34]; the variation of the polarization angles suggests that the "uniform" magnetic field lines are curved (Fig. 8). The morphology is suggestive of an hourglass, with the left half missing due to lack of polarization sensitivity. The lower limit for $B_\perp$, using

the position angle dispersion before taking out the curvature of the field ($\delta\phi = 13.1^\circ \pm 4.6^\circ$), is $B_{\perp,min}$ ∼1.9 mG. This conservative estimate of the dispersion in ϕ leads to a small turbulent to magnetic energy ratio, $\beta_{turb} < 0.14$. Therefore, the magnetic field appears to dominate the turbulence in the core region of NGC 2024 FIR 5.

Finally, Padoan et al. [44] have suggested that the phenomenon of depolarization of dust emission toward cores is due to grains in high-density cores not producing polarization, because of a change in grain properties. Specifically, they proposed that grains in regions with $A_V > 3$ magnitudes produce only unpolarized emission, and that an envelope region surrounding a core produces polarization up to 15%. A core embedded in this envelope would add unpolarized emission, so that the polarized map would show an anticorrelation of percentage polarization with total intensiy.

There are several reasons to question this conclusion. Perhaps the strongest is that very strong polarization is seen in the interferometric maps, such as those of DR21OH, W51, and NGC 2024. An interferometer is a spatial filter, sensitive only to small-scale structure, so interferometer mapping does not detect emission, polarized or unpolarized, on large spatial scales. In the case of the BIMA maps, that scale is typically $\sim 20''$. Hence, the maps show polarized and unpolarized emission only from regions with small spatial structure – the cores themselves. Furthermore, maps of the total polarized flux $\sqrt{Q^2 + U^2}$ do not support the suggestion that the polarized flux comes only from diffuse envelope regions. Figure 8 shows in grey scale the total polarized flux toward NGC 2024. The strongest peak is just north of the total intensity core. However, there is no polarized flux detected (to the sensitivity limit) over most of the map, including the area to the east of the core where the total intensity is quite strong. This total polarized flux map does not look like a map of a low density envelope region surrounding the dense core. In addition, Matthews et al. [37] argued that the depolarization seen in OMC3 was not correlated with the dense cores, but appeared to be explainable by the field morphology – a helical field. Toward DR21OH the same phenomenon is seen. The percentage polarization decreases from south to north to a minimum on the two cores (see Fig. 6), but then does not increase immediately north of the cores but declines further to below the sensitivity limit. Still further to the north, the polarization reappears strongly, but the position angles are substantially different, suggesting a morphological explanation. We therefore believe that the millimeter and submillimeter wavelength polarization maps do indeed probe magnetic fields in dense cores.

5 Conclusions

There are a number of observational tools that make it possible to study interstellar magnetic fields on all scales. Although the various probes sample different phases and regions of the interstellar medium, a fairly consistent picture has emerged. Here we give a summary of the conclusions, with a notation of the section in this paper where the conclusion is discussed.

- The Galactic interstellar magnetic field has a spiral pattern, with reversals in field direction in interarm regions. Near the sun $B_t \approx 6$ μG (or perhaps less) (Sect. 3.2). The field strength increases inward in the Galaxy (Sect. 3.2), with milligauss fields in the Galactic center (Sect. 3.5). The Galactic magnetic field is probably maintained by a Galactic dynamo, although the details are very unclear.
- The Galactic field has a ratio of uniform to random components $B_u/B_r \sim 0.5$ (Sects. 3.1, 3.2, 3.3, 3.7). Recent evidence suggests that the random component is primarily parallel to the mean field (Sect. 3.7).
- In diffuse H I interstellar clouds, $B_{\|} \sim 10$ μG, with wide cloud-to-cloud variations. The magnetic field strength appears to increase little with gas density up to $n \sim 10^3$ cm^{-3}, although observational sensitivity limits the strength of this conclusion (Sects. 3.4, 3.7).
- At higher densities, $n(H_2) > 10^3$ cm^{-3}, the data are consistent with $B \propto \sigma_v\sqrt{n}$, which is the theoretical prediction for flattened cores supported by a combination of a uniform magnetic field pressure and turbulence. In cores, motions are approximately Alfvénic, and $\lambda_B \approx 1-2$ (a critical to slightly supercritical mass-to-magnetic flux ratio). (Sect. 4.2).
- The ratio of B_r/B_u appears to decline with density. At low densities the random field dominates, at high densities in molecular cores the uniform field dominates (Sects. 3.1, 3.2, 3.3, 3.7, 4.2, 4.3).
- In some molecular cores there is evidence for the "hourglass" pinch that would be produced by cloud contraction with the magnetic field frozen into the matter (Sect. 4.5).

Acknowledgments

It is with great pleasure that C. Heiles acknowledges helpful discussions with Ellen Zweibel and Josh Walawender. This work was supported in part by NSF grants AST-9530590, AST-0097417, AST-9820641, and AST-9981363.

References

1. S. Basu: ApJ **540**, L103 (2000)
2. R. Beck: Space Science Reviews **99**, 243 (2001)
3. T. L. Bourke, P. C. Myers, G. Robinson, A. R. Hyland: ApJ **554**, 916 (2001)
4. J. C. Brown, A. R. Taylor: ApJ **563**, L31(2001)
5. C. Chandler, J.E. Carlstrom: ApJ **466**, 338 (1996)
6. S. Chandrasekhar, E. Fermi: ApJ **118**, 113 (1953)
7. G. E. Ciolek, S. Basu: ApJ **529**, 925 (2000)
8. R. M. Crutcher: ApJ **234**, 881 (1979)
9. R.M. Crutcher, T. H. Troland, A. A. Goodman, C. Heiles, I. Kazes, P. C. Myers: ApJ **407**, 175 (1993)
10. R. M. Crutcher, T. Troland, B. Lazareff, G. Paubert, I. Kazes: ApJ **514**, l121 (1999)
11. R.M. Crutcher: ApJ **520**, 706 (1999)

12. R. M. Crutcher, T. H. Troland: ApJ **537**, L139 (2000)
13. J. Davis, J. L. Greenstein: ApJ **114**, 206 (1951)
14. B. Draine: 'Electromagnetic Properties of Grains Related to Grain Alignment'. In: *Polarimetry of the Interstellar Medium, A.S.P. Conference Series Volume 97*, ed. By W.G. Roberge, D.C.B. Whittet (ASP, San Francisco 1996) pp. 16
15. J.D. Fiege, R.E. Pudritz: MNRAS **311**, 85 (2000)
16. J.D. Fiege, R.E. Pudritz: MNRAS **311**, 105 (2000)
17. T. Gold: Nature **169**, 322 (1951)
18. P. Goldreich, N.D. Kylafis: ApJ **243**, 75 (1981)
19. A. A. Goodman, T. J. Jones, E.A. Lada, P. C. Lada: ApJ, **448**, 748 (1995)
20. J.L. Han, R.N. Manchester, E.M. Berkhuijsen, R. Beck: A & Ap, **322**, 98 (1997)
21. J.L. Han, R.N. Manchester, G.J. Qiao: MNRAS **306**, 371 (1999)
22. J.L. Han: ' Magnetic fields in our Galaxy: How much do we know? (II) Halo fields and the flobal field structure'. In: *Astrophysical Polarized Backgrounds, American Institute of Physics Conference Proceedings 609*, ed. by S. Cecchini, S. Cortiglioni, R. Sault, C. Sbarra (AIP, 2002), pp. 96
23. C. Heiles, W. T. Reach, B. Koo: ApJ, **466**, 191 (1996)
24. C. Heiles: ApJ, **462**, 316 (1996)
25. C. Heiles: 'A Comprehensive View of the Galactic Magnetic Field, Especially near the Sun'. In: *Polarimetry of the Interstellar Medium, A.S.P. Conference Series Volume 97*, ed. By W.G. Roberge, D.C.B. Whittet (ASP, San Francisco 1996) pp. 457
26. C. Heiles: AJ, **119**, 923 (2000)
27. C. Heiles: PASP **113**, 788 (2001)
28. C. Heiles: ApJ **551**, L105 (2001)
29. C. Heiles, T. Troland: to be submitted to ApJ (2002)
30. F. Heitsch, E.G. Zweibel, M.-M. MacLow, P.S. Li, M.L. Norman: ApJ **561**, 800 (2001)
31. R. H. Hildebrand, J. A. Davidson, J. L. Dotson, C. D. Dowell, G. Novak, J. E. Vaillancourt: PASP, **112**, 1215 (2000)
32. C. Indranin, A.A. Deshpande: New Astronomy, **4**, 33 (1998)
33. N.D. Kylafis: ApJ **275**, 135 (1983)
34. S.-P. Lai, University of Illinois Ph.D. thesis (2001)
35. A. Lazarian: 'Gold Alignment'. In: *Polarimetry of the Interstellar Medium, A.S.P. Conference Series Volume 97*, ed. By W.G. Roberge, D.C.B. Whittet (ASP, San Francisco 1996) pp. 433
36. A. Lazarian, A. A. Goodman, P. C. Myers: ApJ **490**, 273 (1997)
37. B. C. Matthews, C. D. Wilson, J. D. Fiege: ApJ **562**, 400 (2001)
38. C.F. McKee: 'The Dynamical Structure and Evolution of Giant Molecular Clouds'. In: *The Origin of Stars and Planetary Systems*, ed. by C.J. Lada, N.D. Kylafis (Kluwer, Dordrecht 1999) pp. 29-66
39. C. F. McKee, E. G. Zweibel: ApJ **399**, 551 (1992)
40. T.Ch. Mouschovias, L. Spitzer, Jr.: ApJ **210**, 326 (1976)
41. T.Ch. Mouschovias, G.E. Ciolek: In: *The Origin of Stars and Planetary Systems*. ed. by C.J. Lada & N.D. Kylafis (Kluwer, Dordrecht 1999) pp. 305-340
42. P. Myers, A.A. Goodman: ApJ **329**, 392 (1988)
43. E.C. Ostriker, J.M. Stone, C.F. Gammie: ApJ **546**, 980 (2001)
44. P. Padoan, A. Goodman, B. T. Draine, M. Juvela, A. Nordlund, O. E. Rognvaldsson: ApJ **559**, 1005 (2001)
45. R. Rao, R. M. Crutcher, R. L., Plambeck, M. C. H. Wright: ApJ **502**, L75 (1998)

46. D.A. Schleuning: ApJ **493**, 811 (1998)
47. F.H. Shu, A. Allen, H. Shang, E.C. Ostriker, Z.-Y. Li: In: *The Origin of Stars and Planetary Systems*. ed. by C. J. Lada & N. D. Kylafis (Kluwer, Dordrecht 1999) pp. 193-226
48. T. H. Troland, R. M. Crutcher, A. A. Goodman, C. Heiles, I. Kazes, P. C. Myers: ApJ **471**, 302 (1996)
49. B. E. Turner, R. H. Gammon: ApJ **198**, 71 (1975)
50. J.M. Walawender, E.G. Zweibel, C. Heiles: ApJ, in preparation (2002)
51. W. D. Watson, D. S. Wiebe, R. M. Crutcher: ApJ **549**, 377 (2001)
52. D. S. Wiebe, W. D. Watson: ApJ **549**, L115 (2001)
53. D.A. Weintraub, A.A. Goodman, R.L. Akeson In 'Protostars and Planets IV'. ed. V. Mannings, A.P. Boss & S.S. Russell (University of Arizona Press), in press
54. D. P. Woody, S. L. Scott, N. Z. Scoville, L. G. Mundy, A. I. Sargent, S. Padin, C. G. Tinney, C. D. Wilson: ApJ **337**, L41 (1989)
55. E.G. Zweibel, C. Heiles: Nature **385**, 131 (1997)

MHD Turbulence in Star-Forming Regions and the Interstellar Medium

Mordecai-Mark Mac Low

Dept. of Astrophysics, American Museum of Natural History, 79th Street and Central Park W., New York, NY, 10024-5192, USA;
e-mail: mordecai@amnh.org

Abstract. MHD turbulence plays a central role in the physics of star-forming molecular clouds and the interstellar medium. MHD turbulence in molecular clouds must be driven to account for the observed supersonic motions in the clouds, as even strongly magnetized turbulence decays quickly. Driven MHD turbulence can globally support gravitationally unstable regions, but local collapse inevitably occurs. Differences in the strength of driving and the gas density may explain the very different rates of star formation observed in different galaxies. Two types of comparisons to observations are reviewed. First, the use of wavelet transform methods suggests that the driving comes from scales larger than observed molecular clouds. Second, comparison of simulated spectral cubes from models to real observations suggests that Larson's mass-size relationship is an observational artifact. The driving mechanism for the turbulence is likely a combination of field supernovae in star-forming sections of galactic disks, and magnetorotational instabilities in outer disks and low surface brightness galaxies. Supernova-driven turbulence has a broad range of pressures with a roughly log-normal distribution. High-pressure, cold regions can be formed even in the absence of self-gravity.

1 Introduction

One of the big questions in star formation is what determines the rate of star formation in galaxies? Another, more pointed way of phrasing this question is to ask why the star formation rate in normal galaxies is so low, and why it varies so strongly, over orders of magnitude from low surface brightness galaxies, through normal galaxies, to starburst galaxies. The free-fall time for gas at typical interstellar densities is

$$t_{\rm ff} = \sqrt{\frac{3\pi}{32G\bar{\rho}}} \approx (3.4 \times 10^7 \text{ yr}) \left(\frac{n}{1 \text{ cm}^{-3}}\right)^{-1/2}, \quad (1)$$

where $\bar{\rho}$ is the mean mass density of the gas, G the gravitational constant and $n = \bar{\rho}/\mu$ the number density, with $\mu = 2.36 m_H$. Yet galactic ages range up to 10^{10} yr, and star formation continues today. What has delayed star formation sufficiently to allow it to continue?

In what might be called the standard theory of star formation, magnetic fields are invoked to answer both of these questions. If fields are strong enough, they can magnetostatically support clouds against collapse. The star formation

E. Falgarone and T. Passot (Eds.): LNP 614, pp. 182–212, 2003.
© Springer-Verlag Berlin Heidelberg 2003

rate would then be determined by the rate of ambipolar drift of neutral gas past ions tied to the magnetic field towards the centers of self-gravitating cores [62,79]. Furthermore, if the fields are strong enough that the Alfvén speed v_A reaches the rms velocity v, then strong shocks will be converted to MHD waves. As linear Alfvén waves are lossless, it was thought that motions remaining from the initial formation of the clouds might be enough to explain the observation of strongly supersonic motions in molecular clouds [2]. In this review I will explain why both of these ideas now appear questionable.

2 Decaying Turbulence

First let us consider the decay of supersonic turbulence as shown in Fig. 1. I call on computations performed with two different methods: Eulerian hydrodynamics and MHD on a grid, using the code ZEUS-3D [83,84,27,36], available from the Laboratory for Computational Astrophysics at

http://zeus.ncsa.uiuc.edu/lca_home_page.html,

and Lagrangian hydrodynamics using a smoothed particle hydrodynamics (SPH) code derived from that described by Benz [21] and Monaghan [61], running on special purpose GRAPE processors [28,82], and incorporating sink particles [17].

We chose initial conditions for our models inspired by the popular idea that setting up velocity perturbations with an initial power spectrum $P(k) \propto k^{\alpha}$ in Fourier space similar to that of developed turbulence would be in some way equivalent to starting with developed turbulence [68,73,74]. Observing the development of our models, it became clear to us that the loss of phase information in the power spectrum [3] allows extremely different gas distributions to have the same power spectrum. This is particularly important for supersonic flows. Supersonic, HD turbulence has been found in simulations [74] to have a power spectrum $\alpha = -2$. However, any single, discontinuous shock wave will also have such a power spectrum, as that is simply the Fourier transform of a step function, and taking the Fourier transform of many shocks will not change this power law. Nevertheless, most distributions with $\alpha = -2$ do not contain shocks.

After experimentation, we decided that the quickest way to generate fully developed turbulence was with perturbations having a flat power spectrum $\alpha = 0$ for $0 < k_d < 8$. We set up velocity perturbutions drawn from a Gaussian random field fully determined by its power spectrum in Fourier space following the standard procedure: for each wavenumber $\mathbf{k_d}$ we randomly select an amplitude from a Gaussian distribution centered on zero and with width $P(k_d) = P_0 k_d^{\alpha}$ with $k_d = |\mathbf{k_d}|$, and a phase between zero and 2π. We then transform the field back into real space to obtain the velocity in each zone. This is done independently for each velocity component. For the SPH calculation the velocities defined on the grid are assigned onto individual particles using the "cloud-in-cell" scheme [41]. In all of our models we take $c_{\rm s} = 0.1$, initial density $\rho_0 = 1$, and we use a periodic grid with sides $L = 2$ centered on the origin. These parameter choices define our unit system. Our choice of periodic boundary conditions corresponds

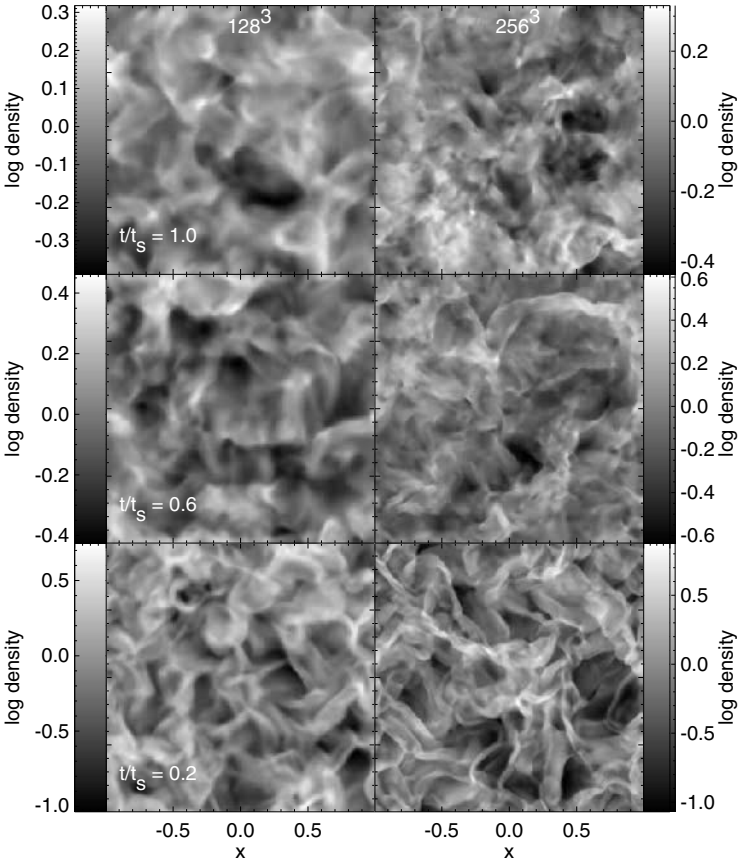


Fig. 1. Log of density is shown at times $t/t_s = 0.2$, 0.6 and 1.0 on slices through the decaying supersonic hydrodynamic models C and D with initial rms Mach number $M = 5$ described in [55] at standard resolution (128^3) and high resolution (256^3), where t_s is the sound crossing time of our numerical box. Note that each image is scaled to its own maximum and minimum to enhance morphological features (From [52])

to the case of free turbulence discussed above, at least initially. Thereafter, the appropriate treatment is less clear.

We next performed resolution studies using ZEUS for three different cases with no field, weak field and strong field as shown in Fig. 2. The weak field models have an initial ratio of thermal to magnetic pressure $\beta = 2$, while the strong field models have $\beta = 0.08$. We ran the same unmagnetized model with the SPH code to demonstrate that our results are truly independent of the details of the viscous dissipation, and so that our lack of an explicit viscosity does not affect our results.

The kinetic energy decay curves for the four resolution studies are shown in Fig. 2. For each of our runs we performed a least-squares fit to the slope of the power-law portion of the kinetic energy decay curves. These results appear

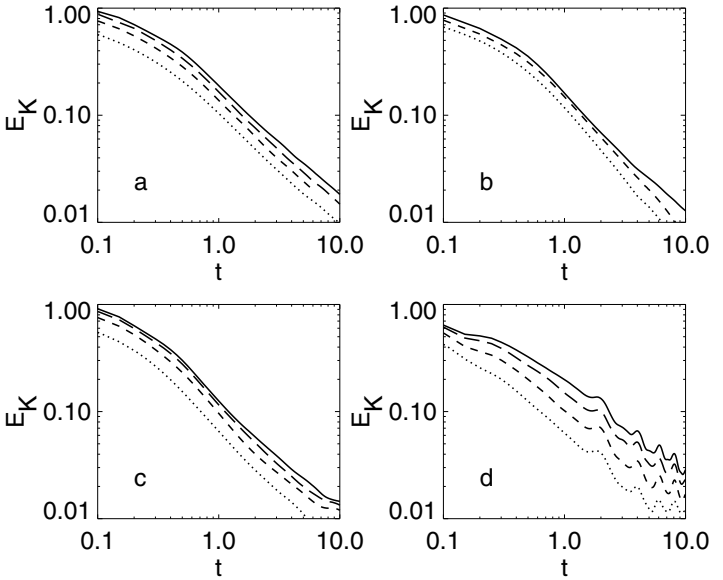


Fig. 2. Three-dimensional resolution studies for initial rms Mach number $M = 5$, isothermal models. ZEUS models have 32^3 (*dotted*), 64^3 (*short dashed*), 128^3 (*long dashed*), or 256^3 (*solid*) zones, while the SPH models have 7000 (*dotted*), 50,000 (*short dashed*), or 350,000 (*solid*) particles. Panels show *a)* hydro runs with ZEUS, *b)* hydro runs with SPH, *c)* initial rms Alfvén number $A = v/v_A = 5$ MHD runs with ZEUS, and *d)* $A = 1$ MHD runs with ZEUS (From [55])

converged at the 5–10% level; it is very reassuring that the different numerical methods converge to the same result for the unmagnetized case.

We find that highly compressible, isothermal turbulence decays close to linearly in time, with $\eta = 0.98$. Adding magnetic fields decreases the decay rate only slightly to $\eta \sim 0.85$–0.9, with very slight dependence on the field strength or adiabatic index. Similar results have been reported by Stone, Ostriker, & Gammie [85] and Padoan & Nordlund [68] for compressible MHD turbulence, and by Biskamp & Müller [22] for incompressible MHD turbulence.

The clear astrophysical implication of these models is that even strong magnetic fields, with the field in equipartition with the kinetic energy, cannot prevent the decay of turbulent motions on dynamical timescales. If molecular clouds live for longer than their dynamical time of roughly a megayear, as is generally believed even by those arguing for lifetimes under 10 Myr [6,35] rather than the more classical 30 Myr [23], then the significant kinetic energy observed in their gas must be supplied more or less continuously.

3 Driven Turbulence

To compute the energy dissipation from uniformly driven turbulence we initialize the turbulent flow with a narrow band of k values, using a top-hat function with roughly the same behavior as the steeply peaked curve used by Stone et al. [85] in most of their models. We set the velocity field up as described above [55]. To drive the turbulence, we then normalize this fixed pattern to produce a set of perturbations $\delta\boldsymbol{\nu}(x, y, z)$, and at every time step add a velocity field $\delta\boldsymbol{v}(x, y, z) = A\delta\boldsymbol{\nu}$ to the velocity $\boldsymbol{v}$, with the amplitude A now chosen to maintain constant kinetic energy input rate $\dot{E}_{\rm in} = \Delta E/\Delta t$. For compressible flow with a time-dependent density distribution, maintaining a constant energy input rate requires solving a quadratic equation in the amplitude A at each time step. For a grid with N zones on a side, each of volume ΔV, the equation for A is

$$\Delta E = \frac{1}{2}\Delta V \sum_{i,j,k=1}^{N} \rho_{ijk} A\delta\boldsymbol{\nu}_{ijk} \cdot (\boldsymbol{v}_{ijk} + A\delta\boldsymbol{\nu}_{ijk}). \tag{2}$$

We take the larger root of this equation to get the value of A. The resulting flow is shown in Fig. 3.

We find that the best description of these compressible models comes by taking a length scale $\mathcal{L} = \lambda_d$ the driving wavelength, and a velocity scale $\mathcal{V} = v_{\rm rms}$ the rms velocity, rather than any of the other options available. Happily, these are just the length and velocity scales that would be expected from the theory of *incompressible* turbulence. Figure 4 shows equilibrium energy dissipation rates for all the models described by Mac Low [52], compared to the quantity $k_d v_{\rm rms}^3 \sim v_{\rm rms}^3/\lambda_d$. A fit to the hydrodynamic models gives a relation with slope 1.02. Let us define a dimensionalized wavenumber $\tilde{k} = (2\pi/L)k_d = 2\pi/\lambda_d$. A very good approximation is then the linear relation

$$\dot{E}_{\rm kin} \simeq -\eta_v m \tilde{k} v_{\rm rms}^3, \tag{3}$$

with $\eta_v = 0.21/\pi$, where the assumption is made that in equilibrium $\dot{E}_{\rm kin} = \dot{E}_{\rm in}$. The dependence on the mass of the cube m comes strictly from dimensional arguments, as all of the runs shown have the same mass $m = \rho_0 L^3 = 8$. The referee of this review points out that $\eta_v k_d \simeq 0.42\lambda_d^{-1}$, remarkably close to the value of $1/(2\lambda_d)$ that would be expected by direct application of the theory of incompressible turbulence. This implies that the density variations average out almost independently of velocity variations.

The MHD models that fit (3) most closely are the strong field cases, with $v_A/c_s = 10$. The weak field cases appear to follow a relation similar to (3), but with values of η_v up to a factor of two higher, as shown in Fig. 5. Without further computation, it remains unclear how much of the variation seen among the models with the same β is due to random fluctuations or the remaining lack of numerical convergence, and how much is real. The higher dissipation seen in the high-β, weak-field cases could be explained by noting that weak fields will be more strongly influenced by the flow, generating more dissipative MHD waves.

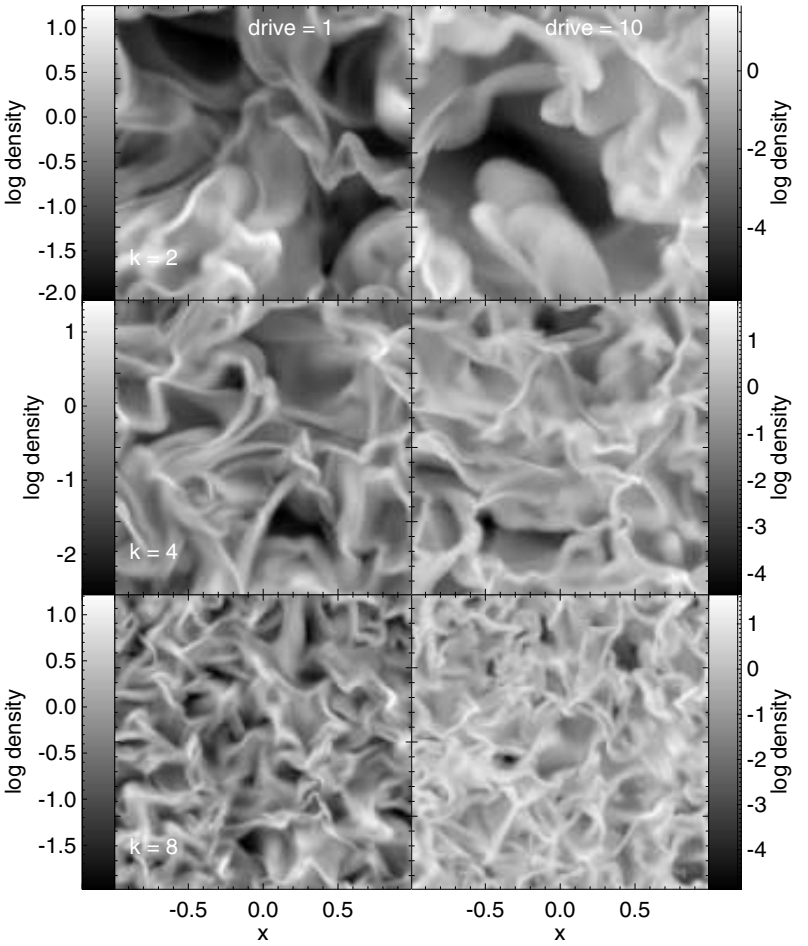


Fig. 3. Models showing the appearance of turbulence with different characteristic size scales and driving energy input. Log of density on slices through hydrodynamic models from [52] at a resolution of 128^3 grid points. The value of "drive" given in the figure is the driving energy input rate $\dot{E}_{\rm in}$ for that model, while $k = L/\lambda_d$ is the wavenumber of the driving pattern, and the size of the cube $L = 2$ for all runs

We can use these results to discuss whether decaying turbulence can delay gravitational collapse. This can formally be examined by determining whether the ratio

$$\tau = t_d/t_{\rm ff} > 1, \tag{4}$$

where the turbulent decay time $t_d = E_{\rm kin}/\dot{E}_{\rm kin}$, and the free-fall time $t_{\rm ff}$ for the gas is given by (1). Because t_d depends not only on the strength of the turbulence, but also on the driving wavelength, the value of τ also depends on the ratio

$$\kappa = \lambda_d/\lambda_J, \tag{5}$$

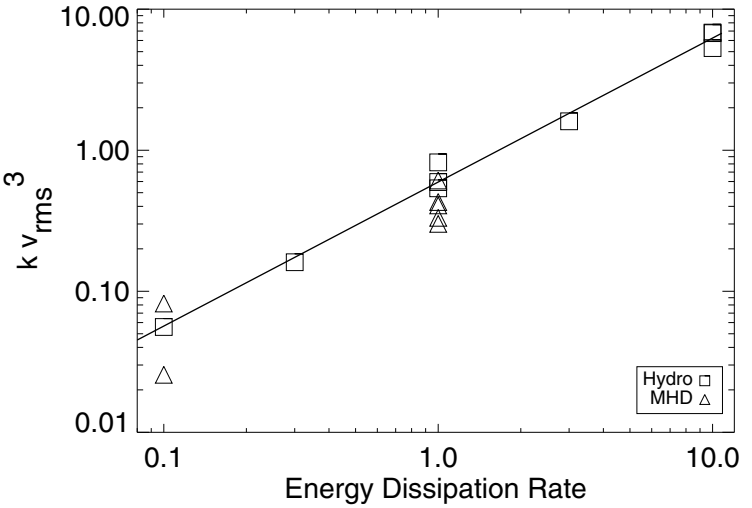


Fig. 4. Energy dissipation rate for hydro and MHD models compared to $k_d v_{\rm rms}^3$ where k_d is the dimensionless driving wavenumber, and the size of the cube $L = 2$ for all runs. The line has slope of unity, and is a fit to the hydro models, yielding the value for the dissipation coefficient $\eta_v = 0.21/\pi$ (see equation [3]). Hydrodynamical models are indicated by squares, MHD models by triangles (From [52])

where the Jeans wavelength $\lambda_J = c_s\sqrt{\pi/G\rho_0}$. Numerical models that I discuss in the next section show that turbulence cannot support the gas against collapse at wavelengths longer than the driving wavelength [50,47], so that $\kappa \leq 1$ is necessary for the turbulence to fully support against collapse.

Substituting for the values in (4), we can write

$$\tau = \frac{E_{\rm kin}}{\dot{E}_{\rm kin}} \frac{c_s}{\lambda_J} \sqrt{\frac{32}{3}}. \tag{6}$$

We can now use (3) for $\dot{E}_{\rm kin}$, and, somewhat less accurately, take $E_{\rm kin} \sim m v_{\rm rms}^2/2$, noting that this introduces no more than a 20–30% error. Substituting and using the definition of κ given in (5), I find that the dissipation time scaled in units of the free fall time is

$$\tau(\kappa) = \frac{1}{4\pi\eta_v} \left(\frac{32}{3}\right)^{1/2} \frac{\kappa}{M_{\rm rms}} \simeq 3.9 \frac{\kappa}{M_{\rm rms}}, \tag{7}$$

where $M_{\rm rms} = v_{\rm rms}/c_s$ is the rms Mach number of the turbulence. In molecular clouds, $M_{\rm rms}$ is typically observed to be of order 10 or higher, while $\kappa < 1$ is required for support, so turbulence will decay long before the cloud collapses and not markedly influence its collapse.

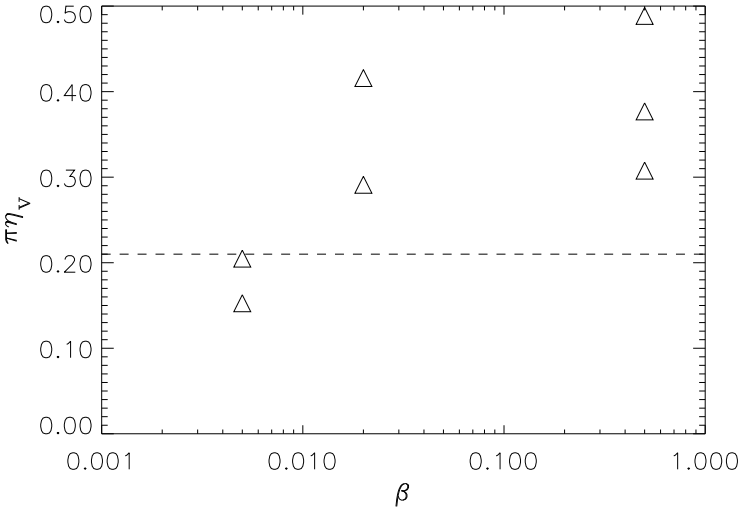


Fig. 5. Dependence of the dissipation coefficient $\eta_v = \dot{E}_{\rm kin}/(\tilde{k}mv_{\rm rms}^3)$ on the plasma β, the ratio of thermal to magnetic pressure. The dashed line shows the value derived from unmagnetized models. Models with weaker fields appear to have as much as a factor two higher dissipation rate, while strong field models approach the unmagnetized rate (From [52])

4 Self-gravitating Turbulence

Now that it is clear that driven turbulence must be present in molecular clouds, we can ask whether it alone is sufficient to support clouds against gravitational collapse.

The virial theorem provides a first way of examining this question. In equilibrium the total kinetic energy in the system adds up to half its potential energy, $E_{\rm kin} + 1/2\,E_{\rm pot} = 0$. If $E_{\rm kin} + 1/2\,E_{\rm pot} < 0$ the system contracts, while $E_{\rm kin} + 1/2\,E_{\rm pot} > 0$ implies expansion. In turbulent clouds, the total kinetic energy includes not only the internal energy but also the contributions from turbulent gas motions. If this is taken into account, simple energy considerations can already provide a qualitative description of the collapse behavior of turbulent self-gravitating media [24].

A more thorough investigation, however, requires a linear stability analysis. For the case of an isothermal, infinite, homogeneous, self-gravitating medium at rest (i.e. without turbulent motions) Jeans [43] derived a relation between the oscillation frequency ω and the wave number k of small perturbations,

$$\omega^2 - c_{\rm s}^2 k^2 + 4\pi G\,\rho_0 = 0\,, \tag{8}$$

where $c_{\rm s}$ is the isothermal sound speed, G the gravitational constant, and ρ_0 the initial mass density. Note that the derivation includes the ad hoc assumption that the linearized version of the Poisson equation describes only the relation between the perturbed potential and the perturbed density, neglecting the potential of

the homogeneous solution. This is the so-called 'Jeans swindle'. Perturbations are unstable against gravitational contraction if their wave number is below a critical value, the Jeans wave number $k_{\rm J}$, i.e. if

$$k^2 < k_{\rm J}^2 \equiv \frac{4\pi G\rho_0}{c_{\rm s}^2} \ , \tag{9}$$

or equivalently if the wave length of the perturbation exceeds a critical size given by $\lambda_{\rm J} \equiv 2\pi k_{\rm J}^{-1}$. This directly translates into a mass limit. All perturbations with masses exceeding the Jeans mass,

$$M_{\rm J} \equiv \rho_0 \lambda^3 = \left(\frac{\pi}{G}\right)^{3/2} \rho_0^{-1/2} c_{\rm s}^3, \tag{10}$$

will collapse under their own weight.

These and subsequent analytical approaches (reviewed in [47]) make a strong assumption that substantially limits their reliability, namely that the equilibrium state is homogeneous, with constant density ρ_0. However, observations clearly show that molecular clouds are extremely non-uniform. One way to achieve progress and circumvent the restrictions of a purely analytical approach is to perform numerical simulations. Bonazzola et al. [24], for example, used low resolution (32×32 collocation points) calculations with a 2-dimensional spectral code to support their analytical results. Also restricted to two dimensions were the hydrodynamical studies by Passot et al. [69], Léorat, Passot & Pouquet [50], Vázquez-Semadeni et al. [89] and Ballesteros-Paredes, Vázquez-Semadeni & Scalo [8], although performed with far higher resolution. Magnetic fields were introduced in two dimensions by Passot, Vázquez-Semadeni, & Pouquet [71], and extended to three dimensions with self-gravity (though at only 64^3 resolution) by Vázquez-Semadeni, Passot, & Pouquet [90]. A careful analysis of 1-dimensional computations including both MHD and self-gravity was presented by Gammie & Ostriker [33], who extended their work to 2.5 dimensions more recently [66].

We use ZEUS-3D and SPH to examine the gravitational stability of three-dimensional hydrodynamical turbulence at higher resolution than before, and include magnetic fields using ZEUS-3D. The use of both Lagrangian and Eulerian numerical methods to solve the equations of self-gravitating hydrodynamics in three dimensions (3D) allows us to attempt to bracket reality by taking advantage of the strengths of each approach. This also gives us some protection against interpreting numerical artifacts as physical effects.

SPH can resolve very high density contrasts because it increases the particle concentration, and thus the effective spatial resolution, in regions of high density, making it well suited for computing collapse problems as shown in Fig. 6. By the same token, though, it resolves low-density regions poorly. Shock structures tend to be broadened by the averaging kernel in the absence of adaptive techniques. The correct numerical treatment of gravitational collapse requires the resolution of the local Jeans mass at every stage of the collapse [18]. In the current code, once an object with density beyond the resolution limit of the code has formed

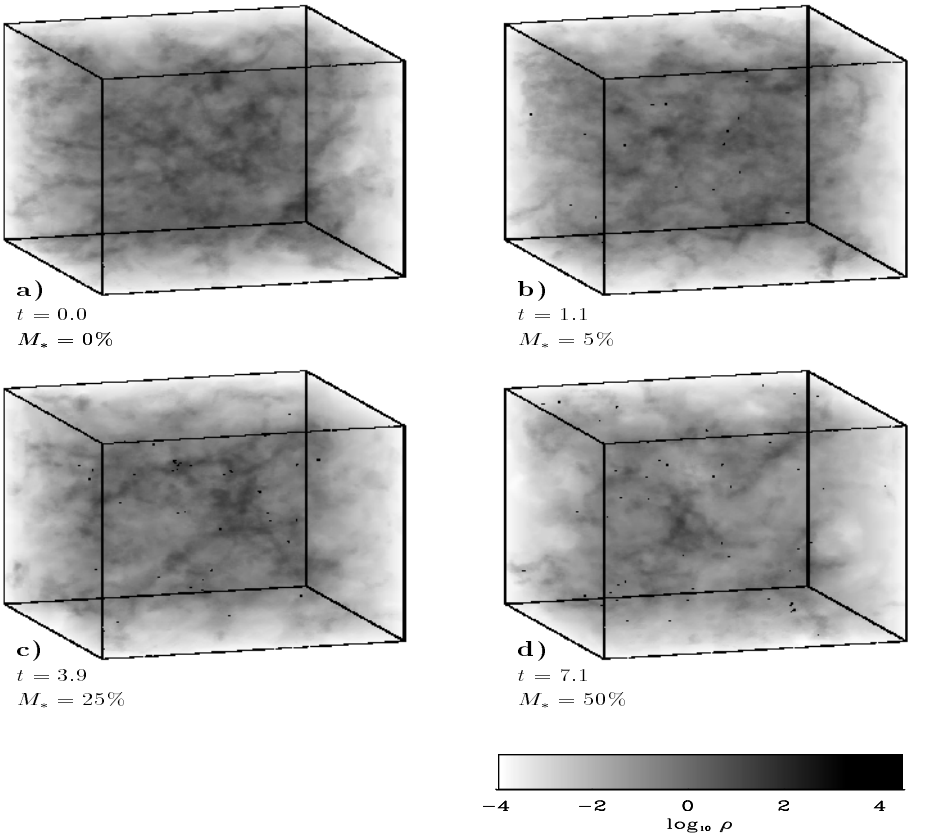


Fig. 6. SPH density cubes for model driven in the wavenumber interval $3 \leq k_d \leq 4$, shown (**a**) at the time when gravity is turned on, (**b**) when the first dense cores are formed and have accreted $M_* = 5\%$ of the mass, (**c**) when the mass in dense cores is $M_* = 25\%$, and (**d**) when $M_* = 50\%$. Time is measured in units of the global system free-fall time scale $\tau_{\rm ff}$ (From [47])

in the center of a collapsing gas clump it is replaced by a 'sink' particle [17]. Adequately replacing high-density cores and keeping track of their further evolution in a consistent way prevents the time step from becoming prohibitively small. We are thus able to follow the collapse of a large number of cores until the overall gas reservoir becomes exhausted.

ZEUS-3D, conversely, gives us equal resolution in all regions, and allows us to resolve shocks well everywhere, as well as allowing the inclusion of magnetic fields, as shown in Fig. 7. On the other hand, collapsing regions cannot be followed to scales less than one or two cells. We must again consider the resolution required for gravitational collapse. For a grid-based simulation, the criterion given by Truelove et al. [87] holds. Equivalent to the SPH resolution criterion, the mass contained in one grid zone has to be rather smaller than the local Jeans mass throughout the computation.

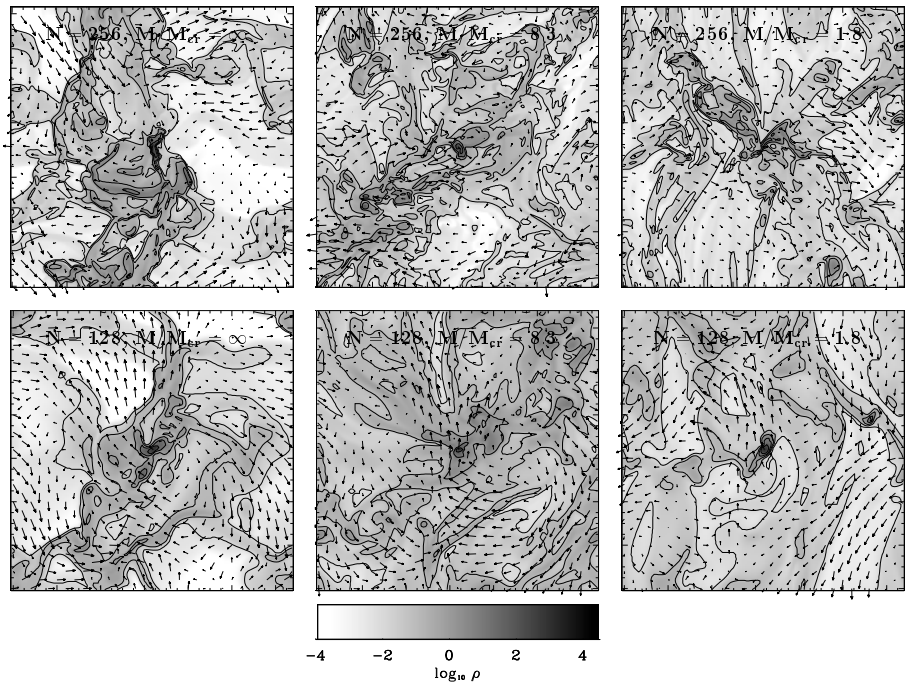


Fig. 7. Two-dimensional slices of 256^3 models from [39] driven with wavenumber $k = 2$ hard enough that the mass in the box represents only $1/15\ \langle M_{\rm J}\rangle_{\rm turb}$, and with magnetic fields strong enough to give critical mass fractions as shown, corresponding to $\beta = \infty$, 0.9, and 0.04. Slices are taken at the location of the zone with the highest density at the time when 10% of the total mass has been accreted onto cores. The plot is centered on this zone. Arrows denote velocities in the plane. The length of the largest arrows corresponds to a velocity of $v \sim 20c_s$. The density greyscale is given in the colorbar. (From [39])

Applying this criterion strictly would limit our simulations to the very first stages of collapse, as we have not implemented anything like sink particles in ZEUS. We have therefore extended our models beyond the point of full resolution of the collapse, as we are primarily interested in the formation of collapsed regions, but not their subsequent evolution. Thus, in the ZEUS models, the fixed spatial resolution of the grid implies that strongly collapsed cores have a larger cross-section than appropriate for their mass. In encounters with shock fronts the probability for these cores to get destroyed or lose material is overestimated. Cores simulated with ZEUS are therefore more easily disrupted than they would be physically. SPH, on the other hand, underestimates the disruption probability, because sink particles cannot lose mass or dissolve again once they have formed. The physical result is thus *bracketed* by these two numerical methods.

We use the same driving method as described above for both the SPH and ZEUS models. We define an effective turbulent Jeans mass $\langle M_{\rm J}\rangle_{\rm turb}$ by substituting $c_{\rm s}^2 \longrightarrow c_{\rm s}^2 + 1/3\,\langle v^2\rangle$ for the thermal sound speed $c_{\rm s}$ in (10) where we

approximate the rms velocity of the flow $\langle v^2 \rangle$ by $2E_{\rm kin}/M$. The turbulent Jeans mass $\langle M_{\rm J} \rangle_{\rm turb}$ must be compared to the total system mass $M \equiv 1$ in order to determine whether global stability is reached.

We find that *local* collapse occurs even when the turbulent velocity field carries enough energy to counterbalance gravitational contraction on global scales, as shown in Fig. 8. This confirms the results of two-dimensional (2D) and low-resolution (64^3) 3D computations with and without magnetic fields by Vázquez-Semadeni et al. [90]. An example of local collapse in a globally supported cloud is given in Fig. 6. The presence of shocks in supersonic turbulence drastically alters the result from analytic models of incompressible turbulence, as was first noted by Elmegreen [29] and studied numerically by Vázquez-Semadeni et al. [90]. The density contrast in isothermal shocks scales quadratically with the Mach number, so the shocks driven by supersonic turbulence create density enhancements with $\delta\rho \propto \mathcal{M}^2$, where $\mathcal{M}$ is the *rms* Mach number of the flow. In such fluctuations the local Jeans mass is *decreased* by a factor of $\mathcal{M}$ and therefore the likelihood for gravitational collapse *increased*.

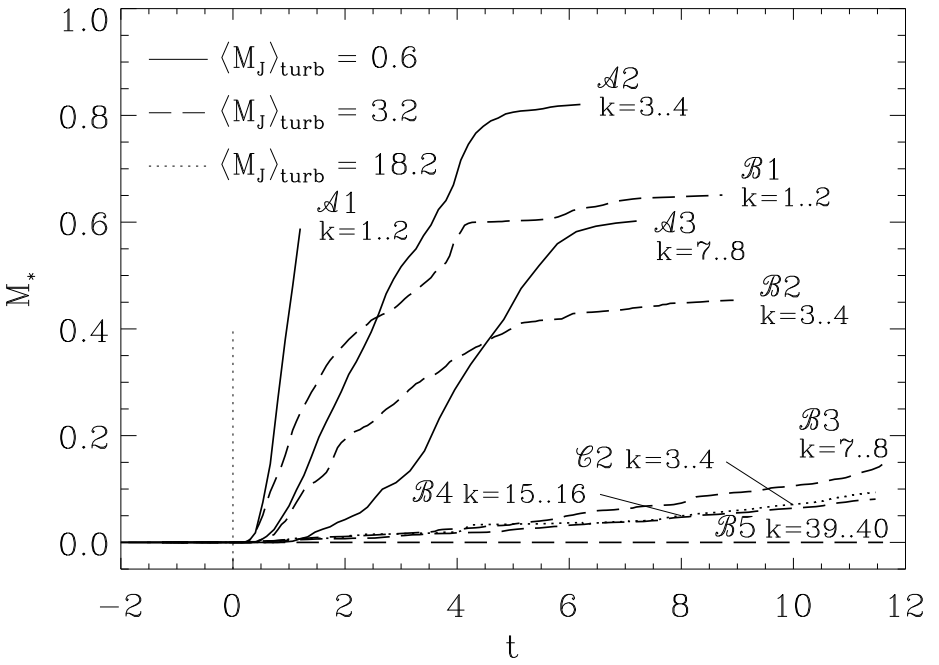


Fig. 8. Fraction of mass M_* in dense cores as function of time. All models are computed using SPH with sink particles replacing dense, collapsed cores. The different model parameters from [47] are indicated in the figure. The mass in the box is initially unity, so the solid curves are unsupported formally, while the others are supported. The figure shows how the efficiency of local collapse depends on the scale and strength of turbulent driving. Time is measured in units of the global system free-fall time scale $\tau_{\rm ff}$, while the driving wavelength $\lambda_d = 2\pi/k$ (From [47])

To test this explanation numerically, we designed a test case driven at short enough wave length and high enough power to support even fluctuations with $\delta\rho \propto \mathcal{M}^2$, and ran it with both codes, driven with a wave number $k = 39 - 40$. Within 20 $\tau_{\rm ff}$ this model shows no signs of collapse. All the other globally supported models with less extreme parameters that we computed did form dense cores during the course of their evolution, supporting our hypothesis that local collapse is caused by the density fluctuations resulting from supersonic turbulence.

Magnetic fields might alter the dynamical state of a molecular cloud sufficiently to prevent gravitationally unstable regions from collapsing [59]. They have been hypothesized to support molecular clouds either magnetostatically or dynamically through MHD waves. Mouschovias & Spitzer [63] derived an expression for the critical mass-to-flux ratio in the center of a cloud for magnetostatic support. Assuming ideal MHD, a self-gravitating cloud of mass M permeated by a uniform flux Φ is stable if the mass-to-flux ratio

$$\frac{M}{\Phi} < \left(\frac{M}{\Phi}\right)_{cr} \equiv \frac{c_\Phi}{\sqrt{G}}. \tag{11}$$

with c_Φ depending on the geometry and the field and density distribution of the cloud. A cloud is termed *subcritical* if it is magnetostatically stable and *supercritical* if it is not. Mouschovias & Spitzer [63] determined that $c_\Phi = 0.13$ for a spherical cloud.

We include magnetic fields in our models of driven, self-gravitating turbulence to test their effectiveness in supporting against self-gravity. The MHD simulations start with a uniform magnetic field in the z-direction. We must consider the resolution required to accurately follow magnetized collapse. Numerical diffusion can reduce the support provided by a static or dynamic magnetic field against gravitational collapse. Increasing the numerical resolution decreases the scale at which numerical diffusion acts. For strong, subcritical fields, the resolution should ensure that numerical diffusion remains unimportant even for the dense, shocked regions. We did a suite of models varying the sound speed and the mass in the cube while holding the magnetic field strength constant, thus varying M/M_{cr} and the number of zones in a Jeans length λ_J. From these models we conclude that for a self-gravitating magnetostatic sheet to be well resolved with the algorithm under study, its Jeans length must exceed four zones [39].

Collapse occurs in both unmagnetized and magnetized cases at all resolutions, as shown in Fig. 9. However, increasing the resolution makes itself felt in different ways in hydrodynamical and MHD models. In the hydrodynamical case, higher resolution results in thinner shocks and thus higher peak densities. These higher density peaks form cores with deeper potential wells that accrete more mass and are more stable against disruption. If we increase the resolution in the MHD models, on the other hand, we can better follow short wavelength MHD waves, which appear to be able to delay collapse, although not to prevent it. This result extends to models with 512^3 zones [40].

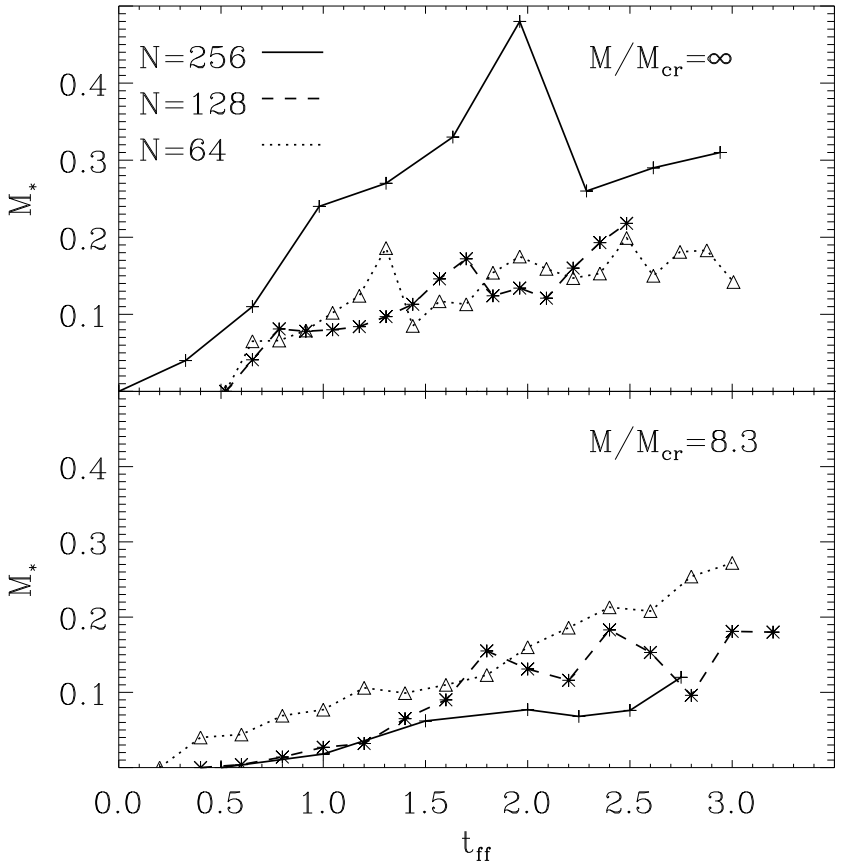


Fig. 9. Comparison of the mass accretion behaviour for runs driven at wavenumber $k = 1 - 2$ with varying resolution. Pure hydro runs are shown in the upper panel, and MHD runs in the lower panel, with resolutions 643 (*dotted*), 1283 (*dashed*), and 2563 (*solid*). M_* denotes the sum of masses found in all cores determined by the modified `clumpfind` algorithm [92,47]. Note that these cores are subject to destruction as they can only collapse to the grid scale. Times are given in units of free-fall time. Although the collapse rate varies, we get collapse in all cases (From [39])

5 Characterization of Turbulence

5.1 Wavelet Transforms

To obtain clues to the true physical nature of interstellar turbulence, characteristic scales and any inherent scaling laws have to be measured and modelled. A major problem with characterizing both the observations and the models is to determine what scaling behaviour, if any, is present in complex turbulent structures. Both the velocity and density fields need to be considered, but only the radial velocity and column densities can be observed.

One measure useful for characterizing structure and scaling in observed maps of molecular clouds is the Δ-variance, $\sigma 2_\Delta$, an averaged wavelet transform method for measuring the amount of structure at different scales introduced by Stutzki et al. [86]. The Δ-variance spectrum clearly shows characteristic scales and scaling relations, and its logarithmic slope can be analytically related to the spectral index of the corresponding Fourier power spectrum (see Fig. 10).

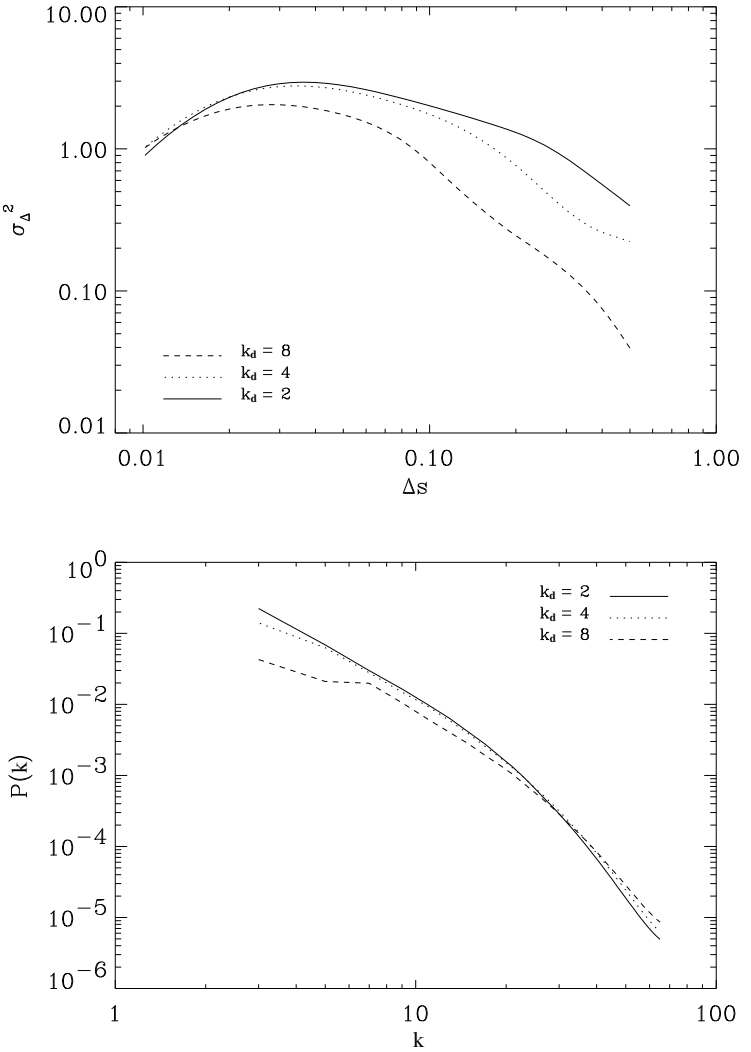


Fig. 10. Comparison of 3D Δ-variance spectra (*above*) to Fourier power spectra (*below*) for 128^3 models of turbulence driven at wavenumbers of $k_d = 2$ (solid), $k_d = 4$ (dotted), and $k_d = 8$ (dashed), demonstrating that the analytic relation between power spectrum slope and Δ-variance spectrum slope also holds for the local behavior of structures not showing a straight power-law (From [56])

In order to understand the physical significance of the characterization of the observational maps by Δ-variance spectra, we apply the same analysis to observations of the Polaris Flare at three different scales [30,19,38], as analyzed by [86,20], and to simulated observations of gas distributions resulting from our MHD models [56,65], as well as to the actual density distributions from the models. In Fig. 10 we compare the power spectrum and Δ-variance for the density distribution from three simulations with different driving scales. Each shows a different characteristic scale formed by the driving, visible as a turn-over at large lags in the Δ-variances and at small wavenumbers in the power spectrum, respectively. At smaller lags and higher wavenumbers power laws can be seen in both cases for the models with longer wavelength driving, with their slopes related by the analytic relation mentioned above. A steep drop-off follows at the smallest scales indicating the resolution limit of the simulation. The model with the shortest wavelength ($k_d = 8$) driving shows no power law because the driving scale of 16 zones is only marginally larger than the dissipation scale of about 10 zones.

The Δ-variance analysis of astronomical maps was extensively discussed by Bensch et al. [20]. In all the observations they analyzed, the total cloud size was the only characteristic scale detected by means of the Δ-variance. Below that size they found a self-similar scaling behaviour reflected by a power law with index α =0.5–1.3 corresponding to a Fourier power spectral index ζ =2.5–3.3.

In Fig. 11 we show how numerical resolution, or equivalently the dissipation scale, influences the Δ-variance spectrum that we find from our simulations. We test the influence of the numerical resolution on the structure by comparing a simple hydrodynamic problem of decaying turbulence computed at resolutions from 64^3 to 256^3, with an initial rms Mach number $M = 5$. In contrast to the results from [52] which showed little dependence of the energy dissipation rate on the numerical resolution, we find here clear differences in the scaling behaviour of the turbulent structures. At small scales we find a very similar decay in the relative structure variations up to scales of about 10 times the pixel size (0.03, 0.06, and 0.1 for the resolutions 256^3, 128^3, and 64^3, respectively) in all three models. This constant length range starting from the pixel scale clearly identifies this decay as being due to the numerical viscosity acting at the smallest available size scales.

At the largest lags structure variations decay for all three simulations at a length scale of roughly a quarter the cube size. This structure reflects the original driving of the turbulence with wavenumbers $1 < k_d < 8$, which produces power in the Δ-variance at about half of the corresponding length scales. Structures larger than about half the cube size are suppressed by the periodicity of the simulation cubes.

We can now attempt to interpret the observations using the behavior seen in the model Δ-variance spectra, and equivalent behavior seen in velocity centroid Δ-variance spectra of the models, as described by Ossenkopf & Mac Low [65]. Figure 12 shows the square root of the Δ-variance for the Polaris Flare velocity centroid maps. An upturn in power at the smallest lags in each map produced by

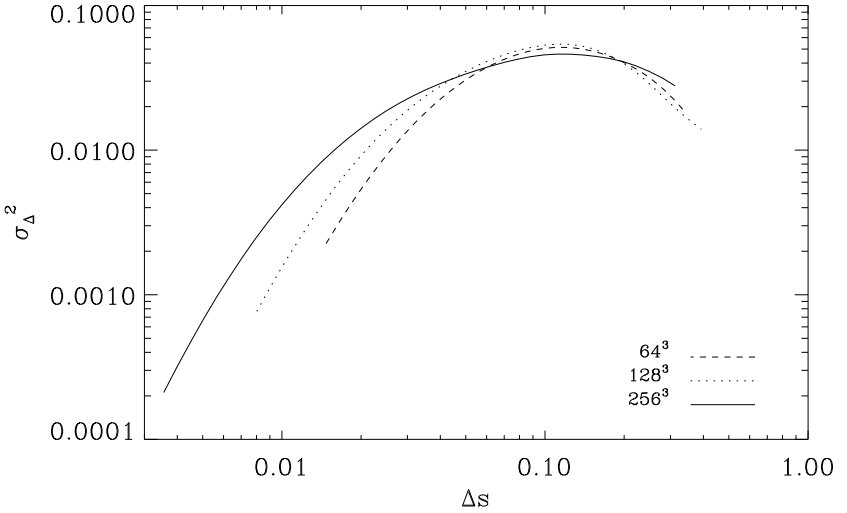


Fig. 11. Δ-variance of the density cubes (projected to two-dimensions) for a model of decaying hydrodynamic turbulence at a time $t/t_s = 0.1$, where t_s is the sound-crossing time, computed using numerical resolutions of 64^3, 128^3, and 256^3 zones (From [56])

the observational noise adding power at small scales was subtracted to prepare these spectra [20]. The velocity structure in the Polaris Flare maps shows power-law behaviour from scales of about 0.05 to several parsecs. Below that range the slope of the Δ-variance spectrum of the velocity centroids steepens, perhaps because of physical dissipation through ambipolar diffusion or other mechanisms, while above that range it flattens because of the lack of emission on larger scales. Bensch et al. [20] demonstrated similar scaling behaviour in the intensity maps, reflecting the density structure.

Comparison between the observations of the Polaris Flare and the turbulence simulations constrains the mechanisms driving the turbulence in this cloud. Any mechanism that drives at an intermediate length scale, such as jets from embedded protostars, should produce a peak in the Δ-variance at that scale, which is not observed. The approximately self-similar, power-law behaviour seen in the observations is best reproduced by models where the energy is injected at large scales and dissipated at small scales. Driving by interactions with superbubbles and field supernova remnants [64,54] would provide such a driving mechanism. The Polaris Flare molecular cloud lies in the wall of a large cylindrical structure representing one of the nearest H I supershells, the North Celestial Polar Loop [60], adding additional support to this proposal.

The dominant physical mechanism for dissipation in molecular clouds was first shown by Zweibel & Josafatsson [94] to be ambipolar diffusion. Klessen et al. [47] showed that the length scale on which ambipolar diffusion will become important can be found by examining the ambipolar diffusion Reynolds number

$$R_A = \mathcal{M}_A \tilde{L} \nu_{ni} / v_A \tag{12}$$

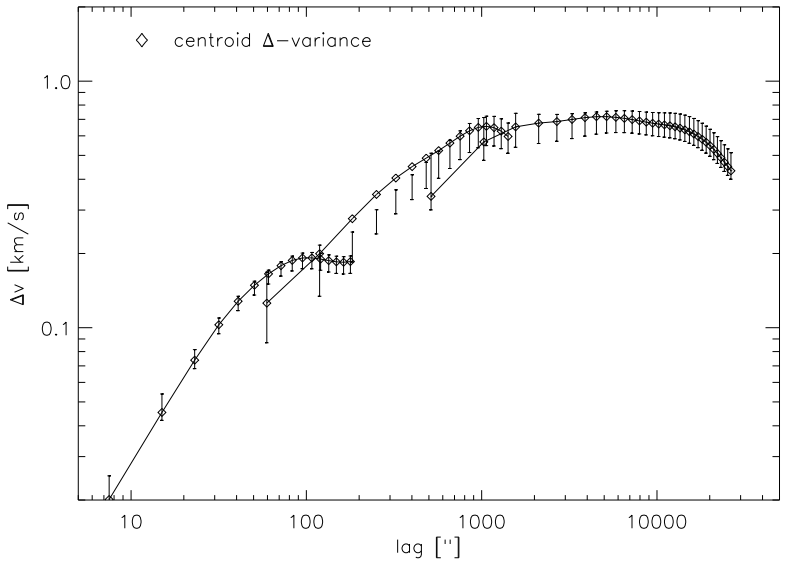


Fig. 12. The square root of the Δ-variance of the smoothed Polaris Flare velocity centroid maps. The size of the smoothing filter was adapted to remove only the average noise level. Spectra are shown from three nested maps obtained with the 1.2 m CfA telescope with half power beam width (HPBW) 8.7' [38], the 3 m KOSMA with HPBW 2.2' [19], and the 30 m IRAM with HPBW 0.35' [30]. At a distance of 150 pc, $1000'' = 0.73$ pc. The three individual maps are assumed to sample the same underlying spectrum at overlapping scales. (From [65])

defined by Balsara [10] and Zweibel & Brandenburg [93], where $\tilde{L}$ and $\mathcal{M}_A$ are the characteristic length and Alfvén Mach number, $\nu_{ni} = \gamma\rho_i$ is the rate at which each neutral is hit by ions, and $v_A^2 = B^2/4\pi\rho$ approximates the effective Alfvén speed in a mostly neutral region with total mass density $\rho = \rho_i + \rho_n$ and magnetic field strength B. The coupling constant depends on the cross-section for ion-neutral interaction, and for typical molecular cloud conditions has a value of $\gamma \approx 9.2 \times 10^{13}$ cm^3 s^{-1} g^{-1} (e.g. [81]).

Setting the ambipolar diffusion Reynolds number $R_A = 1$ yields a diffusion length scale of

$$L_D = v_A/\mathcal{M}_A\nu_{ni} \tag{13}$$

$$\approx (0.041\text{pc})\mathcal{M}_A \left(\frac{B}{10\ \mu\text{G}}\right)\left(\frac{10^{-6}}{x}\right)\left(\frac{10^3\ \text{cm}^{-3}}{n_n}\right)^{3/2} \tag{14}$$

with the ionization fraction $x = \rho_i/\rho_n$ and the neutral number density $n_n = \rho_n/\mu$, with $\mu = 2.36m_H$. If the ionization level in the Polaris Flare is low enough and the field is high enough, this length scale of order 0.05 pc would be directly resolved in the IRAM observations. We cannot yet unambiguously say whether the steepening observed at the smallest scales in the velocity spectra is due to beam smearing at the observational resolution limit or to a detection of the

actual dissipation scale, similar to the downturn at the dissipation scale in the numerical models. If better observations do continue to show such a downturn in the future, that will be an indication of the dissipation scale.

5.2 Clump Characterization

Using 2D numerical simulations, Ballesteros-Paredes et al. [8] showed that observed clumps frequently come from the superposition of several physically disconnected regions in the line of sight at the same radial velocity (see also [67]) but not necessarily at the same position or three-dimensional velocity. This had earlier been suggested by workers including Burton [25], Issa et al. [42], Falgarone, Puget, & Pérault [31], and Adler & Roberts [1].

We used the models described above to examine this effect in 3D using simulated observations. We calculate the radiative transfer through the density and velocity fields given by the computations, assuming local thermodynamic equilibrium (LTE) for the population of the molecular energy levels, which is a sufficiently good approximation to study qualitatively the effects of projection and superposition (velocity crowding) of structure in the cloud. In order to determine differences when observing the same region with different tracers, we also set minimum density thresholds below which the molecules might be underexcited.

Velocity crowding contributes to the generation of clumps in the observational space, so observed clumps frequently contain emission from physically separated regions [8,72,67,49]. We demonstrate this effect using a typical MHD simulation with intermediate values of driving strength ($L = 1$) and wavenumber ($k = 4$).

Figure 13 shows physical and observed maps for ^{13}CO(1-0) and (2-1). We see that clumps in real space (letters A, B, and C in panel $y - x(1 - 0)$) do not necessarily have a counterpart in observational space. Clumps in observational space, on the other hand, (letter D in panel $Vy{-}x(1{-}0)$) do not necessarily come from isolated regions in real space, but have contributions from many different regions along the same line of sight. In Fig. 13 we plot dotted lines that show the places where the emission of a physical clump lies in observational space, and where the emission from an observational clump is generated in physical space. For reference, we use the same lines in the ^{13}CO(1-0) map as in the ^{13}CO(2-1) map.

If clumps in the observational space are the result of the contribution of multiple regions in the physical space it is of primary importance to understand whether relationships reported for observed clumps in molecular clouds are also valid for the actual physical clumps.

Larson [48] studied the dependence with size of the mean density, velocity dispersion, and mass spectrum of the clouds in a sample of observational data

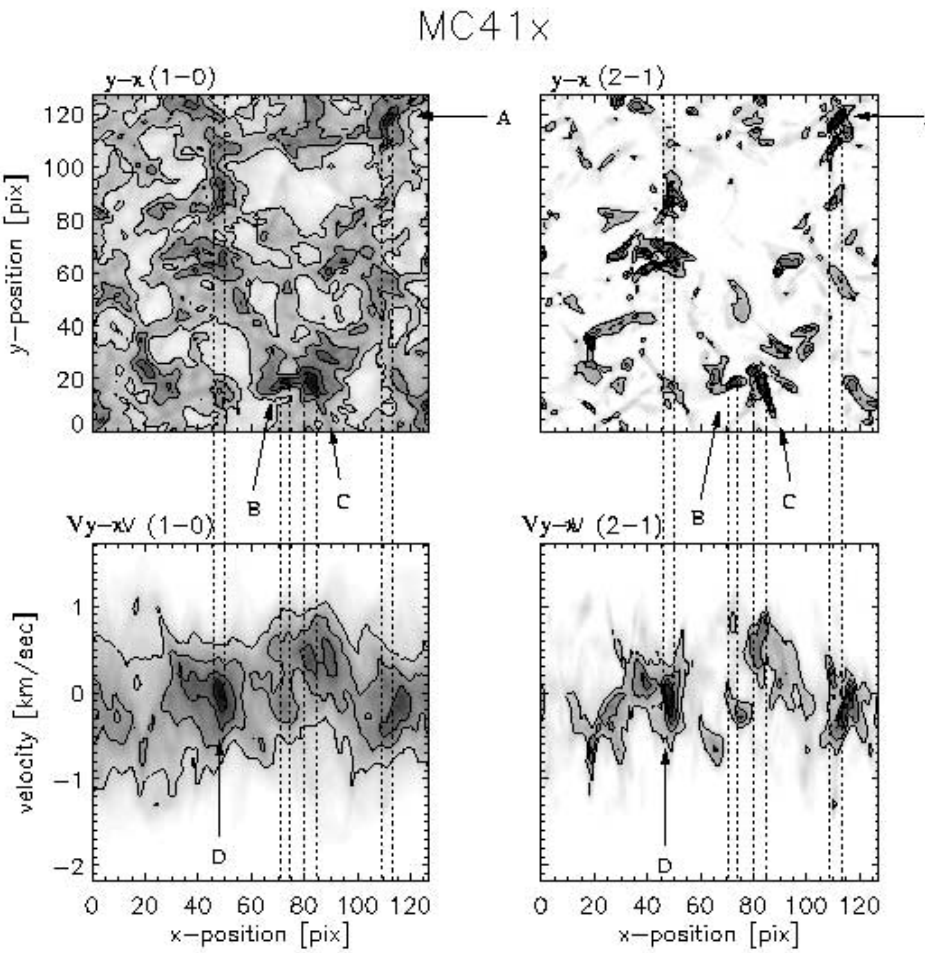


Fig. 13. Physical and velocity space maps for ^{13}CO(1-0) and (2-1) in a run with $\dot{E} = 1$, $k = 4$ and $B = 0.1$. Clumps in physical space (A, B, C) do not necessarily correspond to clumps in the observed velocity space. Observed clumps in velocity space (D) are not necessarily formed by emission from a single region in physical space. From [7]

taken from the literature. He found

$$\rho \propto R^{\alpha}, \tag{15}$$

$$\delta v \propto R^{m}, \tag{16}$$

$$\frac{dN}{d\log M} \propto M^{\gamma}. \tag{17}$$

The most commonly quoted values in the literature are $\alpha \sim -1$, $m \sim 1/2$, and $\gamma \sim -0.5$. However, there is some discrepancy in the values reported [26,51].

Larson [48] himself mentioned that the relationships he found might be due to the limitations of the observations. Kegel [44] first demonstrated that the ob-

served mean density-size relationship could be due to observational effects, and that the observed and physical values of properties such as radius or volumetric density might be quite different. Scalo [77] showed that CO and extinction saturate at roughly the same column density, which forms the upper envelope of the mean density-size relationship. Vázquez-Semadeni et al. [88] reported the lack of a mean density-size relationship, confirming numerically the analysis by Kegel ([44], see also the discussions in [48] and [77]) in the sense that there are clouds with small sizes and low column density that will be undetected in observational surveys.

In Fig. 14 we use the computational model described above to compare the density-size relationships in physical and observational spaces. We define clumps as a connected set of points below a local maximum following the intensity *only* downwards until the threshold is reached, a scheme implemented in the code called `clumpfind` [92]. We note three points about Fig. 14. First, there is no relation between mean density and size (15), confirming the 2D results of Vázquez-Semadeni et al. [88]. Second, there is a minimum density below which there are no clumps identified. This minimum is just given by the density threshold we used in `clumpfind`. Third, even though the simulations exhibit a large dynamical range in density ($\rho_{\max}/\rho_{\min} \sim 3.5 \times 10^4$), the dynamical range in the mean density-size relationship is small, because in constructing such a plot we choose the clumps around the local density maxima.

In the middle and lower panels of Fig. 14 we show the mean density-size relationship for clumps in simulated observational maps of the same run, integrated along the x axis. The middle panel is the one obtained by using CS(1-0), and the lower panel is the relationship obtained by using ^{13}CO(1-0). The observed clumps do exhibit approximately the relationship given by Larson [48], despite the lack of correlation exhibited by the physical clumps in this model. We conclude from this demonstration that the observed density-size relationship (15) is an observational artifact.

Kegel [44] suggested one mechanism that appears able to explain our results: only clouds with emission exceeding some intensity threshold will be detected, effectively setting a column density cutoff, rather than the physical density cutoff imposed in the physical density-size relationship. (This threshold is determined by the noise level in observations, or by some average intensity in the simulated observations). A constant column density cutoff produces a cutoff with slope -1 in the mean density-size plane (middle and lower panels), just as the constant physical density cutoff in physical space produces a flat cutoff.

6 Supernova-Driven Turbulence

Both support against gravity and maintenance of observed motions appear to depend on continued driving of the turbulence, as I have described. What then is the energy source for this driving?

Motions coming from gravitational collapse have often been suggested, but fail due to the quick decay of the turbulence as described above. If the turbulence

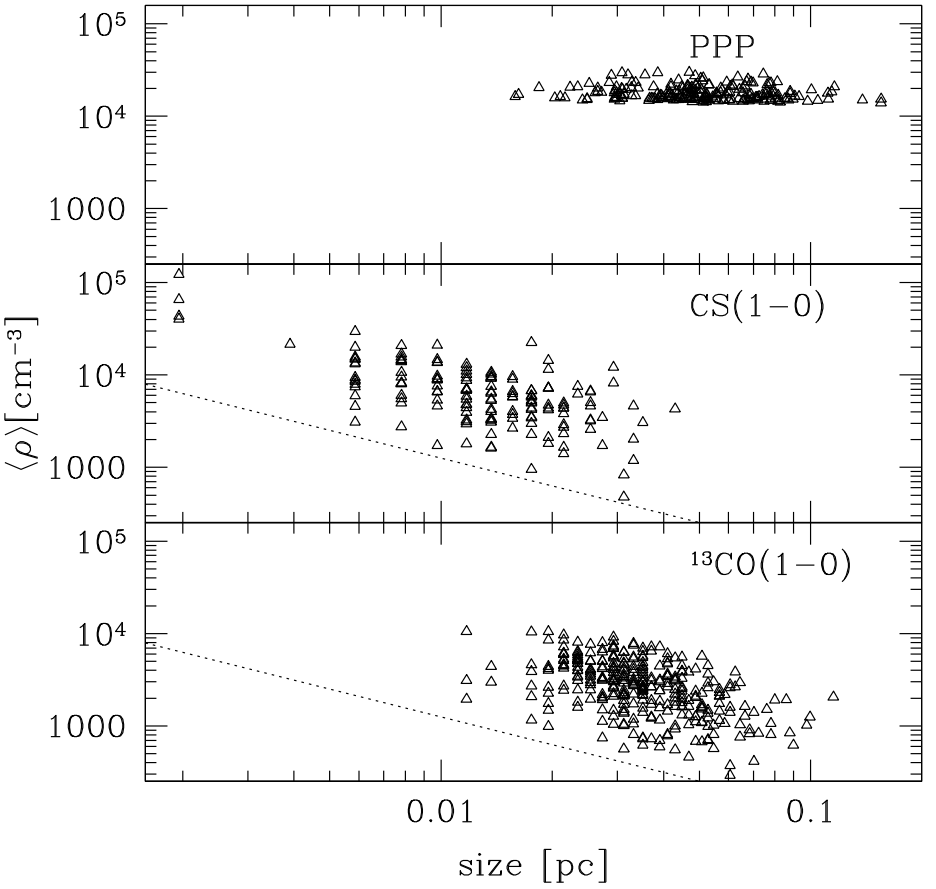


Fig. 14. Mean density-size relationship for physical clumps in physical (PPP, upper panel); and simulated observational clumps in two different density tracers in observational coordinates (PPV, middle and lower panels). The dotted line has a slope of $\alpha = -1$. In physical space we find no correlation, verifying the results of Vázquez-Semadeni et al. [88], but nevertheless the simulated observations show such a correlation, as found by Larson [48] and many others. The selection of two different density tracers was chosen to show that the apparent correlation does not depend on the selection of the density threshold. (From [7])

decays in less than a free-fall time, then it cannot delay collapse for substantially longer than a free-fall time [46].

Protostellar jets and outflows are another popular suspect for the energy source of the observed turbulence. They are indeed quite energetic, but they deposit most of their energy into low density gas, as is shown by the observation of multi-parsec long jets extending completely out of molecular clouds [9]. Furthermore, as described in the previous section, the observed motions show increasing power on scales all the way up to and perhaps beyond the largest

scale of molecular cloud complexes [65]. It is hard to see how such large scales could be driven by protostars embedded in the clouds.

Another energy source that has long been considered is shear from galactic rotation. Work by Sellwood & Balbus [78] has shown that magnetorotational instabilities [4,5] could couple the large-scale motions to small scales efficiently. For parameters appropriate to the far outer H I disk of the Milky Way, they derive a resulting velocity dispersion of 6 km s^{-1}, close to that observed. This instability may provide a base value for the velocity dispersion below which no galaxy will fall. If that is sufficient to prevent collapse, little or no star formation will occur, producing something like a low surface brightness galaxy with large amounts of H I and few stars.

In active star-forming galaxies, however, clustered and field supernova explosions appear likely to dominate the driving, raising the velocity dispersion to the 10–15 km s^{-1} observed in star-forming portions of galaxies (see work cited in [53] for example). These explosions will be predominantly from B stars no longer associated with their parent gas, as they are far more numerous than more massive O stars, have explosions that are just as powerful, and live long enough (up to 50 Myr) to either drift away from or ionize their parent clouds. This provides a large-scale self-regulation mechanism for star formation in disks with sufficient gas density to collapse despite the velocity dispersion produced by the magnetorotational instability. As star formation increases in such galaxies, the number of OB stars increases, ultimately increasing the supernova rate and thus the velocity dispersion, which will restrain further star formation.

Supernova driving not only determines the velocity dispersion, but may actually form molecular clouds by sweeping gas up in a turbulent flow. Clouds that are turbulently supported will experience inefficient, low-rate star formation, while clouds that are too massive to be supported will collapse [45], undergoing efficient star formation to form OB associations or even starburst knots.

We study supernova driving in a magnetized medium numerically, using the RIEMANN framework for computational astrophysics, which is based on higher-order Godunov schemes for MHD [76,11,12], and incorporates schemes for pressure positivity [15], and divergence-free magnetic fields [16]. (The framework also includes parallelized, divergence-conserving, MHD adaptive mesh refinement [14,13], though no results using that capability are shown here.) In the models presented here, we solve the ideal MHD equations including both radiative cooling and pervasive heating in a $(200\ \mathrm{pc})^3$ periodic computational box, using a grid of 128^3 cells. We start the simulations with a uniform density of 2.3×10^{-24} g cm^{-3}, threaded by a uniform magnetic field in the x-direction with strength 5.8 μG, a factor of roughly two stronger than that observed in the Milky Way disk. This very strong field maximizes the effects of magnetization on the turbulence.

For the cooling, we use a tabulated version of the radiative cooling curve shown in Fig. 1 of MacDonald and Bailey [57], which is based on the work of Raymond, Cox & Smith [75] and Shapiro and Moore [80]. (It falls smoothly from temperatures of order 10^5 K to 10^2 K, not incorporating a sharp cutoff at 10^4 K

due to the turnoff of Lyα cooling.) In order to prevent the gas from cooling below zero, we set the lower temperature cutoff for the cooling at 100 K. We also include a diffuse heating term to represent processes such as photoelectric heating by starlight, which we set constant in both space and time. We set the heating level such that the initial equilibrium temperature determined by heating and cooling balance is 3000 K.

We explode SNe at a rate of one every 0.1 Myr in our box, twelve times higher than our present Galactic rate, corresponding to a mild starburst like M82. The SNe are permitted to explode at random positions. Each SN explosion dumps 10^{51} erg thermal energy into a sphere with radius 5 pc. The evolution of the system is determined by the energy input from SN explosions and diffuse heating and the energy lost by radiative cooling. An example of the resulting turbulence is shown in Fig. 15.

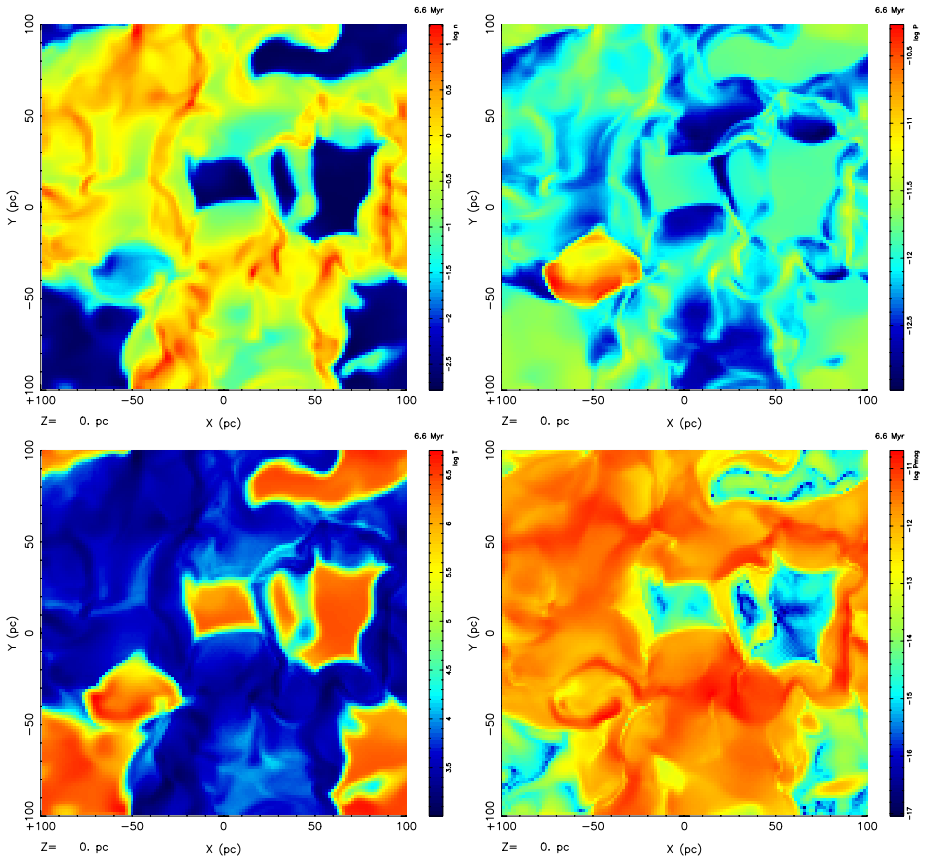


Fig. 15. Two-dimensional slices through the three-dimensional MHD model M2, parallel to the magnetic field at a time of 6.6 Myr, showing density (upper left), thermal pressure (upper right), temperature (lower left), and magnetic pressure (lower right). Color bars indicate the scale of each quantity. (From [54])

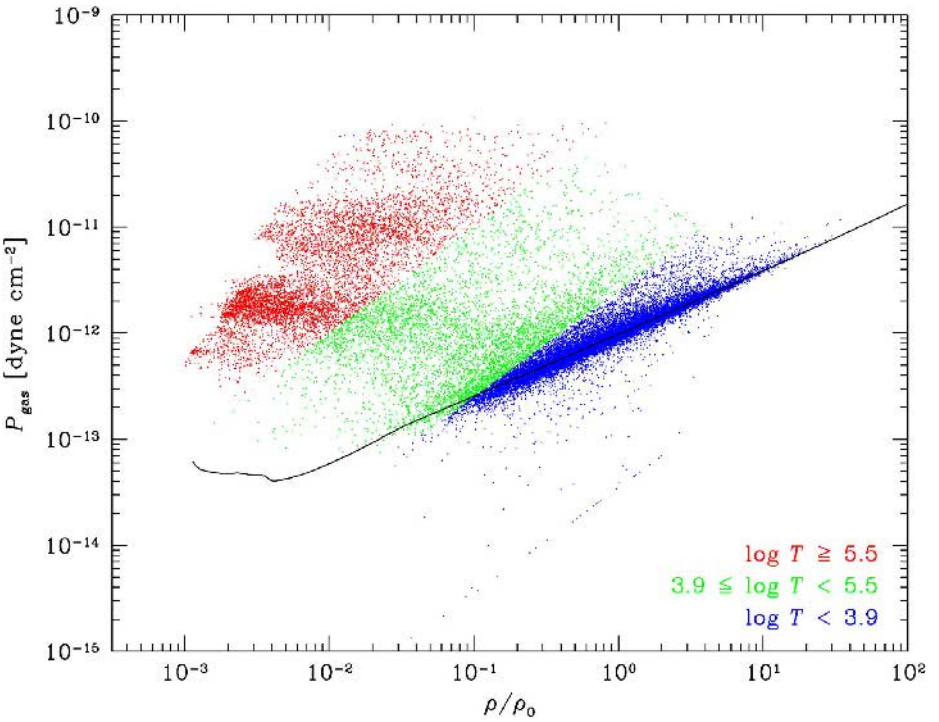


Fig. 16. Scatter plot of pressure vs. density at $t = 6.6$ Myrs in the MHD simulation M2, showing 32^3 points sampled at intervals of four points in each direction. Note that for each density a wide variation in pressure is seen. Cool gas with $\log T < 3.9$ is shown in blue, warm gas with $3.9 < \log T < 5.5$ in green, and hot gas with $\log T > 5.5$ in red. The thermal equilibrium curve for the cooling and heating functions in this simulation is overlaid as a black line. (The line of points at the very bottom right corresponds to an absolute cutoff in the cooling at 100 K that was enforced on the temperature in this model.) Note that our cooling curve may artificially prevent much low-temperature gas from forming in this model. From [54]

The first theories of the multi-phase ISM, such as Field, Goldsmith, & Habing [32], postulated an isobaric medium. Since then, multi-phase models have commonly been interpreted as being isobaric, although McKee & Ostriker [58] and Wolfire et al. [91] actually assume only local pressure equilibrium, not global, and McKee & Ostriker [58] considered the distribution of pressures. In typical multi-phase models, the heating and cooling rates of the gas have different dependences on the temperature and density, so that the balance between heating and cooling determines allowed temperatures and densities for any particular pressure. This balance can be shown graphically in a phase diagram, showing, for example, the allowed densities for any pressure ([32]; for a modern example, see Fig. 3(a) of [91]). In Fig. 16, the thermal-equilibrium curve for the heating and cooling mechanisms included is shown as a black line. Only a single phase is predicted at high densities as our cooling curve did not include the

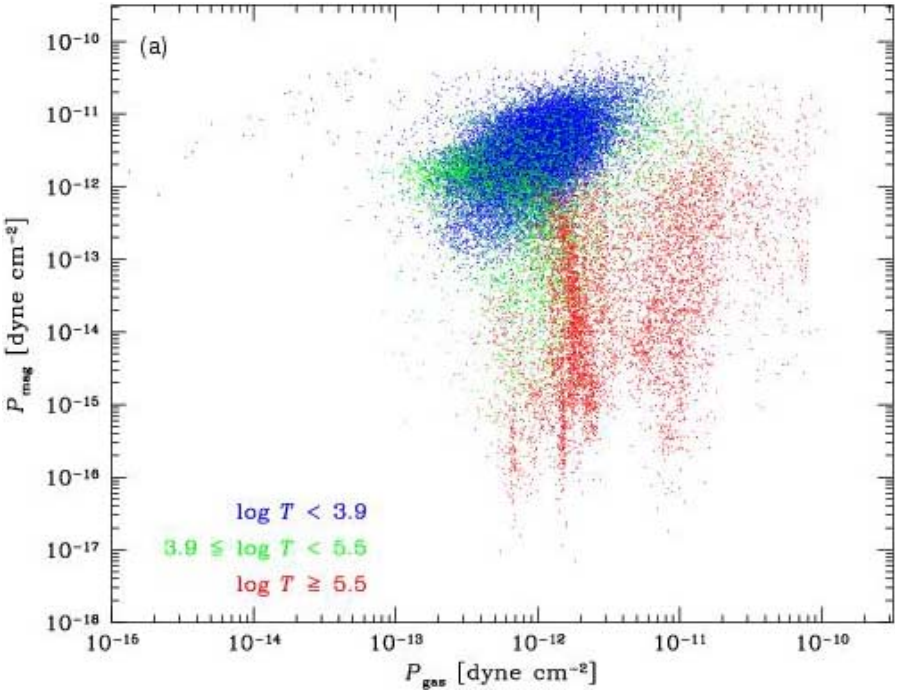


Fig. 17. Scatter plot of magnetic vs. thermal pressure at $t = 6.6$ Myrs in the MHD simulation. We again plotted a subset of 32^3 points sampled at intervals of four points in each direction. Note that regions of very low thermal pressure have substantial magnetic pressures. From [54]

physically-expected unstable region at temperatures of order 10^3 K [91]. Thus, if our model produced an isobaric medium, it would be expected to have a single low-temperature phase in uniform density given by the point at which the thermal-equilibrium curve crosses that pressure level. (Effectively, we would have the hotter two of the three phases proposed by McKee & Ostriker [58].)

The scattered points in Fig. 16 show the actual density and pressure of individual zones in the model. Many zones at low temperature do lie on the thermal equilibrium curve, but scattered all up and down it at many different pressures and densities. Furthermore, a substantial fraction of the gas has not had time to reach thermal equilibrium at all after dynamical compression. It appears that pressures are determined dynamically, and the gas then tries to adjust its density and temperature to reach thermal equilibrium at that pressure. Most gas will land on the thermal equilibrium curve when dynamical times are long compared to heating and cooling times.

Even if we include the proper physics to allow multiple phases, the behavior observed in our model will lead to all points within the range of pressures available along the thermal equilibrium line being occupied, rather than the appearance of discrete phases (also see the Chapter by Vázquez-Semadeni et al.

in this volume). Unstable regions along the thermal equilibrium curve and off it will also be populated [34], as observed by Heiles [37], but not as densely, as gas will indeed attempt to heat or cool to a stable thermal equilibrium at its current pressure. In particular, cold high pressure regions can be formed dynamically, without the influence of self-gravity, perhaps giving a method for forming molecular clouds with observed properties that are not in hydrostatic equilibrium.

The relation between magnetic and thermal pressure is shown in Fig. 17. In this Figure, the relative strength of thermal and magnetic pressure is shown at one time for the magnetized simulation. The scattering of regions at very low thermal pressure all have substantial magnetic pressures, demonstrating that magnetically supported regions can occur. However, their relative importance is rather low, as shown by the small number of points in that regime. Hot gas can be seen, on the other hand, to be dominated by thermal pressure, with low magnetic pressures.

As we have shown, the supernova-driven models have broad ranges of pressures. We can quantify this by examining the pressure probability density function (PDF), as shown in Fig. 18. In both cases, these show roughly log-normal pressure PDFs, very unlike the power-law distributions predicted by the analytic theory derived by McKee & Ostriker [58]. The observed distributions rather more resemble the log-normal density distributions suggested by Passot & Vázquez-

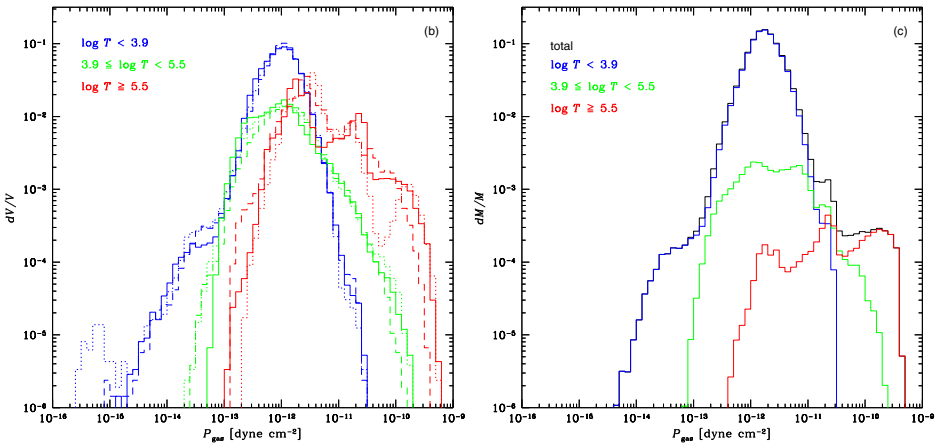


Fig. 18. (*b*) Volume-weighted PDFs of pressure from the MHD models for cool gas with $\log T < 3.9$ (blue), warm gas with $3.9 < \log T < 5.5$ (green), and hot gas with $\log T > 5.5$ (red) at different times of 5.55 Myr (dashed), 6.06 Myr (dotted), and 6.55 Myr (solid) in the 1.6 pc resolution model M2. This model has twelve times the galactic rate of SNe, which results in a very broad pressure distribution. (*c*) Mass distribution of pressure for the MHD model M2 for the full distribution (black), and for cool gas with $\log T < 3.9$ (blue), warm gas with $3.9 < \log T < 5.5$ (green), and hot gas with $\log T > 5.5$ (red) at a time of 6.55 Myrs. Most of the mass is found in cold gas, with a broad distribution around the peak pressure. From [54]

Semadeni [70]. Mac Low et al. [54] show that the cool gas can actually be modeled quite successfully with this heuristic theory.

7 Conclusions

- Even relatively strong magnetic fields, with the field in equipartition with the kinetic energy, cannot prevent the decay of turbulent motions on dynamical timescales far shorter than the observed lifetimes of molecular clouds. The significant kinetic energy observed in molecular cloud gas must be supplied more or less continuously.
- Supersonic turbulence strong enough to globally support a molecular cloud against collapse will usually cause *local* collapse. The turbulence establishes a complex network of interacting shocks. The local density enhancements in fluctuations created by converging shock flows can be large enough to become gravitationally unstable and collapse. The probability for this to happen, the efficiency of the process, and the rate of continuing accretion onto collapsed cores are strongly dependent on the driving wave length and on the rms velocity of the turbulent flow, and thus on the driving mechanism.
- Interstellar clouds driven on large scales or without even global turbulent support very rapidly form stars in clusters. On the contrary, in gas that is supported by turbulence, local collapse occurs sporadically over a large time interval, forming isolated stars. The total star formation efficiency before the cloud dissolves due to stellar feedback or external shocks will probably be low. Thus, the strength and nature of the turbulence may be fully sufficient to explain the difference between the observed isolated and clustered modes of star formation.
- Magnetorotational instabilities may provide a base value for the velocity dispersion below which no galaxy will fall. If that is sufficient to prevent collapse, little or no star formation will occur, producing something like a low surface brightness galaxy with large amounts of H I and few stars. In star-forming galaxies, however, clustered and field supernova explosions, predominantly from B stars no longer associated with their parent gas, appear likely to dominate the driving, raising the velocity dispersion to some 10–15 km s^{-1}.
- In a supernova-driven interstellar medium, we find a broad range of pressures with a log-normal distribution, and a substantial fraction of associated densities far from the thermal equilibrium values. This limits the predictive usefulness of phase diagrams based on thermal equilibrium, although thermal equilibrium at the local pressure will still be the mildly favored state. Gas pressures appear to be determined dynamically, and each individual parcel of gas seeks local thermal equilibrium at the pressure imposed on it by the turbulent flow. Inferences that molecular clouds must be gravitationally bound because of their high observed confinement pressures are called into question by these results. Regions with densities approaching the overall densities of GMCs, and pressures an order of magnitude above the average interstellar pressure appear in our simulations even in the absence of self-gravity.

Acknowledgements

I thank the referee of this review for a detailed and thoughtful report, my collaborators M. A. de Avillez, J. Ballesteros-Paredes, D. Balsara, A. Burkert, F. Heitsch, J. Kim, R. S. Klessen, V. Ossenkopf, and M. D. Smith for their participation in different parts of the work reviewed here, and the organizers of the conference for their partial support of my attendance. This work was also partially supported by the NSF under CAREER grant AST99-85392 and by the NASA Astrophysical Theory Program under grant NAG5-10103. This research has made use of NASA's Astrophysics Data System Abstract Service.

References

1. D. S. Adler, W. W. Roberts: Astrophys. J. **384**, 95 (1992)
2. J. Arons, C. E. Max: Astrophys. J. **196**, L77 (1975)
3. L. Armi, P. Flament: J. Geophys. Res. C, **90**, 11779 (1985)
4. S. A. Balbus, J. F. Hawley: Astrophys. J. **376**, 214 (1991)
5. S. A. Balbus, J. F. Hawley: Rev. Mod. Phys. **70**, 1 (1998)
6. J. Ballesteros-Paredes, L. Hartmann, E. Vázquez-Semadeni: Astrophys. J. **527**, 285 (1999)
7. J. Ballesteros-Paredes, M.-M. Mac Low: Astrophys. J., **570**, 734 (2002)
8. J. Ballesteros-Paredes, E. Vázquez-Semadeni, J. Scalo: Astrophys. J. **515**, 286 (1999)
9. J. Bally, D. Devine: Astrophys. J. **428**, L65 (1994)
10. D. S. Balsara: Astrophys. J. **465**, 775 (1996)
11. D. S. Balsara: Astrophys. J. Supp. **116**, 119 (1998a)
12. D. S. Balsara: Astrophys. J. Supp. **116**, 133 (1998b)
13. D. S. Balsara: J. Comput. Phys. **174**, 614 (2001)
14. D. S. Balsara, C. Norton: Parallel Comput. **27**, 37 (2001)
15. D. S. Balsara, D. S. Spicer: J. Comput. Phys. **148**, 133 (1999a)
16. D. S. Balsara, D. S. Spicer: J. Comput. Phys. **149**, 270 (1999b)
17. M. R. Bate, I. A. Bonnell, N. M. Price: 1995, Monthly Not. Roy. Astron. Soc., **277**, 362 (1995)
18. M. R. Bate, A. Burkert: Monthly Not. Roy. Astron. Soc. **288**, 1060 (1997)
19. F. Bensch, J.-F. Panis, J. Stutzki, A. Heithausen, E. Falgarone: Astron. Astrophys. **365**, 275 (2001)
20. F. Bensch, J. Stutzki, V. Ossenkopf: Astron. Astrophys. **366**, 636 (2001)
21. W. Benz: in *The Numerical Modelling of Nonlinear Stellar Pulsations*. ed. by J. R. Buchler, (Kluwer, Dordrecht, 1990), p. 269
22. D. Biskamp, W.-C. Müller: Phys. Plasmas **7**, 4889 (2000)
23. L. Blitz, F. H. Shu: Astrophys. J. **238**, 148 (1980)
24. S. Bonazzola, E. Falgarone, J. Heyvaerts, M. Perault, J. L. Puget: Astron. Astrophys. **172**, 293 (1987)
25. W. B. Burton: Astron. Astrophys. **10**, 76 (1971)
26. J. S. Carr: Astrophys. J. **323**, 170 (1987)
27. D. Clarke: NCSA Technical Report (1994)
28. T. Ebisuzaki, J. Makino, T. Fukushige, M. Taiji, D. Sugimoto, T. Ito, S. K. Okumura: Publ. Astron. Soc. Japan, **45**, 269 (1993)
29. B. G. Elmegreen: Astrophys. J. **419**, L29 (1993)

30. E. Falgarone, J.-F. Panis, A. Heithausen, M. Perault, J. Stutzki, J.-L. Puget, F. Bensch: Astron. Astrophys. **331**, 669 (1998)
31. E. Falgarone, J.-L. Puget, M. Pérault: Astron. Astrophys. **257**, 715 (1992)
32. G. B. Field, D. W. Goldsmith, H. J. Habing: Astrophys. J. **155**, L149 (1969)
33. C. F. Gammie, E. C. Ostriker: Astrophys. J. **466**, 814 (1996)
34. A. Gazol, E. Vázquez-Semadeni, F. J. Sánchez-Salcedo, J. Scalo: Astrophys. J. **557**, L121 (2001)
35. L. Hartmann, J. Ballesteros-Paredes, E. A. Bergin: Astrophys. J. **562**, 852 (2001)
36. J. F. Hawley, J. M. Stone: Comp. Phys. Comm., **89**, 1 (1995)
37. C. Heiles: Astrophys. J. **551**, L105 (2001)
38. A. Heithausen, P. Thaddeus: Astrophys. J. (Letters) **353**, L49 (1990)
39. F. Heitsch, M.-M. Mac Low, R. S. Klessen: Astrophys. J. **547**, 280 (2001)
40. F. Heitsch, E. G. Zweibel, M.-M. Mac Low, P. Li, M. L. Norman: Astrophys. J. **561**, 800 (2001)
41. R. W. Hockney, J. W. Eastwood: *Computer Simulation Using Particles* (Institute of Physics, Bristol, England, 1988)
42. M. Issa, I. MacLaren, A. W. Wolfendale: Astrophys. J. **352**, 132 (1990)
43. J. H. Jeans: Phil. Trans. A. **199**, 1 (1902)
44. W. H. Kegel: Astron. Astrophys. **225**, 517 (1989)
45. W.-T. Kim, E. C. Ostriker: Astrophys. J. **559**, 70 (2001)
46. R. S. Klessen, A. Burkert: Astrophys. J. Supp. **128**, 287 (2000)
47. R. S. Klessen, F. Heitsch, M.-M. Mac Low: Astrophys. J. **535**, 887 (2000)
48. R. Larson: Monthly Not. Roy. Astron. Soc. **194**, 809 (1981)
49. A. Lazarian, D. Pogosyan, E. Vázquez-Semadeni, B. Pichardo: Astrophys. J. **555**, 130 (2001)
50. J. Léorat, T. Passot, A. Pouquet: Monthly Not. Roy. Astron. Soc. **243**, 293 (1990)
51. R. B. Loren: Astrophys. J. **338**, 902 (1989)
52. M.-M. Mac Low: Astrophys. J. **524**, 169 (1999)
53. M.-M. Mac Low: in *Stars, Gas, & Dust in Galaxies*, ed. by D. Alloin, K. Olsen, G. Galaz (ASP, San Francisco, 2000), p. 55
54. M.-M. Mac Low, D. Balsara, M. A. de Avillez, J. Kim: Astrophys. J., submitted (2002, astro-ph/0106509)
55. M.-M. Mac Low, R. S. Klessen, A. Burkert, M. D. Smith: Phys. Rev. Lett. **80**, 2754 (1998)
56. M.-M. Mac Low, V. Ossenkopf: Astron. Astrophys. **353**, 339 (2000)
57. J. MacDonald, M. E. Bailey: Monthly Not. Roy. Astron. Soc. **197**, 995 (1981)
58. C. F. McKee, J. P. Ostriker: Astrophys. J. **218**, 148 (1977)
59. C. F. McKee: in *The Origin of Stars and Planetary Systems*, ed. by C. J. Lada and N. D. Kylafis (Kluwer, Dordrecht, 1999) p.29
60. H. Meyerdierks, A. Heithausen, K. Reif: Astron. Astrophys. **245**, 247 (1991)
61. J. J. Monaghan: Ann. Rev. Astron. Astrophys. **30**, 543 (1992)
62. T. Ch. Mouschovias: Astrophys. J. **211**, 147 (1977)
63. T. C. Mouschovias, L. Spitzer: Astrophys. J. **210**, 326 (1976)
64. C. A. Norman, A. Ferrara: Astrophys. J. **467**, 280 (1996)
65. V. Ossenkopf, M.-M. Mac Low: Astron. Astrophys.,**390**, 307 (2002)
66. E. C. Ostriker, C. F. Gammie, J. M. Stone: Astrophys. J. **513**, 259 (1999)
67. E. Ostriker, J. Stone, C. Gammie: Astrophys. J. **546**, 980 (2001)
68. P. Padoan, and Å. Nordlund: Astrophys. J. **526**, 279 (1999)
69. T. Passot, A. Pouquet, P. R. Woodward: Astron. Astrophys.**197**, 392 (1988)
70. T. Passot, & E. Vázquez-Semadeni: Phys. Rev. E **58**, 4501 (1998)

71. T. Passot, E. Vázquez-Semadeni, A. Pouquet: Astrophys. J. **455**, 536 (1995)
72. B. Pichardo, E. Vázquez-Semadeni, A. Gazol, T. Passot, J. Ballesteros-Paredes: Astrophys. J. **532**, 353 (2000)
73. D. H. Porter, A. Pouquet, P. R. Woodward: Phys. Rev. Lett. **68**, 3156 (1992)
74. D. H. Porter, A. Pouquet, P. R. Woodward: Phys. Fluids **6**, 2133 (1994)
75. J. C. Raymond, D. P. Cox, B. W. Smith: Astrophys. J. **204**, 290 (1976)
76. P. L. Roe, D. S. Balsara: SIAM J. Appl. Math. **56**, 57 (1996)
77. J. Scalo: in *Physical Processes in Fragmentation and Star Formation*. ed. by R. Capuzzo-Dolcetta, C. Chiosi, A. di Fazio (Kluwer, Dordrecht, 1990) p. 151
78. J. A. Sellwood, S. A. Balbus: Astrophys. J. **511**, 660 (1999)
79. F. H. Shu: Astrophys. J. **214**, 488 (1977)
80. P. R. Shapiro, R. T. Moore: Astrophys. J. **207**, 460 (1976)
81. M. D. Smith, M.-M. Mac Low: Astron. Astrophys. **326**, 801 (1997)
82. M. Steinmetz: Monthly Not. Roy. Astron. Soc. **278**, 1005 (1996)
83. J. M. Stone, M. L. Norman: Astrophys. J. Supp. **80**, 753 (1992a)
84. J. M. Stone, M. L. Norman: Astrophys. J. Supp. **80**, 791 (1992b)
85. J. M. Stone, E. C. Ostriker, C. F. Gammie: Astrophys. J. **508**, L99 (1998)
86. J. Stutzki, F. Bensch, A. Heithausen, V. Ossenkopf, M. Zielinsky: Astron. Astrophys. **336**, 697 (1998)
87. J. K. Truelove, R. I. Klein, C. F. McKee, J. H. Holliman II, L. H. Howell, J. A. Greenough: Astrophys. J. **489**, L179 (1997)
88. E. Vázquez-Semadeni, J. Ballesteros-Paredes, L. F. Rodriguez: Astrophys. J. **474**, 292 (1997)
89. E. Vázquez-Semadeni, T. Passot, A. Pouquet: Astrophys. J. **441**, 702 (1995)
90. E. Vázquez-Semadeni, T. Passot, A. Pouquet: Astrophys. J. **473**, 881 (1996)
91. M. G. Wolfire, C. F. McKee, D. Hollenbach, A. G. G. M. Tielens, E. L. O. Bakes, Astrophys. J. **443**, 152 (1995)
92. J. P. Williams, E. J. de Geus, L. Blitz: Astrophys. J. **428**, 693 (1994)
93. E. G. Zweibel, A. Brandenburg: Astrophys. J. **478**, 563 (1997, err: **485**, 920)
94. E. G. Zweibel, K. Josafatsson: Astrophys. J. **270**, 511 (1983)

Thermal Instability and Magnetic Pressure in the Turbulent Interstellar Medium

Enrique Vázquez-Semadeni[1], Adriana Gazol[1], Thierry Passot[2], and Javier Sánchez-Salcedo[3]

[1] Instituto de Astronomía, UNAM, Campus Morelia, Apdo. Postal 3-72, Morelia, Michoacán, MEXICO
[2] CNRS, Observatoire de la Côte d'azur, B.P. 4229, 06304, Nice, Cédex 4, France
[3] Instituto de Astronomía, UNAM, Apdo. Postal 70-264, México, D.F., 04510, México

Abstract. We review recent results on the nonlinear development of thermal instability (TI) in the context of the turbulent atomic interstellar medium (ISM), in which correlated density and velocity fluctuations are present, as well as forces other than the thermal pressure gradient. First, we present a brief summary of the linear theory, remarking that, in the atomic ISM, the condensation mode is unstable but the wave mode is stable at small scales. Next, we revisit the growth of isolated entropy perturbations in initially unstable gas, as a function of the ratio of the cooling to the dynamical crossing times η. The time for the dynamical transient state to subside ranges from 4 to 30 Myr for initial density perturbations of 20% and sizes 3 to 75 pc. When $\eta \ll 1$, the condensation produces locally supersonic motions and a shock propagates off the condensation, bringing the surrounding medium out of thermal equilibrium. Third, we consider the evolution of *velocity* perturbations, maintained by a random forcing, representing turbulent energy injection to the ISM from stellar sources. These perturbations correspond to the wave mode, and are stable at moderate amplitudes and small scales, as confirmed numerically.

We then consider the behavior of magnetic pressure in turbulent regimes. Various observational and numerical results suggest that the magnetic pressure does not correlate well with density at low and intermediate densities. We propose that this is a consequence of the slow and fast modes of nonlinear MHD waves being characterized by different scalings of the magnetic field strength versus density. This lack of correlation suggests that, in fully turbulent regimes, the magnetic field may not be a very efficient source of pressure, and that polytropic descriptions of magnetic pressure are probably not adequate.

Finally, we discuss simulations of the ISM (and resolution issues) tailored to investigate the possible existence of significant amounts of gas in the "lukewarm" temperature range between the warm and cold stable phases. The mass fraction in this range increases, and the phase segregation decreases, as smaller scales are considered. We attribute this to two facts: the enhanced stability of moderate, adiabatic-like velocity fluctuations with $\eta \gg 1$ and the recycling of gas from the dense to the diffuse phase by stellar energy injection. Moreover, the magnetic field is not strongly turbulent there, possibly providing additional stability. We conclude by suggesting that the gas with unstable temperatures can be observationally distinguished through simultaneous determination of two of its thermodynamic variables.

E. Falgarone and T. Passot (Eds.): LNP 614, pp. 213–251, 2003.
© Springer-Verlag Berlin Heidelberg 2003

1 Introduction

The fact that the neutral atomic interstellar medium (ISM) is most likely thermally bistable [68,23,82] has had a great impact on our picture of interstellar structure formation. Indeed, in two of the most influential models of the ISM to date, the famous two- and three-phase models of the ISM of Field, Goldsmith & Habing [23] and McKee & Ostriker [53], the concepts of thermal and pressure equilibria played a fundamental role, so that distinct *phases* (thermodynamic regimes with different density and temperature, but the same pressure) were predicted to coexist in pressure equilibrium. These phases correspond to stable thermal-equilibrium (i.e., heating-cooling balance) temperature regimes, and are separated by unstable regimes that, in those models, were therefore not expected to be present in the ISM. An opposite view was taken in the so-called time-dependent model of the ISM of Gerola, Kafatos & McCray [27], which was presented as an alternative to the pressure equilibrium two-phase model, and which made radically different assumptions: a constant density in the presence of stochastic, local heating events that caused strong local fluctuations of pressure and temperature, because the cooling and recombination times are comparable or shorter than the time between successive exposures of a given gas parcel to those heating events. This model predicted that significant amounts of gas should be in the unstable range, as they cooled after the transient heating events. More recently, Lioure & Chiéze [48] have considered models with a continuous recycling of gas among the various gas phases due to stellar energy injection, also concluding that significant amounts of gas should populate the unstable temperature range in the ISM. Note that the three-phase model [53] did consider the existence of local fluctuations in the pressure, although it was still based on the premise of "rough pressure balance".

Nevertheless, both the equilibrium and the time-dependent models omitted a number of important aspects in the ISM budget. The multiphase equilibrium models essentially neglected the possibility of large pressure fluctuations in the ISM. The time dependent model instead included this possibility as a fundamental premise, but neglected the fact that such pressure fluctuations should induce motions, which should in general be turbulent (i.e., spanning a wide range of scales), and in turn cause strong density fluctuations [18,3]. Moreover, both the time-dependent and the three-phase models omitted other important agents of the ISM, such as magnetic fields, rotation, and cosmic rays. Elmegreen [16,19] performed a combined instability analysis including self-gravity, cooling and heating, and magnetic fields, but the effects of turbulence, which is an inherently nonlinear phenomenon, can only be dealt with by means of numerical simulations of the gas dynamics in the Galactic disk in the presence of thermal instability (TI). The role of the turbulent motions may be crucial. In fact, realistic cloud/intercloud structure has been reported in models incorporating turbulence from stellar-like driving and cooling, but not necessarily a thermally bistable regime [4,10,11,69,78,66,25]. Similar results have been reported for pressureless (Burgers-like) models with stellar driving [73,9], and for simulations of interacting nonlinear MHD waves [20]. Thus, it is important to investigate the

role of TI in determining the distribution of the physical variables (density, temperature, velocity) of the flow, and, in particular, the degree to which phase segregation, as was proposed in the multiphase models, is realized, in the context of a turbulent ISM with multiple sources of turbulent energy at a variety of scales, such as stellar winds, supernova explosions, spiral arm passage, magnetorotational instabilities [75], etc., besides TI.

Although the nonlinear development of TI has been studied extensively for decades now (e.g., [29,74,56,57,61,70,13,58,59,7,44,37,38,8]), only recent work has started to investigate the interplay between TI and the turbulent nature of the ISM, such as, for example, the triggering of TI by external compressions ([31,32,42]) and the possibility that the TI itself may contribute to the generation of turbulence in the ISM [81,43,45].

However, the fact that additional energy sources feeding ISM turbulence besides TI itself, such as stellar energy injection, or large-scale gravitational or magnetic instabilities, has additional implications. First, the very presence of strong motions implies that transport (advection) should be important, while traditionally conductive processes have received more attention (e.g., [83,5,55]). Second, these transport processes may imply the existence of constant fractions of gas transiting through the unstable regime, and erase, to some extent, the phase segregation expected in multiphase models. These expectations are furthered by several observational studies (e.g., [15,36,77,24,30]) that have suggested that the fraction of gas in the unstable range between the cold and warm phases of the atomic ISM is substantial.

Another important issue to consider is the fact that the ISM is magnetized, which suggests the possibility that turbulent magnetic pressure may supplement thermal pressure and somehow counteract TI. The condensation process in a magnetized medium has been studied by several workers (see, e.g. [22,20,61,50,32]; see also the references given in [32]), concluding that, although condensation can be inhibited under some circumstances, it is in general possible. However, those studies have not considered the case of TI developing in an externally-driven turbulent medium, except for [32].

In this paper we review recent work and present new results concerning the interplay between TI and turbulence in the warm and cool ISM. First, we review the main aspects of the instability in Sect. 2. Then, in Sect. 3, we revisit the growth of isolated density fluctuations, focusing in particular on the late stages and the state of the gas surrounding the condensation. Next we discuss the development or suppression of growth in the presence of random velocity fluctuations, stressing that these probably constitute the most common way of inducing density fluctuations in the ISM. We then briefly discuss the nature of magnetic pressure on turbulent media (Sect. 5) and its dependence on density. In Sect. 6, we discuss the role of TI in numerical models of the ISM in the presence of magnetic fields, the Coriolis force, modeled star formation, self-gravity and TI, aiming at determining the fraction of unstable gas, and at interpreting the results in the light of the previous sections. We include an extensive discussion of

numerical tests to maximize the reliability of the results. Finally, we summarize the results and mention a number of implications in Sect. 7.

2 Review of the Linear Theory and the Role of the Ratio of Cooling Time to Crossing Time

The ISM in general is subject to heating and cooling processes that cause it to have an effective thermal behavior very different from that of an ideal gas. Our current understanding of the physical processes responsible for the heating and cooling of the neutral atomic ISM can be found in Wolfire et al. [82]. Their effect is customarily collected in a net cooling function per unit mass $\mathcal{L} = \rho\Lambda - \Gamma$, where ρ is the gas mass density, Λ is the cooling rate, and Γ is the heating rate. In this paper we will use a piecewise power-law fit to the "standard" cooling curve of [82], of the form

$$\Lambda = C_{i,i+1} T^{\beta_{i,i+1}} \quad \text{for} \quad T_i \leq T < T_{i+1}, \tag{1}$$

with the coefficients and exponents given in Table 1, and a constant heating rate $\Gamma = \Gamma_0$, which is a reasonable approximation to the weak dependence $\Gamma \sim \rho^{0.2}$ across the density range of the atomic medium [82]. The fit is obtained by first fitting a piecewise power law to the standard thermal equilibrium-pressure vs. density curve of [82] to obtain the exponents of the cooling curve, and then determining the coefficients by equating the cooling rate to the constant heating $\Gamma_0 = 0.015$ erg $\mathrm{s}^{-1}\mathrm{g}^{-1}$. The values of the coefficients have units of erg $\mathrm{s}^{-1}\mathrm{g}^{-2}\mathrm{cm}^3\mathrm{K}^{-\beta_{i,i+1}}$.

Table 1. Cooling function parameters

Interval	T_i/K	n_i/cm^{-3}	$C^{\mathrm{a}}_{i,i+1}$	$\beta_{i,i+1}$
	15.0	80.0		
(1,2)			3.42×10^{16}	2.13
	141.	7.00		
(2,3)			9.10×10^{18}	1.00
	313.	3.16		
(3,4)			1.11×10^{20}	0.565
	6101.	0.59		
(4,5)			2.00×10^{8}	3.67
	10^5			

[a] In units of erg $\mathrm{s}^{-1}\mathrm{g}^{-2}\mathrm{cm}^3\mathrm{K}^{-\beta_{i,i+1}}$.

Subject to these processes, and in the absence of magnetic fields, self-gravity and other agents, the dynamics of the gas are described by the equations [22,55]

$$\frac{d\rho}{dt} + \rho\nabla\cdot\mathbf{u} = 0, \tag{2}$$

$$\rho\frac{d\mathbf{u}}{dt} = -\nabla P \tag{3}$$

$$\frac{1}{\gamma-1}\frac{dP}{dt} + \frac{\gamma}{\gamma-1}P\nabla\cdot\mathbf{u} + \rho\mathcal{L}(\rho,\mathcal{T}) - \nabla\cdot(K\nabla T) = 0, \tag{4}$$

where $\mathbf{u}$ is the fluid velocity, P is the thermal pressure, T is the temperature, γ is the heat capacity ratio, $K = K(T)$ is the thermal conductivity, and the ideal-gas equation of state, $P = \rho RT/\mu$, is assumed, with R being the universal gas constant and μ the mean molecular weight in units of the hydrogen mass. For clarity, we remind the reader that $R = k_{\rm B}/m_{\rm H}$, where $k_{\rm B}$ is the Boltzmann constant, and $m_{\rm H}$ is the mass of the Hydrogen atom.

The linear analysis leading to TI was first performed in full in the classic paper of Field [22], and later generalized to the case of a flow in motion [33,34]. Other useful, more recent presentations may be found in [17,76,55]. The analysis, assuming perturbations proportional to $\exp(nt + i\mathbf{k}\cdot\mathbf{x})$ in all variables, yields the cubic dispersion relation

$$n^3 + n^2\left[\frac{N_\rho}{c_V} + \frac{ck^2}{k_K}\right] + nc^2k^2 + c^2k^2\left[\frac{N_P}{c_P} + \frac{ck^2}{\gamma k_K}\right] = 0, \tag{5}$$

where n is the growth rate, k is the wavenumber, $c \equiv \gamma kT/\mu$ is the adiabatic sound speed, and

$$N_\rho \equiv \left(\frac{\partial\mathcal{L}}{\partial T}\right)_\rho, \qquad N_P \equiv \left(\frac{\partial\mathcal{L}}{\partial T}\right)_P = \left[N_\rho - \frac{\rho_0}{T_0}\left(\frac{\partial\mathcal{L}}{\partial\rho}\right)_T\right], \tag{6}$$

with the subindex "0" denoting values in the unperturbed equilibrium state, and

$$k_K \equiv \frac{R}{\mu}\frac{c\rho_0}{(\gamma-1)K}. \tag{7}$$

As pointed out by Field, k_K is the mean free path of the gas molecules (see also [76], p.35).

The dispersion relation (5) has three roots, one of them being always real, and the other two being either a complex conjugate pair, or a pair of real numbers. There is instability whenever $\Re(n) > 0$, where $\Re()$ denotes the real part. The positive real root corresponds to exponential growth without propagation, and is thus called the *condensation* mode. The pair of complex roots corresponds to oscillatory behavior, and is thus called the *wave mode*. This mode grows in amplitude (i.e., is *overstable*) when the real part of those roots is positive. This nomenclature was extended by Field to the case when the roots are real, in which case he said the wave mode is *overdamped* (i.e., does not oscillate).

The condensation mode is unstable if the so called "isobaric" criterion, namely

$$N_P < 0, \tag{8}$$

is satisfied. However, when only this criterion is satisfied, the growth rate vanishes as $k \to 0$. For the growth rate to remain finite at long wavelengths, it is necessary to also satisfy the "isochoric" criterion, which reads

$$N_\rho < 0. \tag{9}$$

The corresponding instability criterion for the wave mode is the so called "isentropic" (or "adiabatic") criterion, reading

$$\frac{N_P}{c_P} - \frac{N_\rho}{c_V} > 0, \tag{10}$$

where c_P and c_V are the specific heats at constant pressure and at constant volume, respectively. Note that $c_V = R/\left[(\gamma-1)\mu\right]$. For the cooling and heating functions adopted here, the isobaric, isochoric and isentropic criteria respectively imply $\beta < 1$, $\beta < 0$ and $\beta < 1/(1-\gamma) = -3/2$.

Since in the atomic ISM only the isobaric condition (8) is satisfied, and given the focus of this paper on that medium, we will concentrate on this case in what follows. It is worth recalling that isobaric and isochoric perturbations (accomplished, for example, by setting up density fluctuations at constant pressure or temperature fluctuations at constant density, respectively), are generically referred to as *entropy* preturbations, since they imply a variation of the ratio P/ρ^γ, which remains constant for reversible isentropic processes. We now discuss the evolution of entropy perturbations in detail.

2.1 Entropy Perturbations

Figure 1 shows the growth rate as a function of wavenumber for our fit to the "standard" cooling curve of [82], (1) and a realistic value of the conductivity $K = K_0 = 5/3(k_B T_0 l/v_{rms}) n_0 (3k_B/2m)$ [47], with $T_0 = 2400$ K, $n_0 = 1$ cm^{-3}, $m = m_H$, $l = 3.2\times10^{-3}$ pc and $v_{rms} = 5.7$ kms^{-1}. In the case of the pure development of the instability, without any external forcing processes, there are three clearly distinct scale ranges for the wavelength λ that arise from the presence of three characteristic time scales [55]: the dynamical time, $\tau_{\rm d}$, in this case given by the sound crossing time, $\tau_{\rm s} = c/\lambda$; the cooling time, which for isobaric processes is given by[1]

$$\tau_{\rm c} = \frac{\gamma R}{(\gamma-1)\mu|N_P|}, \tag{11}$$

but which, to order of magnitude, is in general $\tau_{\rm c} \approx c_V T/(\rho\Lambda)$; and the conductive time, $\tau_\kappa = \lambda^2/\kappa_0$, where $\kappa_0 = K_0/\rho_0$ is the thermal diffusivity. These

[1] This expression is obtained by linearizing the energy equation (4) and then computing $\tau_{\rm c} = \delta e/(d\delta e/dt)$ for an isobaric process.

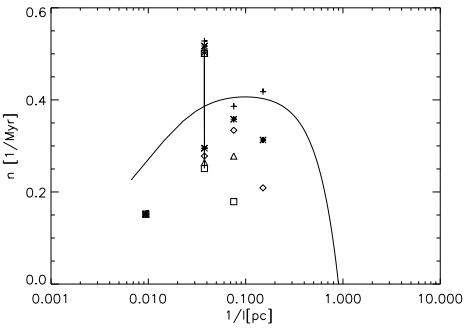


Fig. 1. (*Solid line*): Analytical growth rate as a function of wavenumber as given by (5) using our piecewise power-law fit to the "standard" cooling function of Wolfire et al. [82] and a constant heating rate. The wavenumber at which the growth rate becomes zero corresponds to the Field length. (*Points*): Numerical growth rates in simulations of the linear growth of 2.5% isobaric density perturbations in boxes of physical size 150 pc (see Sect. 6.1) for several values of the mass diffusion coefficient (see Sect. 3): $\mu_m = 0$ (*crosses*); $\mu_m = 0.001$ (*asterisks*); $\mu_m = 0.002$ (*diamonds*); $\mu_m = 0.004$ (*triangles*); $\mu_m = 0.008$ (*squares*). For perturbations with $l = 1/0.05 = 18.75$ pc (1/8 of the box size), which are in the transition regime between long and short wavelengths, growth rates for perturbations with zero velocity (appropriate for the short-wavelength regime) are shown by points lying below the theoretical curve, while rates for perturbations with self-consistent density, velocity and pressure (appropriate for the long-wavelength regime) are shown by points lying above the curve.

time scales then define two characteristic length scales. First, the so called Field length [5] $\lambda_{\rm F} = 2\pi/k_{\rm F}$, with $k_{\rm F} = (|N_P|/\kappa_0)^{1/2}$, at which $\tau_\kappa \sim \tau_{\rm c}$, and below which the growth of perturbations is inhibited by thermal conduction. Second, the scale at which $\tau_{\rm d} \sim \tau_{\rm c}$, which we shall denote $\lambda_{\rm eq}$. In the remainder of the paper we shall use the notation $\eta \equiv \tau_{\rm c}/\tau_{\rm d}$. The three scale ranges are then the "long" ($\lambda \gg \lambda_{\rm eq}$), "intermediate" ($\lambda_{\rm eq} \gg \lambda \gg \lambda_{\rm F}$) and "short" ($\lambda_{\rm F} \gtrsim \lambda$), where $\lambda = 2\pi/k$ is the perturbation wavelength. For the long-wavelength range, $\eta \ll 1$, while for the intermediate- and short-wavelength cases, $\eta \gg 1$. In what follows we will often discuss in terms of η, as we consider that the relevant physical quantities involved are the cooling and sound-crossing time scales, even though it is customary in the literature to base the discussion on the perturbation wavelength.

In the case of short- and intermediate-wavelength perturbations [2], the condensation mode evolves nearly isobarically, as $\tau_{\rm c} \gg \tau_{\rm d}$ and thus sound waves have ample time to restore pressure equilibrium while the gas cools. This also has the consequence that, in the case of zero diffusivity, the growth rate asymptotically approaches the cooling rate in the limit of large wavenumbers. In the presence of diffusivity, the growth rate decreases again in the short-wavelength

[2] For convenience, in the remainder of this paper, we will group intermediate- and small-scale perturbations into the small-wavelength ($\eta > 1$) category.

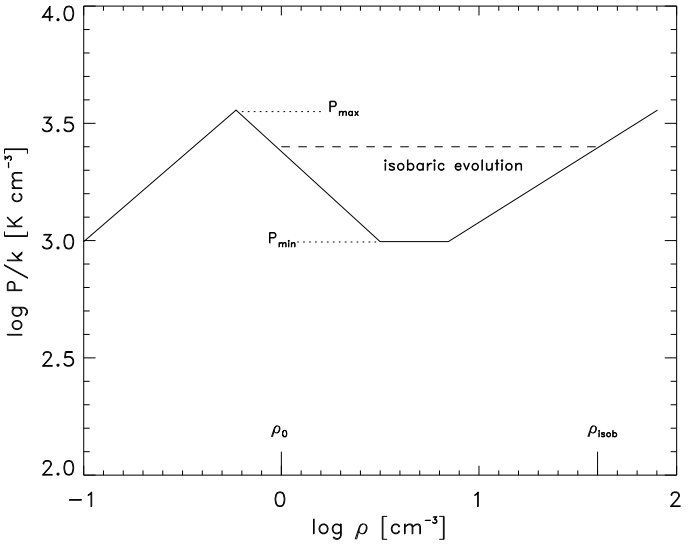


Fig. 2. Thermal-equilibrium piecewise polytropic behavior of the pressure for the cooling functions used in Papers I and IV. $\rho_{\rm isob}$ is the value of the density within the cold phase whose corresponding thermal-equilibrium pressure equals that of the mean density ρ_0.

range, due to the action of thermal conduction. These properties are illustrated in Fig. 1.

In the opposite limit of very large wavelengths, the condensation mode initially behaves isochorically (even though the isochoric criterion is not satisfied), since $\eta \ll 1$, so that the cooling acts much more rapidly than the sound waves can travel accross the perturbation to restore pressure balance. This implies that large pressure gradients can be set up, which in turn can trigger strong motions that can become locally supersonic [22,54,2,55]. Moreover, in this regime, the thermal pressure is given by the condition of thermal equilibrium, because the rapid cooling always allows its establishment. Figure 2 shows the equilibrium-pressure versus density for our piecewise cooling function (1). The slope of this graph constitutes an effective polytropic exponent given by $\gamma_{i,i+1} = 1 - 1/\beta_{i,i+1}$ (c.f. (1)), so that the pressure behaves as $P \propto \rho^{\gamma_{i,i+1}}$. In this figure, we denote by $\rho_{\rm isob}$ the density value within the cold phase that corresponds to the same pressure as that at the mean density. The instability under the isobaric mode is seen as the negative-slope range 0.6 cm$^{-3} \lesssim \rho \lesssim 3.2$ cm^{-3} (equivalent in thermal equilibrium to the temperature range $300 \lesssim T \lesssim 6000$ K). Note also the marginally stable, $\gamma_{\rm eff} = 0$ behavior in the range 3.2 cm$^{-3} \lesssim \rho \lesssim 7.1$ cm^{-3}.

In this same limit (long wavelengths), the growth rate asymptotically approaches [55]

$$n_{\rm long} = \pm kc \left(-\frac{N_P}{N_\rho} \right)^{1/2}, \tag{12}$$

which is of the order of the inverse of the sound crossing time, explaining its vanishing as $k \to 0$ (Fig. 1). Note that the situation is different when the isochoric criterion is satisfied, in which case, the growth rate remains roughly constant at $\tau_{\rm c}^{-1}$ over the long- and intermediate-wavelength ranges.

2.2 Adiabatic Perturbations

The case of adiabatic perturbations is perhaps the most relevant for a turbulent ISM since the adiabatic condition implies that the density and pressure gradients have the same sign, a situation which is naturally accomplished by means of a compressive motion acting on time scales much shorter than the cooling time. In contrast, for entropy perturbations the density and pressure gradients can have opposite signs. We refer to the latter case as a "reversed" pressure gradient, while we say that a pressure gradient with the same sign as the density gradient is "regular".

In the presence of velocity fluctuations, a new characteristic time scale appears in the system, namely the bulk-velocity crossing time, $\tau_{\rm u} = \lambda/u$, where u is the characteristic velocity of the perturbation. Thus, it is convenient to redefine the dynamical time scale as $\tau_{\rm d} = \min(\tau_{\rm s}, \tau_{\rm u})$, so that the definition $\eta = \tau_{\rm c}/\tau_{\rm d}$ can be preserved. Adiabatic perturbations with $\eta \gg 1$ become unstable when the adiabatic instability criterion (10) is satisfied [22,76], and excite the wave mode of the instability, which in this case consists of nearly adiabatic sound waves with growth rate $n = ikc + (1/2)(N_P/c_P - N_\rho/c_V)$, where the imaginary part gives the propagation speed and the real part gives a modulation that grows only if the adiabatic criterion is satisfied (overstability). However, at $\eta < 1$ (long wavelengths), the complex conjugate roots become real, and the wave mode becomes a condensation mode, which is unstable whenever $N_P/(\gamma N_\rho) < 0$ [76]. For $N_\rho > 0$, as is usually the case, this reduces to the isobaric criterion. Thus, in the atomic ISM, *wave-like (or "velocity") perturbations are linearly unstable only for* $\eta < 1$*, i.e, in the long-wavelength limit.* For nonlinear perturbations, however, the density increase in the waves can accelerate the cooling and locally cause η to become < 1, allowing the instability to proceed even in cases of perturbations of initial short-wavelength perturbations. A similar effect can occur even when the compression acts on the warm stable phase [31,42]. We will discuss this further in Sect. 4.

2.3 Entropy vs. Adiabatic Fluctuations

It is now important to ask what kind of processes in the real world can generate the two kinds of fluctuations: entropy or adiabatic. In the absence of initial fluid motions, entropy fluctuations can only be produced by locally varying the cooling-to-heating ratio. On the other hand, if velocity fluctuations are used as the driver of the density fluctuations, as occurs in a turbulent medium, then the density fluctuations can either behave as entropy or as adiabatic fluctuations depending on scale and on the velocity amplitude. The production of entropy-like perturbations clearly requires that $\tau_{\rm u} \gg \tau_{\rm c}$, so that the thermal response

of the flow proceeds under thermal equilibrium. At small scales, where $\eta \gg 1$, this then implies that $\tau_{\rm u} \gg \tau_{\rm c} \gg \tau_{\rm s}$, so that the motions have to be essentially quasi-static. At large scales, where $\eta \ll 1$, we see that even supersonic motions can produce entropy-like fluctuations, as long as $\tau_{\rm u}, \tau_{\rm s} \gg \tau_{\rm c}$. The remaining possibility, i.e, $\tau_{\rm u} \sim \tau_{\rm s} \ll \tau_{\rm c}$, occurring for finite-amplitude velocity fluctuations at small scales, causes adiabatic-like perturbations.

Another important distinction is that, for entropy perturbations, the motions are driven by the thermal pressure gradient generated by the instability, and tend to restore pressure equilibrium. Instead, in the case of externally-driven velocity perturbations, the motions drive the density and pressure gradients, which then feed back on the cooling and the motions themselves. Thus, the cause-effect relationship between the motions and the thermal pressure gradient are essentially reversed in the two cases.

2.4 The Magnetic Case

The linear instability analysis in the presence of a uniform magnetic field B was also studied by Field [22]. Here we just briefly summarize his main results, and then discuss some recent work in the nonlinear regime.

Qualitatively, Field concluded that the inclusion of the magnetic field should introduce three main modifications to the non-magnetic results. First, the wave mode splits into three modes, which correspond to the three modes of MHD waves: Alfvén, slow and fast. Of these, the Alfvén mode is "neutral", in the sense that it does not interact with the instability, since it is strictly non-compressive, while the fast and slow modes are governed by the same isentropic criterion. Second, the condensation mode is unaffected when the vector wavenumber $\mathbf{k}$ is parallel to $\mathbf{B}$, while the field has a stabilizing effect when $\mathbf{k}$ is perpendicular to $\mathbf{B}$. Finally, heat conduction is greatly reduced in the direction perpendicular to the field because of the spiraling of electrons between collisions. In summary, no major modifications to the overall picture were foreseen by Field, even though the dispersion relation changes from cubic to fifth-degree. These results were verified numerically by Goldsmith [29].

More recently, Loewenstein [50] has extended Field's linear analysis to the case of a stratified backgound medium, showing that condensation modes do exist in cooling flows, and that, over a certain wavenumber range, the presence of the magnetic field suppresses the damping of the instability due to conduction. In work more closely related to our focus in this paper, Hennebelle & Pérault [32] have investigated the role of a strong compression wave, interpreted as a turbulent velocity fluctuation, on triggering the *nonlinear* instability in the linearly stable diffuse phase in a magnetized medium already segregated into two phases. These authors showed that the instability can indeed be triggered when the directions of the compression and of the initial magnetic field are oblique, giving the threshold values of the angle between them for condensation to occur at various values of the compression Mach number and of the magnetic field strength. In the forthcoming sections we will first describe results concerning the development of TI from an initially unstable medium under both quiescent and

turbulent conditions, in the non-magnetic case. Next we discuss the nature of magnetic pressure in turbulent media, and finally we will consider more complete models of the ISM incorporating all of these agents and processes.

3 Nonlinear Evolution of Entropy Perturbations

As a first step in our discussion of dynamical aspects of the development of TI, in this section we discuss the conditions necessary for the development of large velocities (possibly supersonic) and shocks during the spontaneous (i.e., in the absence of external triggers) condensation process of entropy perturbations in the unstable atomic ISM, as a function of the parameter η. A detailed review of the nature of the shocks has been presented in [55]. We are also interested in the duration of the dynamic phase. Studies dealing with the development of supersonic motions and shocks as a consequence of the condensation process from the unstable regime have mostly focused on the regimes of proto-galaxy-cluster, proto-galactic, and proto-globular cluster clouds (e.g., [13,70,58,59,7,44,38]). In particular, Sasorov [70] pointed out that the development of TI in three dimensions should give rise to flattened structures, similar to the "pancakes" formed by gravitational contraction in the cosmological flow.

In the context of the ISM, the nonlinear development of isolated entropy fluctuations was initially studied by Goldsmith [29] and Schwartz, McCray & Stein [74], who found that the condensation of small-scale ($\eta > 1$) perturbations occurred on time scales of $\sim$ 1–10 Myr, and produced clouds of densities 100$\times$ larger than their initial values. Goldsmith also considered the case of large-scale perturbations, finding the development of transonic velocities, and that the time required for reaching a true steady state is much longer than the time of "initial collapse", not being reached by any of his simulations. However, these works were performed at very low resolutions, and did not discuss the state of the surrounding gas in much detail.

More recently, Burkert & Lin [8] have considered a cooling-only medium (i.e., without background heating), and suggested that a special clump scale can be selected by the following mechanism. In the case of a globally cooling medium which eventually exits the thermally unstable range in roughly one cooling time, large-scale ($\eta < 1$) fluctuations that cool isochorically do not change their density appreciably before exiting the unstable range, so that, after they do, the pressure gradient becomes regular again, and the perturbation is erased. On the other hand, small-scale fluctuations can eventually reach the regime of isochoric cooling with $\eta < 1$ as their density and local cooling rate increase, thus developing large velocities. In this case, advection overtakes the pressure gradient in promoting the compression, which then proceeds at an accelerated rate. Thus, Burkert & Lin suggested that the latter fluctuations are the ones that dominate the fragmentation of a large cloud into clumps, determining the clump properties.

In the presence of both cooling and heating, Sánchez-Salcedo et al. [72] (hereafter Paper I) have recently investigated the evolution of perturbations as a function of size (or, equivalently, η), focusing on the magnitudes of the veloci-

ties that develop, the time scales for reaching a relatively quiescent stage, and the state of the gas surrounding the condensation after the latter has reached a quasi-stationary state. To this end, Paper I performed one-dimensional (1D), high-resolution numerical simulations of the evolution of Gaussian-shaped perturbations in the absence of any other physical processes. The simulations solve the gas dynamic equations in the presence of heating and cooling parameterized as in (1), reading

$$\frac{D\ln\rho}{Dt} = -\frac{\partial u}{\partial x} + \frac{1}{\rho}\mu_m \frac{\partial^2 \rho}{\partial x^2}, \tag{13}$$

$$\frac{D\mathbf{u}}{Dt} = -\frac{1}{\rho}\frac{\partial P}{\partial x} + f + \frac{1}{\rho}\frac{\partial(2\nu\rho S)}{\partial x}, \tag{14}$$

$$T\frac{Ds}{Dt} = \Gamma - \rho\Lambda + 2\nu S^2 + \text{diffusion term}, \tag{15}$$

where $D/Dt = \partial/\partial t + u\partial/\partial x$ is the Lagrangian derivative, $S = (2/3)\partial u/\partial x$ is the generalized strain tensor in the 1D case, s is the entropy per unit mass, f is the random forcing (described and used in Sect. 4), and the other quantities have their usual meanings. A shock-capturing viscosity of the form

$$\nu = \nu_0 + c_\nu \delta x^2 \max(0, -\nabla\cdot\mathbf{u}) \tag{16}$$

is used, with $\nu_0 \sim 10^{-2} c \Delta x_{\min}$ and c_ν being constants. ν_0 is the bulk viscosity, and $\Delta x_{\min}$ is the minimum grid spacing. The last term in (13) is an artificial mass diffusion term, necessary for smoothing excessively large density gradients. The term $2\nu S^2$ in (15) is the viscous heating of the gas. The diffusion term in the same equation has the form $T(\nabla^2 s + \nabla s\cdot(\mu\nabla\ln\rho))$, and is included to guarantee that any mass redistribution due to the mass diffusion term produces a self-consistent entropy redistribution. The coefficient μ_m of these terms is maintained at very low values so that these diffusivities are comparable to the numerical diffusivity, giving a diffusive scale of size 2–3 grid zones. Although the use of numerical diffusivities is inevitable in virtually any numerical study, the simulations use up to 7000 grid points, and non-uniform grid spacing, in order to maximize the resolution at the condensation center, and thus minimize the impact of the diffusivities on the results. Moreover, all the results reported in this section have been subjected to convergence tests to guarantee that they are not altered by changing the resolution (see Paper I).

The simulations start at a density of 1 cm^{-3}, roughly the mean ISM density in our galaxy, as pointed out in Sect. 2.4 from [21], which lies in the unstable range of the cooling curve. The equilibrium temperature at this density is $T \approx 2400$ K. The scale $\lambda_{\rm eq}$ at which the cooling and sound crossing times are equal is $\lambda_{\rm eq} \sim 10$ pc, while the Field length under these conditions is $\lambda_{\rm F} \sim 0.015$ pc (cf. Sect. 2.1 and Fig. 1).

We first consider the case of small-scale ($\eta > 1$) entropy perturbations, as they constitute the paradigm of cloudlet (i.e., small clouds of sizes $\lesssim 1$ pc and densities ~ 50 cm^{-3}) formation by TI in the ISM (e.g., [23,29,74,64,48]). To this

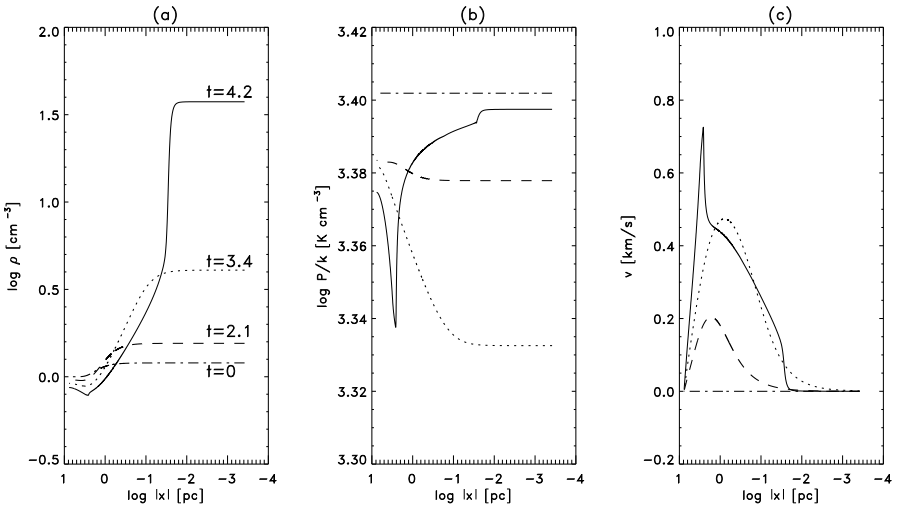


Fig. 3. Time development of run DEN3. The density, pressure and velocity profiles at different times are plotted in panels (a), (b) and (c), respectively. The times corresponding to each line type are indicated in frame (a) in Myr, and the same line labeling is used in frames (b) and (c).

end, we have performed a simulation, labeled DEN3, of the evolution of a Gaussian density perturbation of 20% amplitude and a full width at half maximum (FWHM) of 3 pc in a numerical box of size 10 pc. Figure 3 shows the density, pressure and velocity profiles of the cloud at various times until the time when a "cloud" has formed and the accretion process has mostly subsided. Note the logarithmic x-axis, where x is the distance to the center of the cloud. It is seen that the evolution is indeed quasi-isobaric, with variations in the pressure of less than 2%, and local Mach numbers which do not exceed 0.2. By $t = 4.2$ Myr, the condensation has essentially completed its evolution, and reached the pressure-equilibrium density, $\rho_{\rm isob}$. Figure 4 shows the evolution of this run on the P-ρ diagram, where the quasi-isobaric nature of the condensation is clearly seen.

Note, however, that in Figs. 3 and 4 a population of points is still seen to continue flowing onto the condensation, and as it does, it necessarily remains in the "unstable" range. The accretion and evacuation of the unstable gas will take very long times to complete, because the reservoir of unstable gas outside the cloud is very large ($\sim 90\%$ of the total mass), in agreement with the remark by Goldsmith [29]. Moreover, note that *the inflowing gas is not truly unstable*, as it does not lie on the equilibrium curve anymore. Instead, at $t = 4.2$ Myr, the density and pressure gradients have the same sign throughout this region. Thus this gas does not have a tendency to fragment any further, even though its density is in the "unstable" range. It can be said that this gas is *flowing* because of TI, but once it is doing so it has no tendency to fragment any further.

Finally, note that the cloud formation time is not very short, and is significantly sensitive to the initial amplitude. Simulations with an initial amplitude

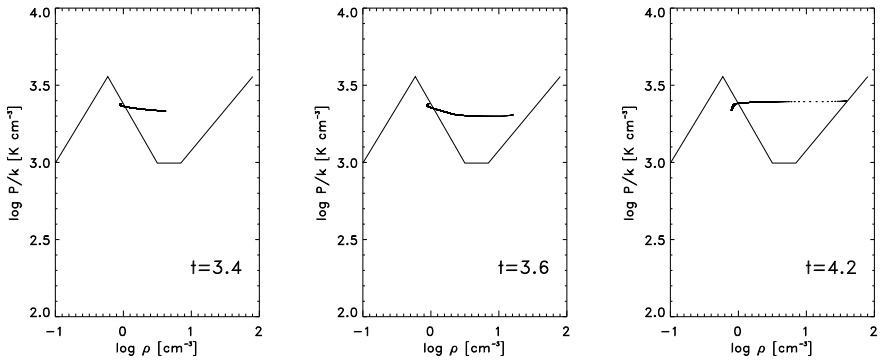


Fig. 4. Time development of run DEN3 in P–ρ phase space, at the times indicated in each frame. At $t = 4.2$ Myr, although the cloud has already formed, a substantial fraction of the points in the simulation are still traversing the unstable range, albeit in a nearly isobaric regime.

of 10% require ~ 5.5 Myr to complete the condensation. This time is comparable to the mean time between successive exposures to passing shock fronts from supernova remnants and superbubbles [40], so that the condensations may have their growth interrupted by external perturbations, as is the case in Sect. 4.

Let us now consider the opposite case of a large-scale entropy perturbation, with FWHM=75 pc and $\eta \sim 0.04$, in a box of 250 pc. We refer to this simulation as run DEN75. Its evolution is shown in Figs. 5 (density, pressure and velocity profiles) and 6 (evolution on the P-ρ plane). In this case, the evolution is significantly different. As dictated by the smaller growth rates of larger-scale perturbations, run DEN75 requires 30 Myr to complete the formation of a cloud, but moreover, throughout the first part of its development, the condensation proceeds along the thermal equilibrium curve (see the first four panels of Fig. 6), developing locally supersonic velocities in the process (maximum Mach number ~ 1.2) that cause a strong overshoot. Thus, this condensation transiently reaches densities $\sim 55\rho_{\rm isob}$. At the time of maximum compression, a strong shock is produced at the cloud boundary that propagates outwards from it. This shock has the important effect of heating the still-infalling gas, bringing it out of thermal equilibrium and closer to isobaric conditions (cf. last two panels of Fig. 6), and reversing the velocity gradient. This shock is located at the peak of the velocity for the various times shown in panel (f) of Fig. 5, and is seen also in the pressure (see panel (e) of the same figure). By the end of the simulation, the accretion ram pressure is still high enough that the condensation has $\rho \sim 3\rho_{\rm isob}$, and this value decreases extremely slowly with time. Again, as in run DEN3, the infalling gas is mostly in the "unstable" density range. Within the infalling region, the pressure and density gradients have the same sign, so this gas again has no further tendency to fragment. We have found from other simulations that the qualitative behavior of run DEN75 occurs down to initial fluctuations with FWHM=15 pc.

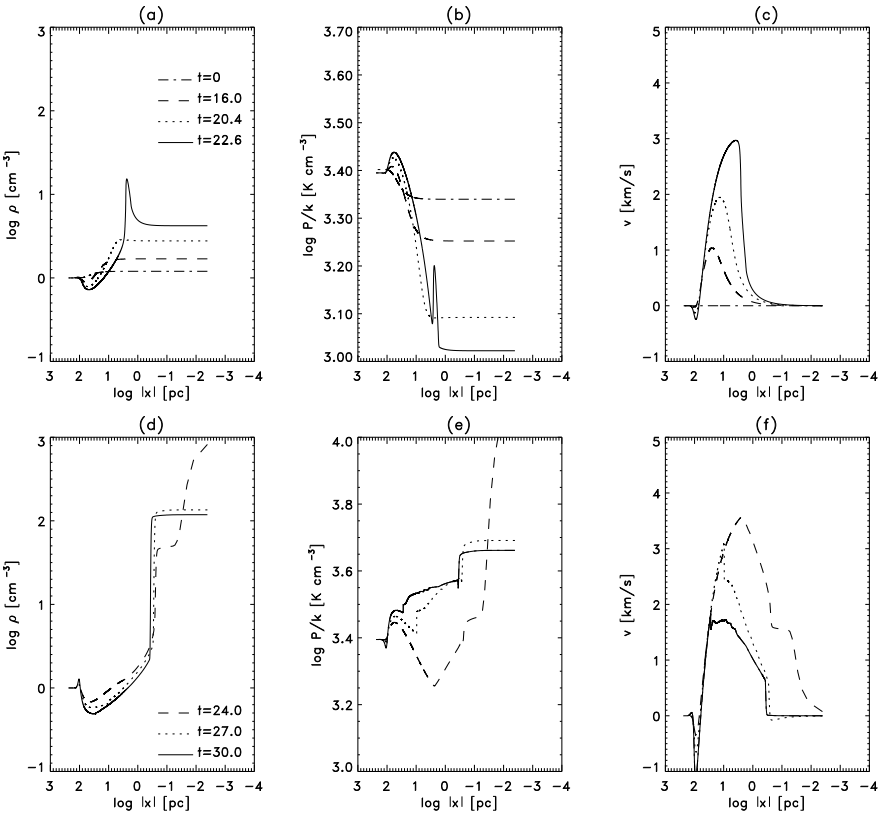


Fig. 5. Same as Fig. 3 but for run DEN75. The first part of the evolution is shown in the upper frames, and the rest in the lower frames. Note the formation of a shock shortly after $t = 22.6$ (frame c), which then propagates outwards from the cloud. At the same time, the density overshoots to over $55\rho_{\rm isob}$ (frame d). After the formation of the cloud, the density relaxes to a value $\sim 2.5\rho_{\rm isob}$, due to the ram pressure of the still infalling gas.

From the evolution of these two simulations, we conclude that large-scale ($\eta \ll 1$) entropy perturbations have such a dynamic evolution that their final central density and pressure are larger than those corresponding to plain thermal-pressure equilibrium with the diffuse phase, and moreover require such long times to evolve (over 20 Myr to the occurrence of the large density overshoot), that they are unlikely to complete their evolution before being disrupted by other perturbations in the real ISM, such as passing shock waves, or simply, general turbulent fluctuations. Small-scale entropy perturbations ($\eta \gtrsim 1$), on the other hand, adhere better to the paradigm of forming near pressure-equilibrium condensations, although we have seen that a significant fraction of the mass still lies in the unstable range after the cloud has formed, and is accreting onto the condensation, causing the presence of (weak) accretion fronts (rather than con-

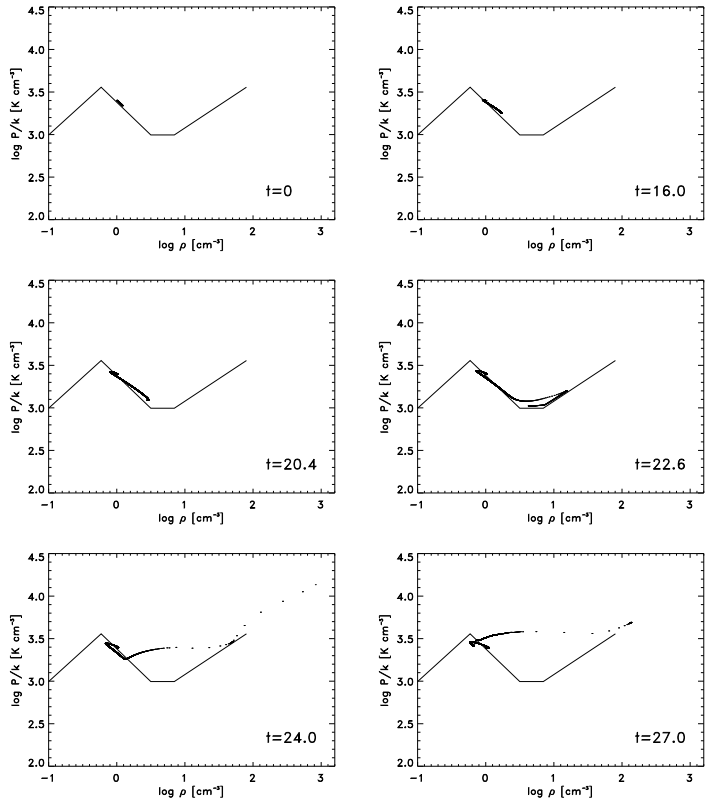


Fig. 6. Time development of run DEN75 in P–ρ phase space. Before the shock formation, the evolution proceeds along the thermal equilibrium curve. Subsequently, the outwards-propagating shock brings the outside medium out of thermal equilibrium, and restores nearly pressure balance. Thus, the infalling gas is traversing the density "unstable" range, but in nearly isobaric (inertial) conditions, rather than along the thermal equilibrium curve.

tact discontinuities) at the cloud boundaries that only subside asymptotically in time. Since the evacuation of the low density regions must proceed in times of order of the sound crossing time, the final fraction of mass in the "unstable" density range, in the more realistic case of multiple fluctuations, should depend on their number. In fact, we have performed simulations with a full spectrum of initial fluctuations, and in those cases the final unstable fraction may be much lower, although still times $\gtrsim 15$ (respectively, 8) Myr are required to evacuate the unstable range when the minimum perturbation size is 12.5 (respectively, 1.25) pc. More importantly, however, small-scale perturbations behave very differently when they are quasi-adiabatic rather than quasi-isobaric, as we discuss in the next section.

4 The Case of Velocity Fluctuations

As mentioned in Sect. 2.2, velocity fluctuations are likely to be the most representative of the actual situation in the turbulent ISM, because in a continuum any density fluctuation must originate from compressive or expansive motions. Such motions are readily available in a compressibly turbulent medium. When the cooling time is long ($\eta \gg 1$), these compressions/rarefactions heat/cool the gas adiabatically, and the perturbations are then isentropic, which, as discussed in Sect. 2.2, are stable in the atomic ISM in the short-wavelength limit. As also mentioned in that section, in the case of velocity fluctuations, the dynamical time in $\eta \equiv \tau_{\rm c}/\tau_{\rm d}$ is given by $\tau_{\rm d} = \min\{\tau_{\rm s}, \tau_{\rm u}\}$, where $\tau_{\rm u}$ is the turbulent crossing time, which is in general also a scale-dependent quantity.

Several studies [56,31,42] have investigated the possibility of triggering TI through the nonlinear compression, either by strong shocks or large-scale, large-amplitude compressive waves of the warm *stable* phase, concluding in general that triggering TI off the stable phase is possible for strong enough compressions, with the possibility of even forming molecular hydrogen in the collapsed region [42]. However, these studies have assumed that the gas has already previously segregated into phases. It is our interest now to discuss the extent to which such segregation can be achieved, starting from unstable conditions. Therefore, in this section we describe the evolution of a thermally unstable medium (with respect to the isobaric criterion) subject to random velocity forcing, as originally presented in Paper I.

We take uniform-density initial conditions, and apply a random forcing f of the form

$$f(x,t) = \mathrm{Re}\left[N \exp\left(ik\left(t\right)x + i\phi\left(t\right)\right)\right], \tag{17}$$

where $k(t)$ is a time dependent wavenumber, $\phi(t)$ is the phase and x the position. Following [6], we take $N = f_0 c_{\rm s} \left(k(t) c_{\rm s}/\delta t\right)^{1/2}$, where f_0 is a constant factor and δt is the length of the timestep. The values of $k/(2\pi/l)$, where l is the box length, and of ϕ are selected at each timestep randomly in the ranges $[3, 10]$ and $[0, 2\pi]$, respectively. The positive exponent $(1/2)$ in N implies that strongest forcing occurs at the highest wavenumbers of the forced range, so that the energy-injection scale scale is $\lambda_{\rm i} = l/10$. We have chosen this forcing for two main reasons. One, it mimics the small-scale stellar forcing acting in the ISM, and, two, it allows us to maintain the desired rms Mach numbers at the (small) scales of interest, since this is difficult to achieve with pure large-scale forcing. We do not consider decaying-turbulence situations, as we are ultimately interested in models of the ISM, which is subject to continued energy-injection processes.

We have performed nearly 20 simulations varying the box size and the scale and amplitude of the forcing (see Paper I for details). From them we conclude that, for the average conditions of the ISM, the presence of turbulent motions with small enough sizes (~ 0.3 pc) and moderate amplitudes ($\mathcal{M}_{\rm rms} \gtrsim 0.3$) such that η is maintained above unity, condensations do not appear. We understand this as a consequence of the turbulent crossing time becoming shorter than the growth time of the condensations, allowing the turbulent fluctuations to both dis-

rupt the incipient condensations and to more than compensate cooling through the heating from shocks and adiabatic compression; i.e., the perturbations become adiabatic, and therefore stable according to the linear analysis in the limit $\eta \gg 1$.

In general, we conjecture that the presence of velocity fluctuations in the ISM, even if unable to completely suppress the development of condensations, may increase the fraction of gas in the "unstable" temperature range. Note also that for small-scale velocity fluctuations, the evolution is *not* along the thermal-equilibrium curve, but rather intermediate between adiabatic and isobaric. Thus, the density range determined by the "unstable" temperature range under these conditions does not exactly coincide with the unstable density range under thermal equilibrium, as given in Sect. 2 and Fig. 2. An interesting possible application for observationally measuring the actual thermodynamic state of the atomic ISM is mentioned in Sect. 7.

5 The Magnetic Pressure in Turbulent Media

We now make a pause in the discussion of the thermal instability and consider the character of the magnetic pressure in turbulent flows, in order to assess the possibility that it may supplement the thermal pressure in the ISM, and thus contribute to weaken the effects of TI. Note that in this section we make no attempt to discuss TI in the presence of a magnetic field. This has been discussed by a number of authors (e.g., [22,61,50,32]). Instead, here we investigate the nature of magnetic pressure in fully turbulent compressible, magnetized isothermal flows. Several works have considered this regime as well, both numerically (see, e.g., the reviews by Mac Low, Ostriker and Nordlund in this volume, and references therein), and theoretically [49]. In particular, the numerical simulations of references [63] and [62] (see also [66] for the nonisothermal case) reported a lack of correlation between the density and the magnetic pressure, B^2, where B is the magnetic field strength, at low and intermediate densities in cases in which the magnetic β parameter[3], equal to the ratio of thermal to magnetic pressure, is ~ 1. Moreover, recent observational data (see, e.g., Crutcher, Heiles & Troland, this volume) suggest a similar lack of correlation at densities below ~ 1000 cm^{-3} in the ISM. Hennebelle & Pérault, in their study of the triggering of TI in the diffuse medium through strong compressions [32], have found that the magnetic pressure decorrelates from the density during this process, even in cases in which no condensation occurs, as long as the angle between the directions of the compression and of the mean field is small enough. This is because in this case, even though the field lines are initially bent by the compression, the magnetic tension eventually redirects the flow along the mean field, and the flow can then proceed unimpeded by the field. More recently, Passot & Vázquez-Semadeni ([67], hereafter Paper II) have attempted to understand the origin of this decorrelation under fully turbulent conditions in terms of the so-called "simple" MHD waves,

[3] The magnetic β should not be confused with the exponent β_{ij} of (1)

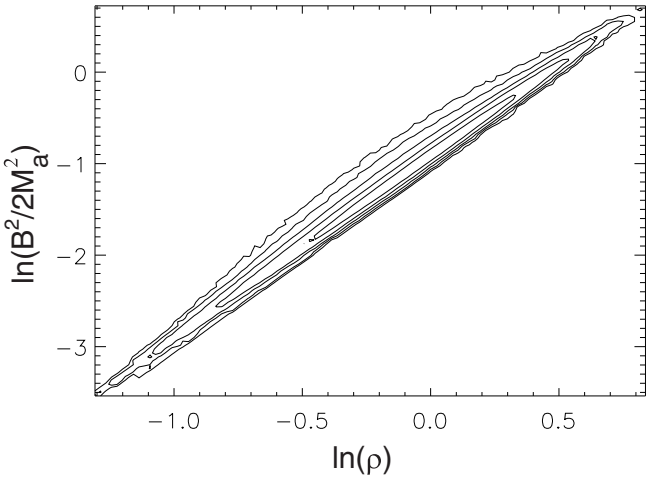


Fig. 7. Magnetic pressure-density correlation, indicated by the two-dimensional histogram of points in log-log coordinates, for a simulation with a magnetic field perpendicular to the direction of propagation (i.e., $\cos\theta = 0$), and forcing parallel to this direction. This configuration allows only the existence of the fast mode of nonlinear MHD waves. The run has an rms field fluctuation $\delta B/B = 0.62$ and $\delta\rho/\rho = 0.62$. The rms Alfvénic Mach number $\tilde{M}_{\rm A} = 5.2$. The magnetic pressure is seen to scale as ρ^2.

and discussed its implications on the role of B^2 as a pressure. In this section we briefly summarize the results of that paper.

We consider isothermal MHD flows in "1+2/3" dimensions, or slab geometry. The direction x is referred to as the direction of wave propagation. In this setup, b_x, the field component along x, is constant. We denote by b the magnitude of the field component perpendicular to x. The initial, uniform magnetic field is chosen to lie in the (x, z) plane, at an angle θ from the x axis, so that $b_x = \cos\theta$ at all times. The treatment in this section is entirely in non-dimensional units, so that the parameters characterizing the flow are the sonic and Alfvénic Mach numbers of the velocity unit, denoted $M_{\rm s}$ and $M_{\rm A}$, respectively, and the propagation angle θ. The plasma beta is then $\beta = M_{\rm A}^2/M_{\rm s}^2$.

"Simple" MHD waves (see, e.g., [46,52]) are finite-amplitude solutions of the equations, characterized by the property that all variables can be expressed as wave profiles, i.e., as a function of a single one of them (say, the density) as in the case of linear MHD waves. The same well-known modes of the linear case, i.e., Alfvén, fast, and slow, exist in the case of simple waves. Only the latter two are associated with the density fluctuations. The propagation velocities of the modes are given by [46,52]

$$v_\pm^2 = \frac{1}{2M_{\rm A}^2\rho}\left(B^2 + \beta\rho\right)\left(1 \pm \sqrt{1 - \frac{4\beta b_x^2\rho}{(B^2+\beta\rho)^2}}\right) \tag{18}$$

and

$$v_A = \pm \frac{b_x}{M_{\rm A}\rho^{1/2}}, \tag{19}$$

where $v_\pm$ denotes the speed of the fast (+) and slow (−) modes, and v_A, that of the Alfvén mode. $B^2 = b_x^2 + b^2$ is the total field strength.

After manipulating the equations to obtain the wave profiles, one finds, in particular, for the dependence of the field with density,

$$\frac{d}{d\rho}\frac{b^2}{2} = \frac{d}{d\rho}\frac{B^2}{2} = (M_{\rm A}^2 v^2 - \beta). \tag{20}$$

In the limit where $4\beta b_x^2\rho \ll (B^2 + \beta\rho)^2$, a limit generally satisfied except when b_x is not too small, for $\beta\rho$ of order unity and small field distortions, this equation can be simplified and integrated using (18) to give the dependences of magnetic pressure with density for the fast and slow modes as

$$b^2 \approx C_1 - C_2\rho \quad \text{(slow mode)} \tag{21}$$

$$B^2 \propto \rho^2 \qquad \text{(fast mode)}, \tag{22}$$

where C_1 and C_2 are constants. The slow-mode behavior can be further determined in some limits. For example, when b_x is small (quasi-perpendicular propagation) or b^2 large, numercially solving (eq:profileb), one finds that there exists a range at small density where $C_2 = 2\beta$, i.e. where the total pressure, $P_{tot} = \frac{b^2}{2M_{\rm A}^2} + \frac{\rho}{M_s^2}$, is rougly constant. At larger density one finds $C_1 = b_x^2$ and $C_2 = \beta$. In any case, the β weighting in C_2 implies that at low β, the slow mode produces large density fluctuations even when the field fluctuations are small. More importantly, the pressures from the two modes depend very differently on density. One can thus expect that, in the large fluctuation amplitude case (i.e., the fully nonlinear regime), *the particular value of the magnetic pressure of a fluid parcel will not be uniquely determined by its density, but instead, that it will depend on the detailed history of how the density fluctuation was arrived at*, causing a lack of correlation between the magnetic pressure and the density.

The latter suggestions have been tested in Paper II by means of numerical simulations with random forcing (actually, an acceleration) applied on wavenumbers 1-19 to all three velocity components or to only the perpendicular ones. Choosing the direction of the forcing and of the uniform magnetic field allows us to highlight either one of the slow and fast modes. Figure 7 shows the b^2-ρ correlation by means of isocontours of the two-dimensional histogram in the $\log(b^2/2M_{\rm A}^2)$–$\log(\rho)$ plane of the points in a simulation with 4096 grid points, forcing applied on all three velocity components, and the initial magnetic field perpendicular to the direction of propagation. This is a case in which only the fast mode exists, consisting of a pure magnetosonic wave, and the correlation exhibits the well known ρ^2 behavior of magnetic pressure in this case. This result holds independently of the value of $M_{\rm A}$.

Figure 8, on the other hand, shows the correlation for a run at the same resolution with the forcing perpendicular to the propagation direction, and the

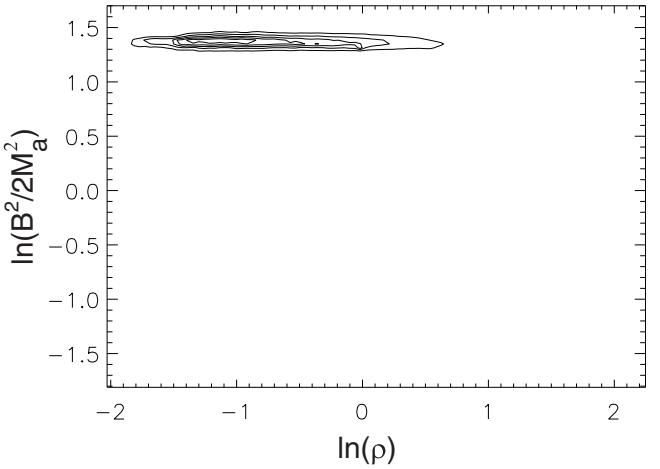


Fig. 8. Magnetic pressure-density correlation for a simulation with $\cos\theta = 0.1$, and forcing perpendicular to the propagation direction, with $\delta B/B = 0.32$ and $\tilde{M}_{\rm A} = 0.48$. This configuration highlights the slow mode. The magnetic pressure is seen to remain constant.

magnetic field almost parallel to the forcing ($b_x = \cos\theta = 0.1$), at low Alfvénic Mach number ($M_{\rm A} = 0.15$). In this case, the density fluctuation production is dominated by the slow mode because $\tilde{\beta} \sim 0.007$ (c.f. (21)). The near constancy of the total pressure in this case is evident in this figure. In this simulation, one observes large oscillating density clumps which do not merge nor pass each other, mostly anti-correlated with b^2. However, for this same field configuration, as $M_{\rm A}$ is increased, the field becomes more strongly distorted, and the fast mode starts acting on the perturbed field. The fluctuations become more random and superpose each other, with the presence of fast shocks and a correlation between ρ and b^2. The result is that at high $M_{\rm A}$ both modes are actively producing density fluctuations, and the correlation between magnetic pressure and the density is lost, as seen in Fig. 9. The situation can be idealized by assuming that the density of each fluid parcel is arrived at through a random sequence of slow and fast waves, which is different for each parcel.

The case of parallel propagation is also interesting to mention as it illustrates the complexity of this problem. For example, the picture of non-interacting clumps observed at small $M_{\rm A}$ (i.e. in presence of weak field distortions) and for large angles is also observed for parallel propagation in the case of large $M_{\rm A}$. But in that case the clumps form inside strong slow shocks. The magnetic field intensity is the weakest inside the clumps which cannot merge due to the high external magnetic pressure. In the general 3D case, we expect all angles between the magnetic field and the propagation direction to be present, and therefore a representative case would be one in which the field is at 45° from the propagation direction. In this case, we recover the trend of a roughly constant magnetic pressure at small $M_{\rm A}$ and an increased scatter between the magnetic pressure

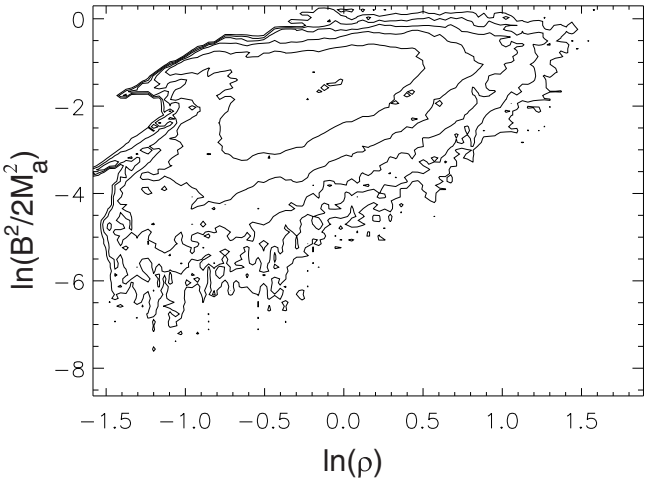


Fig. 9. Magnetic pressure-density correlation for a simulation with $\cos\theta = 0.1$, and forcing in the three directions. This configuration also highlights the slow mode, but increases the field distortions ($\delta B/B = 2.17$ and $\tilde{M}_{\rm A} = 7.29$). The magnetic pressure is seen to decorrelate from the density, although it appears to be bounded above and below by the slow and fast mode dependences, respectively.

and density as $M_{\rm A}$ is increased (not shown). It is found that the level of density fluctuations greatly depends on the dominant mode, has no specific relation with β and is actually the largest when B is strong (and thus slightly perturbed). The density probability density function (PDF) is close to a log-normal when B^2 and ρ are not correlated, and shows an excess at small density when slow waves dominate.

We conclude from this section that the scatter between magnetic pressure and density found in simulations and in observational data can be understood in terms of the different dependence of the magnetic intensity on density for the slow and fast modes of simple nonlinear MHD waves, and of the random sequence of these that a fluid parcel experiences as it evolves in a fully turbulent regime. Moreover, a number of important implications emerge from this lack of correlation. First, it suggests that modeling magnetic pressure by means of a polytropic dependence on density may not be adequate in the fully turbulent case. This relation is seen to apply when scale separation is preserved between small-scale Alfvén waves and large-scale density perturbations. Second, the magnetic "pressure" does not really act as a pressure, as it does not behave as a restoring force in general. Instead, it acts more as a random forcing. Third, the latter point suggests that magnetic pressure should not be effective as a substitute for thermal pressure when the latter behaves "in reverse" (see Sect. 2.2) in thermally unstable situations. A final note is that the isothermality of the flows considered here is of no relevance to the results, which are thus expected to apply equally to non-isothermal flows.

6 TI in Models of the ISM

In the previous sections we have discussed the nonlinear development of fluctuations of various kinds and sizes in the presence of TI under the isobaric criterion, and the nature of magnetic pressure in turbulent media. With this background we can now proceed to discuss the behavior of numerical models of the ISM incorporating the magnetic field, self-gravity, rotation, shear, and stellar-like (localized) energy injection, in the presence of isobaric TI.

As mentioned in the introduction, the classic two- and three-phase models of the ISM were based on the principles of thermal and pressure equilibrium, and thus did not predict the existence of significant amounts of gas in-between the phases. Instead, the time-dependent model [27] did. Furthermore, observations do not clearly support a sharp phase segregation either; instead, they often have found evidence of significant amounts of "lukewarm" gas at temperatures intermediate between those of the cold and warm phases [15,36,77,24,30]. Thus, it is important to perform numerical simulations that quantify this fraction and that allow us to determine whether sharp phase segregation is expected in the ISM, or whether it is more likely a continuum.

Numerical simulations of the ISM including radiative heating and cooling as well as stellar-like energy injection have been performed by a number of authors, starting from the pioneering work of Bania & Lyon [4], and continuing with references [10,11,69,78,66,20,28,25,41,14,80,26,39,51]. These works have included different amounts of physics in the simulations, such as the magnetic field, self-gravity, galactic disk rotation, etc. However, the role of TI had been discussed only in passing until recently. Bania & Lyon, using two-dimensional (2D) simulations of a 180-pc square region on the Galactic plane, at a resolution of 40×40 pixels, including randomly-positioned stellar-like energy sources, pointed out that the inclusion or not of a thermally unstable range had little effect on the resulting structure. Elmegreen [20] pointed out that high-resolution 1D MHD simulations forced with nonlinear magnetic waves can form hierarchical cloud/intercloud structure both if the cooling functions used have a single or double stable equilibria. The 2D non-magnetic simulations of Gerritsen & Icke [28], and the 3D MHD ones of Korpi et al. [41], although not specifically aimed at this issue, already pointed towards the existence of unstable gas. Vázquez-Semadeni, Gazol & Scalo [80] (hereafter Paper III) first discussed the interaction of TI with the turbulent motions in the ISM, suggesting that at large scales (simulation size of 1 kpc) the signature of TI in the mass density histogram is erased when small-scale forcing mimicking expanding HII regions placed at the density peaks is used in low-resolution (128×128) 2D simulations including the magnetic field, self-gravity and rotation. Gazol et al. [26] (hereafter Paper IV) then reported that the temperature histogram in similar simulations at higher resolution (512×512) contains roughly half the mass at unstable temperatures. Recently, Mac Low et al. [51] (see also Mac Low, this volume) have presented 3D MHD simulations spanning the whole range of temperatures existing in the ISM, including the hot ($T \gtrsim 10^6$ K) gas, showing that the effect of supernovae is to introduce large fluctuations (by 2–3 orders of magnitude) in the thermal

pressure of the ISM, in contrast with the multiphase models of the ISM [23,53], and more in agreement with the time-dependent ones [27], and with the recent observational study of Jenkins & Tripp [35]. Finally, in an analytical treatment of the effective equation of state of the ISM and the power spectrum of the energy sources, Norman & Ferrara [60] concluded that the traditional multiphase description should be replaced by a "continuum of phases".

A related line of study has been that taken by Koyama & Inutsuka [43] and Kritsuk & Norman [45]. Both of these groups have recently considered the generation of turbulence in flows in more than one dimension, as a consequence of the nonlinear development of TI. Koyama & Inutsuka considered a shock-compressed layer in 2D between a fast flow with diffuse-gas properties ($n = 0.6$ cm^{-3}, $T = 6000$ K) and a hot gas region, showing that the layer fragments into small cloudlets that have supersonic velocity dispersions with respect to the warm medium in which they are embedded, and coalesce to form larger units. Kritsuk & Norman considered the 3D development of the instability alone, very far from thermal equilibrium ($n = 1$ cm^{-3}, $T = 2 \times 10^6$ K), so that the gas is initially unstable under the isochoric mode and the cooling times are very short (~ 0.3 Myr). They again found that the development of TI generates turbulence, in which roughly 15% of the mass is in the unstable regime. They point out, however, that this turbulence is decaying, because it has no other energy sources than the development of TI itself.

However, none of the works reporting a fraction of unstable gas have discussed in detail the possible suppression of the instability by numerical limitations, raising a concern that perhaps the presence of unstable gas, and therefore the lack of sharp phase transitions in the simulations, are numerical artifacts. In the remainder of this section we first present new results concerning the numerical issues in detail, in order to assess the validity of the ISM simulations and interpret their results. We then present new simulations that, based on the numerical considerations, provide reasonable evidence that significant amounts of gas at unstable temperatures should be expected in the atomic ISM.

The simulations presented in this section have been performed using a pseudo-spectral scheme, described in detail in [79], to solve the MHD equations in the presence of heating, cooling, stellar-like energy sources, and self-gravity, using a hyperviscosity scheme and including a mass diffusion term with coefficient μ_m (cf. (13)) and a thermal diffusivity term with a constant coefficient $K = K_0$ (cf. (4)). In non-dimensional form, the equations read

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{u}) = \mu_m \nabla^2 \rho, \tag{23}$$

$$\frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \cdot \nabla \mathbf{u} = -\frac{\nabla P}{\rho} - \left(\frac{J}{M}\right)^2 \nabla \phi + \frac{1}{\rho}\big(\nabla \times \mathbf{B}\big) \times \mathbf{B} - 2\Omega \times \mathbf{u} - \nu_8 \nabla^8 \mathbf{u} + \nu_2 (\nabla^2 \mathbf{u} + \frac{1}{3} \nabla \nabla \cdot \mathbf{u}), \tag{24}$$

$$\frac{\partial e}{\partial t} + \mathbf{u} \cdot \nabla e = -(\gamma - 1) e \nabla \cdot \mathbf{u} + \frac{K}{c_V} \frac{\nabla^2 e}{\rho} + \Gamma_{\rm d} + \Gamma_{\rm s} - \rho \Lambda, \tag{25}$$

$$\frac{\partial \mathbf{B}}{\partial t} = \nabla \times (\mathbf{u} \times \mathbf{B}) - \nu_8 \nabla^8 \mathbf{B} + \eta_m \nabla^2 \mathbf{B}, \tag{26}$$

$$\nabla^2 \phi = \rho - 1, \tag{27}$$

$$P = (\gamma - 1)\rho e, \tag{28}$$

$$\Gamma_{\rm d}(\mathbf{x}, t) = \Gamma_0, \tag{29}$$

$$\Gamma_{\rm s}(\mathbf{x}, t) = \begin{cases} \Gamma_1 \text{ if } \rho(\mathbf{x}, t_0) > \rho_{\rm cr} \\ \quad \text{and } 0 < t - t_0 < \Delta t_s\,, \\ 0 \quad \text{otherwise,} \end{cases} \tag{30}$$

where M is the Mach number of the velocity unit, taken equal to unity, $J = l/L_{\rm J}$ is the box size in units of the Jeans length, ϕ is the gravitational potential, Ω is the Galactic disk rotation rate, ν_2 and ν_8 are respectively the second-order and eight-order (hyperviscosity) coefficients, η_m is the magnetic diffusivity, e is the specific internal energy, K is the thermal conductivity, $\Gamma_{\rm d}$ and $\Gamma_{\rm s}$ are respectively the diffuse background heating rate and the local stellar heating rate, and Γ_0 and Γ_1 are constants. The equations are non-dimensionalized to the box size L_0, the velocity unit $u_0 = c$, the temperature unit $T_0 = 10^4$K and the magnetic field unit $B_0 = 5\mu$G.

The cooling rate Λ is still given by (1), although the nondimensional coefficients are now given in Table 2, together with the non-dimensional heating rates, for three different box sizes, at the same temperature unit.

Table 2. Cooling function parameters for ISM simulations

Box size (pc)	C_{12}	C_{23}	C_{34}	C_{45}	Γ_0	Γ_1
10	56.2	0.462	0.102	0.474	4.57×10^{-2}	20
150	845.	6.95	1.54	7.14	0.688	38.7
1000	5.62×10^3	46.2	10.2	47.4	4.57	250

The ISM simulations are started with uncorrelated gaussian random fluctuations in all variables and amplitudes of order unity, with characteristic scale $\sim 1/8$ of the box size. The initial magnetic field has a uniform component of 1.5 μG on the x-direction, and an rms fluctuation amplitude of 4.5 μG. The energy injection mechanism consists of small-scale ($\sim$ 10 pixels across) heat sources turned on at sites where the density exceeds a certain threshold (chosen within the cold stable branch of the density range). These sources remain on for $\Delta t = 6$ Myr. This time lapse Δt was shortened by a factor 10 and the energy injection rate increased by roughly the same factor in the simulations with box size = 10

pc, because there, the time unit is too short, and a star would remain on for more than half the duration of the simulation. These sources are intended to mimic the effect of ionization heating from OB stars in HII regions. Supernova-like sources are not included because of limitations of the numerical scheme to handle very strong shocks.

6.1 Numerical Considerations

The numerical simulation of thermally unstable turbulent flows presents a significant numerical challenge. It requires solving simultaneously the regions of sharp gradients occurring in the immediate neighborhood of clouds and the quasi-linear development of perturbations in the more distant, relatively quiescent, unstable medium mediating the clouds and the warm, stable, diffuse phase. The question is whether this medium fragments, to finally end with a state in which the unstable gas virtually disappears from the simulation. The simultaneous solution of both regimes is important, because stellar energy injection recycles gas from the dense phase into the warm phase.

The simulation of regions with strong shocks requires the use of artificial viscosities and diffusivities in order to spread out ("capture") shocks over a few grid points. Unfortunately, such diffusivities also have the effect of damping the growth of perturbations in the relatively smooth regions, because they artificially increase the Field length (Sect. 2.1), reducing the range of unstable scales, as well as their growth rates, in the unstable gas. This problem is also present when finite-difference schemes, which produce numerical diffusivity, are used. This was not a problem, however, in sections 3 and 4 because of the very large resolutions used there, at the expense of using a 1D approach.

The Field length $\lambda_{\rm F}$ has the property of being much larger than the diffusive scale (understood as the molecular mean free path) when the latter is very small, since it can be easily shown [22] that the wavenumber associated to the Field length satisfies

$$k_{\rm F} = \left[\frac{\mu\ (\gamma - 1)}{Rc} k_K |N_P|\right]^{1/2}, \tag{31}$$

where k_K, given by (7), is the wavenumber associated with the molecular mean free path. Thus, $k_{\rm F}$ only grows as the square root of k_K. In a numerical simulation, the scale associated with the artificial thermal diffusivity $\lambda_K = 2\pi/k_K$ (cf. Sect. 2.1) plays the role of the molecular mean free path, and therefore the growth of perturbations with wavelengths $\lambda_K < \lambda_{\rm F}$ can be artificially suppressed if the numerical $\lambda_{\rm F}$ is much larger than the real $\lambda_{\rm F}$ in the atomic ISM. In other words, some of the resolution in a simulation can be "wasted", in the sense that intermediate-wavelength perturbations can be damped even if the diffusive scale is comparable to the smallest resolved scale, as we will show below.

On the other hand, the pseudo-spectral numerical scheme we use in simulations of the ISM has the advantage that it produces no numerical diffusivities at all (the spatial derivatives are calculated exactly in Fourier space, rather than through finite-difference approximations). Thus, all artificial diffusivities

are included explicitely in the equations, and can be controlled through their associated coefficients. This allows us to perform simulations of the non-diffusive case in the linear and weakly nonlinear regimes (recall the diffusivities are only needed to smooth out strong gradients) and of diffusive cases, so that we can both test the code against the predictions of the linear analysis, and then measure the effect of the diffusivities precisely. We have found that actually the mass diffusion has a much stronger damping effect on the growth of perturbations than heat conduction and viscosity. The value of the thermal conductivity used in the ISM simulations discussed below, $K = 7.3 \times 10^{-3}$, was found to not affect the growth rate by more than 10% for the wavenumbers tested below. A similar situation holds for the viscous and hyperviscous coefficients. Note that these coeffcients are chosen as the smallest ones allowing correct treatment of near discontinuities by the numerical code.

We have therefore performed many simulations at low resolution (128×128) to measure the growth rates of pure sinusoidal isobaric perturbations of various wavelengths (from 1 to $1/32\times$ the box size) as a function of the mass diffusion coefficient, in order to investigate at what point the perturbation growth is suppressed and thus full ISM simulations using those values cannot be trusted anymore concerning the instability of regions of sizes comparable to those perturbations. Note that higher resolution is not necessary for this purpose, as all that is needed is to resolve the perturbations themselves and their initial growth. Higher resolution is only necessary in the fully nonlinear case, to resolve shocks while still having a large range of scales between the simulation size and the scale of shock-spreading.

Together with the theoretical growth rate, Fig. 1 shows the growth rates (defined as the inverse of the e-folding time for linear – 2.5% amplitude – perturbations) measured in simulations of a region of 150 pc on a side as a function of wavenumber and of the mass diffusion coefficient. These rates can be compared with the solid line, which gives the solution of the dispersion relation (5) as a function of wavenumber. A good agreement is seen between the theoretical and numerical rates for the zero-diffusivity case. For non-zero diffusivity, the numerical rates are seen to decrease significantly, by roughly a factor of 1.5 at $\mu_m = 4 \times 10^{-3}$ and $\lambda = 1/8$ of the box size. For the same diffusivity, perturbations of size 1/16 of the box size have a zero growth rate (the perturbations remain static, without growing or dispersing).

This damping effect is alleviated somewhat by noting that larger-amplitude perturbations have larger growth rates [74]. We have thus also computed the numerical growth rates for perturbations of initial amplitude of 25% in simulations of box size = 150 pc (not shown), verifying the occurrence of larger growth rates in this case. This is advantageous because then weakly nonlinear perturbations of scales down to sizes $\sim$ 15 pc should grow in simulations with box sizes 150 pc at rates comparable to those of linear perturbations without any diffusivity. Nevertheless, we have found that zero growth still occurs at the same scale (1/16 of the box) as for the linear perturbations at roughly the same value of μ_m. Any scales below this are stabilized by the mass diffusion, and thus may be

wrongly interpreted as stable in a 150-pc ISM simulation using this value of the diffusivity.

Figure 10 shows the numerical growth rates for 2.5% perturbations on simulations of box sizes 10 pc and 1 kpc on top of the theoretical growth rate curve. Remarkably, it is seen that the mass diffusion is much more effective in damping the growth of perturbations for a small box size than for a large one. Specifically, a diffusivity of $\mu_m = 0.001$ is enough to completely damp perturbations of size 1/8 of the box when the box size is 10 pc, while a value $\mu_m = 0.008$ still allows growth of perturbations of size 1/16 of the box for a box size of 1 kpc. This is actually easy to understand because a larger box size is represented, in non-dimensional units, by larger thermal coefficients (see Table 2). Instead the diffusive coefficients are independent of the physical box size, if the diffusive scale is kept at a given fraction of the numerical box (in pixels). This is similar to the effect on the Field length of the thermal conductivity (31). Somewhat surprisingly, we conclude that, at a fixed value of the mass diffusion coefficient, *a simulation with a larger physical box size is more accurate than one with a small physical size because it contains a larger range uf unstable scales.*

A final comment is that, from the results of this section, it appears that the numerical scheme best suited for solving the problem at hand (i.e., capturing strong gradients occurring in dense clouds while not disturbing the growth of perturbations in the mildly turbulent diffuse medium surrounding the dense clouds), is a pseudo-spectral scheme with a position-dependent value of the diffusivities, as in (16), so that strictly zero diffusivity can be used in mild regions, while still smoothing out the strong gradients in the clouds can be achieved. We will attempt such an approach in forthcoming papers. However, our presently available tools for modeling the full ISM problem only include constant-coefficient diffusivities, and we will discuss results using them in the next subsection. Note that not even adaptive-mesh refinement schemes may be better suited for this problem, because the small-amplitude perturbations need to be followed in the diffuse medium in order to determine whether they grow spontaneously, and this would require high refinement levels in relatively large volumes.

6.2 Results

In the light of the results of Sect. 6.1, we can now proceed to present some numerical simulations of the ISM and assess their reliability regarding the mass fraction in the unstable range.

In Paper IV, we presented a 2D simulation of the ISM with a 1-kpc box at a resolution of 512^2 grid points with a mass diffusion coefficient $\mu_m = 0.0075$. From Fig. 10 it can be seen that perturbations of sizes down to 1/16 of the box can grow at this value of μ_m with a physical box size of 1 kpc, while perturbations of size 1/32 of the box are damped. Thus, although this box size is the one that provides the largest range of unstable scales, it nevertheless does not reach down to the fastest-growing scales. Thus, we have performed two more simulations, to cover the entire range of scales of interest: the first one with a box size of 150 pc, and a very-high resolution of 1536^2 grid points, in order to be able to use

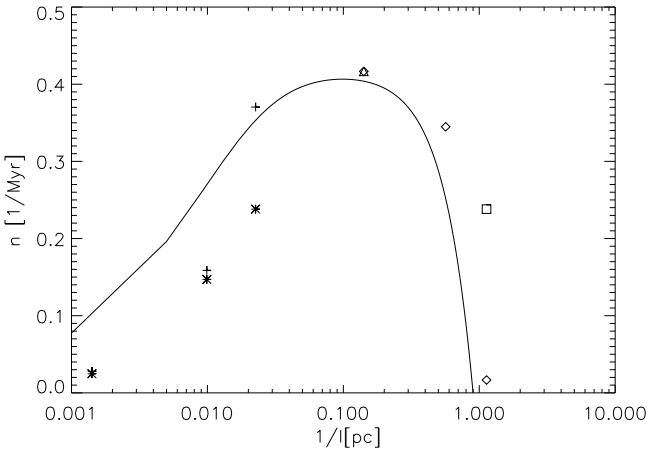


Fig. 10. Numerical growth rates for simulations of the linear growth of 2.5% isobaric density perturbations in boxes of sizes 10 and 1000 pc and various values of the mass diffusion coefficient, superposed on the growth rate curve. *Crosses:* 1000 pc, $\mu_m = 0$. *Asterisks:* Box size=1000 pc, $\mu_m = 0.008$. *Triangles:* Box size=10 pc, $\mu_m = 0.0$. *Diamonds:* Box size=10 pc, $\mu_m = 0.001$. *Squares:* Box size=10 pc, $\mu_m = 0.0005$.

a mass diffusion coefficient $\mu_m = 0.003$ which, according to the data in Fig. 1, allows growth again of perturbations of size 1/16 of the box, thus barely reaching the scales of fastest growth ($\lesssim 10$ pc). The second simulation uses a box of 10 pc, and $\mu_m = 5 \times 10^{-4}$ at a resolution of 512^2, thus allowing the growth of perturbations down to scales 1/8 of the box size. Note that for this simulation, the unstable scale range is actually larger than that allowed by the realistic value of thermal conductivity mentioned in Sect. 2.1. We have opted for performing 2D simulations in order to maintain relatively high resolutions, while still being able to capture the complex vector interactions of the system.

In Fig. 11 we show the density and temperature histograms of these three runs after a stationary regime has been attained. It can be seen that the density histograms do show signatures of TI, such as slope changes and slight peaks at the stable densities, but nevertheless, a sizeable fraction of the gas is in the unstable regime. A similar result was reported by Kritsuk & Norman [45].

Concerning the temperature histograms, we see that, in fact, as the physical box size decreases, the temperature histogram has less pronounced spikes at the temperatures corresponding to the stable phases, suggesting that phase segregation is less pronounced as well. In the case of the 10-pc run, this can be an artifact of the reduced unstable range due to the mass diffusion. However, this is not so for the 1-kpc and 150-pc runs, as both have roughly the same range of unstable scales, and in fact with larger growth rates in the case of the latter, because the corresponding physical scales are smaller. Thus, these two runs suggest that in fact the result is real, due to the decreasing value of η as smaller scales are considered, because in this case the perturbations are adiabatic-like and are stable to first order. Taking the simulations at face value, the mass fractions in

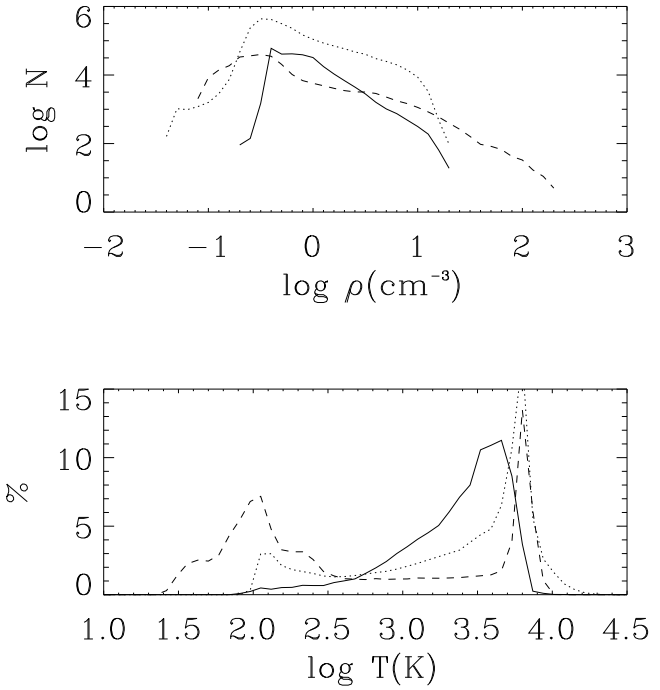


Fig. 11. *Top:* Density histograms for the ISM simulations with box sizes 10 pc (*solid line*), 150 pc (*dotted line*) and 1000 pc (*dashed line*). A clear trend toward higher phase segregation (more strongly bimodal shape) is seen at *large* physical box sizes because the wave mode is unstable at large scales only. *Bottom:* Temperature histograms, with the same labeling.

the "unstable" temperature range are 27% for the 1-kpc run, 58% for the 150-pc run, and 92% for the 10 pc run. Again, we see the trend of a *larger* fraction of unstable gas as smaller scales are considered.

On the other hand, the effect of strong compressions continues to promote the instability. Figures 12 and 13 respectively show the density and pressure fields for the 150-pc run at a typical time. There it can be appreciated that the density maxima (and the filaments connecting them) correspond to pressure *minima*, indicating that these regions are unstable, even when they are much smaller than the smallest linearly unstable scale allowed by the mass diffusion. This is because the strong turbulent compressions locally increase the density and the cooling rate, decreasing η. Thus, *these regions belong to the long-wavelength (small-η) regime in spite of having small physical sizes*, and the pressure behaves closer to the thermal-equilibrium curve. These small-scale, small-η regions are analogous to the compression-induced instability of Hennebelle & Pérault [31], except that in this case they have been pushed from the large- to the small-η regime by the compression, rather than from the stable to the unstable regime.

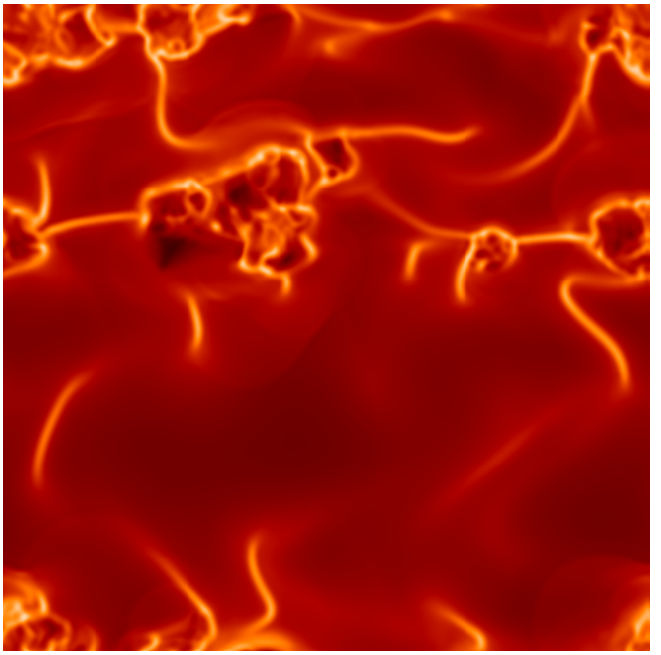


Fig. 12. Density field of the 150-pc simulation of the ISM at the same time for which the density and temperature histograms are shown in Fig. 11. The resolution is 1536^2 grid points.

They also correspond to the transition from isobaric to isochoric cooling as the density increases described by Burkert & Lin [8] and Kritsuk & Norman [45].

It is worth comparing here the present results to those of Hennebelle & Pérault in somewhat greater detail, as those authors did not find significant amounts of unstable gas in their simulations after the condensation process ended. This is because their setup did not consider a globally turbulent medium, but only the effect of a single compressive wave of intermediate-strength (Mach number ~ 2) on the already-segregated warm stable phase. Instead, here we are considering the case of a medium with mean density $n \sim 1$ cm^{-3}, which is close to the ISM average density [21] and lies in the unstable range, so that, even if the two phases are segregated, the mean density remains at that value. Moreover, rather than the effect of a single large-scale compressive mode, we are considering a globally turbulent medium in which the source of energy is stellar-like, having the effect of recycling matter from the dense phase into the diffuse one. As shown in the simulations, a situation similar to that of Hennebelle & Pérault applies at the sites where the compressions are strong (near stellar sources), except that in their case the triggering is in absolutely linearly stable gas, while in our case the triggering is in gas stable only to adiabatic perturbations. On the other hand, at more remote sites, where the turbulence is weaker [1], the fluctuations remain stable in general.

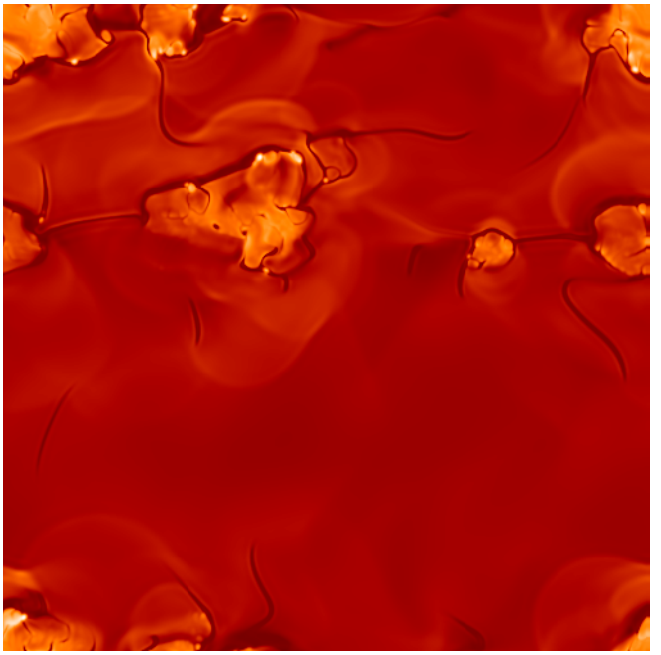


Fig. 13. Thermal pressure field of the 150-pc ISM simulation. The highest pressures correspond to regions of "star formation" clustering, in which the heating from many stars combines additively. Note the high-pressure regions around the clouds, left by the weak shocks propagating into the intercloud medium.

It should be emphasized that, in the context of full ISM simulations, the regions with a reversed pressure gradient are seen to occupy a very small fraction of the volume, with the majority of the space being occupied by a moderately turbulent medium in a nearly isobaric regime, but with significant fluctuations around it. Particularly noteworthy is the existence of weak expanding shock waves which propagate away from the filamentary clouds, behind which *both* the density and the pressure are slightly increased with respect to the intercloud medium, similarly to the case of the wave observed in run DEN75 in Sect. 3. It should be noted, however, that the origin of the outgoing shock waves is not exactly the same in the ISM simulations as it was in run DEN75. In the latter, the shock originates when the condensation reaches its maximum (overshooting) density and terminates the converging motions within it. In the ISM simulations, it originates when star formation suddenly heats and repressurizes the clouds. Nevertheless, the effect on the gas surrounding the clouds is similar. The volume between the wave front and the central clouds contains gas with a "regular" pressure gradient (Sect. 4) and temperatures in the "unstable" range. However, as in the case of run DEN75, this gas is not truly unstable, as it is out of thermal equilibrium, albeit close to pressure equilibrium. The existence of gas with a regular pressure gradient is seen in Fig. 14, which shows the thermal

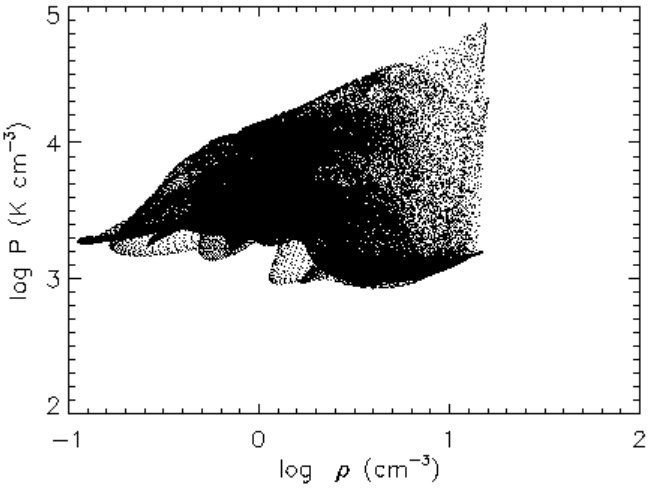


Fig. 14. Pressure vs. density for a vertical stripe of width 200 pixels ($\sim$ 20 pc) in the 150-pc simulation, passing through the large complex slightly up and to the left of the simulation center (cf. Fig. 12).

pressure vs. the density for a vertical stripe of width 200 pixels ($\sim$ 20 pc) and length equal to the box size in the 150-pc simulation including most of the large complex slightly up and to the left of the simulation center. It can be seen that the pressure is far from having a unique value, and, in particular, has a scatter of about one order of magnitude in the "unstable" range, with the upper envelope of the points having a regular dependence of pressure on density. This effect is even more pronounced in the simulations of Mac Low et al. [51], which include supernova energy input (see also the chapter by Mac Low, this volume).

In order to complete the description of the dynamics of the low-thermal-pressure (low-P_T) regions, it is important to also analyze the magnetic pressure within them. This is shown in Fig. 15. Interestingly, some of the low-P_T filaments are also seen as low-magnetic-pressure filaments, while some others are not. However, very few examples of a filament seen also as a high-magnetic-pressure structure are seen, eventhough the magnetic pressure maxima are indeed seen to occur in the star-forming regions. We thus conclude that in general the collapse of the low-P_T regions is ensured. This is confirmed by the fact that those filaments systematically become sites of new star formation at later times as the density threshold is reached.

Also of interest is that the enhanced-density, enhanced-P_T regions surrounding the clouds in general also have an enhanced magnetic pressure, so that these regions are in general slightly over-pressured with respect to the global intercloud medium, and expand somewhat. The intercloud medium is completely permeated by these traveling fronts originating from the clouds. This is most clearly seen in animations, which can be retrieved from our web site `http://www.astrosmo.unam.mx/~e.vazquez/turbulence_HP`. Finally, the magnetic field is rather uniform in the intercloud medium. More discussion con-

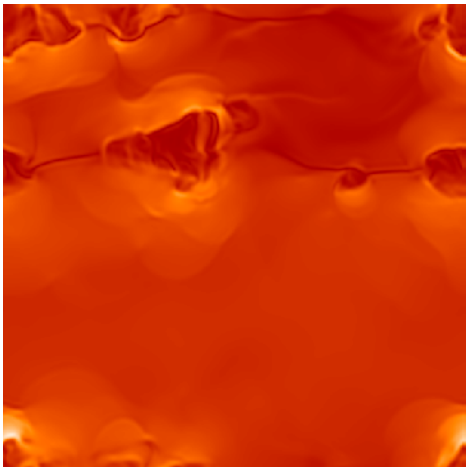


Fig. 15. Magnetic pressure field of the 150-pc ISM simulation. Note that some filaments of low thermal pressure (Fig. 13) also have low magnetic pressure, while some others have undisturbed magnetic pressure.

cerning the magnetic pressure in ISM simulations can be found in the chapter by Mac Low.

6.3 Discussion and Caveats

The results from the previous section suggest a dynamic ISM in which several physical processes are occurring simultaneously besides the pure condensation of thermally-unstable regions, and in which, in fact, the very nonlinear development of the latter leads to complex dynamics.

These results are not free of caveats, however. Three main limitations of the simulations prevent the results from being definitive. First, the need to use artificial diffusivities, especially in the continuity equation, to broaden discontinuities out to a few grid points, causes the Field length to be larger than its real value in the simulations with larger box sizes (1000 and 150 pc). This implies that the fraction of unstable gas could possibly be overestimated because scales that could fragment in the real ISM do not do so in the simulations. However, "unstable" structures larger than those damped by the mass diffusion are seen to exist in the simulations as well, suggesting that the effect is real, if perhaps not as strong.

Here it is important to emphasize that at *large scales* the medium *is* indeed unstable, but this is manifested not in the spontaneous condensation of the structures, but in their null resistance to compression, since typically the turbulence velocity dispersion is supersonic, and the structures are compressed by the turbulence before they can spontaneously condense. At small scales, on the other hand, as indicated by the runs in Sect. 4, a moderate amount of turbulence prevents the spontaneous condensation because the fluctuations are closer to being adiabatic than isobaric.

A second caveat is that the threshold density for star formation (SF) used in the simulations is rather low ($n_{\rm cr} = 15$ cm^{-3} for the 1-kpc and the 150-pc runs, and $n_{\rm cr} = 25$ cm^{-3} for the 10-pc run), again in order to avoid extreme gradients that would form upon the onset of stellar heating in very small, dense clouds. This implies that the clouds are not allowed to reach the pressure-equilibrium density in the dense phase. The SF scheme thus bypasses the process of a cloud (or cloud complex) becoming self-gravitating to start forming stars. This presents the risk that perhaps not enough accumulation of mass is allowed in the clouds before they engage in SF, forcing a greater fraction of the gas to remain in transit in the unstable range towards the clouds. We feel, however, that the essence of the process is the fact that matter is recycled continually among the phases and, given the time scales of the problem, that the existence of sizeable amounts of gas always traversing the unstable range is inevitable to some extent. More accurate estimates of this fraction should be possible upon the application of modifications to the numerical method, as decribed above.

The third caveat is the omission of supernova- (SN-)like energy imput, which should sweep a larger fraction of the volume and induce the formation of a hot phase, which we have neglected. Their inclusion would cause that a larger fraction of the volume would be swept up into shells, forcing the conversion of diffuse gas into the dense phase. Nevertheless, the turbulent mixing ought to be even stronger in this case, and relatively quiescent regions must still exist, as observations suggest. These would undergo slow condensation in the manner outlined here.

In any case, the results presented here should only be considered as suggestive of the existence of sizeable amounts of unstable gas in the ISM, and more specifically designed numerical methods should be applied in order to obtain more definitive confirmation.

7 Summary and Conclusions

In this paper we have reviewed results from various studies aiming at understanding the role of TI in the turbulent atomic ISM, and the behavior of the magnetic pressure in the fully turbulent case. The motivation has been twofold. On the one hand, the classic multi-phase models of the ISM have neglected the implications of the ISM being turbulent, and it is thus important to assess the consequences of advection on the thermal and spatial structure of this medium. On the other hand, observations have often suggested the presence of gas with temperatures in the thermally unstable range, in apparent contradiction with the multi-phase models.

We first reviewed the classic instability analysis of Field [22], emphasizing the different behavior of long- and short-wavelength perturbations (for which the ratio η of the cooling [$\tau_{\rm c}$] to the sound crossing [$\tau_{\rm s}$] time is respectively small and large), and of entropy and isentropic perturbations (which trigger the condensation and the wave modes, respectively). We pointed out that, while much study has been devoted to isobaric entropy perturbations, real-world fluctuations

in the ISM are produced through velocity fluctuations which, in the small-scale limit, belong to the isentropic kind, and are therefore stable to first order at small scales. We also briefly reviewed the magnetic case, in which the presence of a uniform magnetic field can stabilize perturbations with wavenumbers perpendicular to it.

We then reviewed results on the nonlinear stages of evolution of isobaric entropy perturbations, focusing on those that have quantified the magnitude of the speeds developed and the times required for completing the condensation process as a function of the parameter η. For presently-accepted values of the heating and cooling rates [82], large-scale initial perturbations ($\gtrsim 15$ pc, $\eta \lesssim 0.2$) develop supersonic speeds, require times $\gtrsim 10$ Myr to complete the condensation process, and end up with densities and pressures above the thermal equilibrium value due to the ram pressure of the still infalling gas. Those times are long compared with typical times between successive external shock passages, and star formation time scales. We thus concluded that clouds formed from perturbations of such sizes (although the resulting cloud has a size ~ 1 pc) are unlikely to exist in thermal pressure equilibrium with their surroundings. Initial perturbations of sizes $\lesssim 3$ pc, on the other hand, require times ~ 4 Myr to complete their evolution and do not generate supersonic speeds, thus reaching a more quiescent final state, and adhering better to the paradigm of thermal-pressure bounded clouds at the end of their evolution, although, by the time the cloud has formed, accretion is still occurring, so that the clouds are bounded by weak accretion fronts, rather than contact discontinuities. Furthermore, the gas still accreting is necessarily in the unstable temperature range, although it is not in thermal equilibrium; instead, it has a "regular" pressure behavior (P increases with ρ), and thus it is not prone to further fragmentation.

We then described the evolution of perturbations induced by turbulent random forcing. In this case the crossing time entering η should be taken as the minimum of the sound and the turbulent ($\tau_{\rm u}$) crossing times. Thus, small-scale (~ 0.3 pc) *velocity* fluctuations are quasi-isobaric at very small amplitudes, because in this case $\tau_{\rm u} > \tau_{\rm c} > \tau_{\rm s}$ so that the flow can cool in response to the velocity perturbation. As the perturbation amplitude is increased, so that $\tau_{\rm c} > \tau_{\rm s} > \tau_{\rm u}$, the situation changes because now the density is driven by the turbulent velocity rather than by sound waves, and the perturbations become quasi-adiabatic in character, *becoming stable*. We empirically found this to occur roughly when the rms Mach number $\gtrsim 0.3$. Finally, however, if the perturbation amplitude becomes very large, then the density increment induced by it becomes nonlinear and accelerates the cooling rate, effectively causing $\eta < 1$. In this case, velocity fluctuations trigger the condensation mode, which is again unstable, and cause condensations.

Thus, we reached the important conclusion that small-scale fluctuations behave very differently when they are entropy perturbations (caused, for example, by local variations in the heating or cooling rates) and when they are adiabatic (caused by velocity fluctuations), being unstable (and with the fastest growth rates) in the former case, but linearly stable in the latter.

We then considered the magnetic field as an additional source of pressure in the ISM, confirming earlier results that at low and intermediate densities the magnetic pressure is strongly decorrelated from density in fully turbulent cases (large field fluctuations), and proposed an interpretation of this phenomenon in terms of the scaling of B^2 with density for the slow and fast modes of simple nonlinear MHD waves. The decorrelation between magnetic pressure and density has several implications, among which is that the magnetic field probably is ineffective in supplementing thermal pressure in highly turbulent, thermally unstable conditions, and that it is probably inadequate to model magnetic pressure by means of an equivalent polytropic behavior in the fully turbulent case.

Finally, we discussed results from simulations of the ISM in more than one dimension at large and intermediate scales and at various resolutions. To this end, we first performed a detailed study of the competition between numerical diffusivities and the growth of TI, finding that even when the diffusivities (especially the mass diffusion, which is necessarily numerical) are confined to the smallest scales on the numerical grid, they can push the smallest unstable scale (the "Field" length $\lambda_{\rm F}$) to relatively large scales in the simulations with large box sizes, because $\lambda_{\rm F}$ scales as $\lambda_K^{1/2}$, where λ_K is the diffusive scale.

With this information, we discussed the fact that many ISM simulations suggest that the basic structure does not depend sensitively on whether TI is present, as long as there are turbulent motions driven by stellar-like sources (that imply recycling of gas from the cold to the warm phase), and that significant fractions of the gas mass (15-50%) appear to be in the unstable regime. This appears to be a consequence of the fact that the diffuse medium is in a moderately turbulent state, so that a) the fluctuations there have a regular pressure gradient and b) the magnetic field is not strongly turbulent, and therefore may cause additional stability. Of course, when the relatively quiescent intercloud medium is hit by a strong shock from, say, a supernova remnant, then TI can be rapidly induced, as in the studies by Hennebelle & Pérault [31,32] and Koyama & Inutsuka [43].

A final remark of interest is that it may be possible to determine observationally whether the gas seen at unstable temperatures corresponds to the out-of-thermal-equilibrium gas observed in the simulations by either a) simultaneously determing two of its thermodynamic variables, or b) comparing directly observed cooling rates (e.g., fine structure lines) with theoretical estimates of the heating rate (e.g., photoelectric heating) in specific regions (C. Heiles, private communication). If this is confirmed, then it would provide strong evidence that turbulent motions populate all regions of the thermodynamic variable space, preventing a sharp segregation of the atomic ISM into the stable phases of TI.

Acknowledgements

We have greatly benefitted from exchanges with C. Heiles, P. Hennebelle, H. Koyama, J. Scalo and E. Zweibel. The report from an anonymous referee prompted much improvement of the paper and led us to the study of numerical damping of the growth rates. This work has received partial financial support from

CONACYT grant 27752-E, from the French national program PCMI, and from the conference organizers to E.V.-S. We have made extensive use of NASA's Astrophysics Data System Abstract Service.

References

1. V. Avila-Reese, E. Vázquez-Semadeni: ApJ **553**, 645 (2001)
2. S.A. Balbus: 'Thermal Instability'. In: *The Physics of the Interstellar Medium and Intergalactic Medium*, ed. by A. Ferrara, C.F. McKee, C. Heiles, P.R. Shapiro (Astronomical Society of the Pacific, San Francisco 1995), p. 328
3. J. Ballesteros-Paredes, E. Vázquez-Semadeni, J. Scalo: ApJ **515**, 286 (1999)
4. T.M. Bania, J.G. Lyon: ApJ **239**, 173 (1980)
5. M.C. Begelman, C.F. McKee: ApJ **338**, 375 (1990)
6. A. Brandenburg: ApJ **550**, 824 (2001)
7. W. Brinkmann, S. Massaglia, E. Müller: Astron. Ap. **237**, 536 (1999)
8. A. Burkert, D.N.C. Lin: ApJ **537**, 270 (2000)
9. D. Chappell, J. Scalo: MNRAS **325**, 1 (2001)
10. W.-H. Chiang, K.H. Prendergast: ApJ **297**, 507 (1985)
11. W.-H. Chiang, J.N. Bregman: ApJ **328**, 427 (1988)
12. A. Dalgarno, R.A. McCray: Ann. Rev. Astron. Ap. **10**, 375 (1972)
13. L.P. David, J.N. Bregman, C.G. Seab: ApJ **329**, 66 (1988)
14. M.A. de Avillez: MNRAS **315**, 479 (2000)
15. J.M. Dickey, E.E. Salpeter, Y. Terzian: ApJ **211**, L77 (1977)
16. B.G. Elmegreen:, ApJ **378**, 139 (1991)
17. B.G. Elmegreen: 'The Origin and Evolution of Giant Molecular Clouds'. In: *The Physics of Star Formation and Early Stellar Evolution*, ed. by C.J. Lada, N.D. Kylafis (Kluwer, Dordrecht, 1991), p. 35
18. B.G. Elmegreen: ApJ **419**, L29 (1993)
19. B.G. Elmegreen: ApJ **433**, 39 (1994)
20. B.G. Elmegreen: ApJ **480**, 674 (1997)
21. K.M. Ferrière: Rev. Mod. Phys. **73**, 1031 (2001)
22. G.B. Field: ApJ **142**, 531 (1965)
23. G.B. Field, D.W. Goldsmith, H.J. Habing: ApJ **155**, L149 (1969)
24. E.L. Fitzpatrick, L. Spitzer: ApJ, **475**, 623 (1997)
25. A. Gazol, T. Passot: ApJ **518**, 748 (1999)
26. A. Gazol, E. Vázquez-Semadeni, F.J. Sánchez-Salcedo, J. Scalo: ApJ **557**, 121 (2001) (Paper IV)
27. H. Gerola, M. Kafatos, R. McCray, R.: ApJ **189**, 55 (1974)
28. J.P.E. Gerritsen, V. Icke: Astron. Ap. **325**, 972 (1997)
29. D.W. Goldsmith: ApJ **161**, 41 (1970)
30. C. Heiles: ApJ **551**, L105 (2001)
31. P. Hennebelle, M. Pérault: Astron. Ap. **351**, 309 (1999)
32. P. Hennebelle, M. Pérault: Astron. Ap. **359**, 1124 (2000)
33. J.H. Hunter: ApJ **161**, 451 (1970)
34. J.H. Hunter: ApJ **166**, 453 (1971)
35. E.B. Jenkins, T.M. Tripp: ApJS **137**, 297 (2001)
36. P.M.W. Kalberla, U.J. Schwarz, W. M. Goss: Astron. Ap. **144**, 27 (1985)
37. H. Kang, P.R. Shapiro, S.M. Fall,M.J. Rees: ApJ, **363**, 488 (1990)
38. H. Kang, G. Lake, D. Ryu: J. Kor. Astron. Soc. **33**, 111 2000

39. J. Kim, D. Balsara, M.-M. Mac Low: J. Korean Astron. Soc. **34**, 333 (2001)
40. P. Kornreich, J. Scalo: ApJ **531**, 366 (2000)
41. M.J. Korpi, A. Brandenburg, A. Shukurov, I. Tuominen, ÅNordlund: ApJ **514**, L99 (1999)
42. H. Koyama, S.-I. Inutsuka: ApJ **532**, 980 (2000)
43. H. Koyama, S.-I. Inutsuka: ApJ **564L**, 97 (2001)
44. A. Kritsuk: Sov. Astron. **34**, 21 (1990)
45. A. Kritsuk, M.L. Norman: ApJ **569L**, 127 (2002)
46. L.D. Landau, E.M. Lifshitz: *Fluid Mechanics*, 2nd ed. (Pergamon Press, Oxford, 1987)
47. K. R. Lang *Astrophysical Formulae vol. I*, 3th edn., (Springer-Verlag, Berlin, 1999)
48. A. Lioure, J.-P. Chièze: Astron. Ap. **235**, 379 (1990)
49. Y. Lithwick, P. Goldreich: ApJ **562**, 279 (2001)
50. M. Loewenstein: ApJ **349**, 471 (1990)
51. M.-M. Mac Low, D. Balsara, J. Kim, M. de Avillez: astro-ph/0106509
52. G. Mann: J. Plasma Phys. **53**, 109 (1995)
53. C.F. McKee, J.P. Ostriker: ApJ **218**, 148 (1977)
54. B. Meerson, P.V. Sasorov: Sov. Phys. JETP Lett. **65**, 300 (1987)
55. B. Meerson: Rev. Mod. Phys. **68**, 215 (1996)
56. S.L. Mufson: ApJ **193**, 561 (1974)
57. S.L. Mufson: ApJ **202**, 372 (1975)
58. S.D. Murray, D.N.C. Lin: ApJ **339**,933 (1989)
59. S.D. Murray, D.N.C. Lin: ApJ **363**, 50 (1990)
60. C.A. Norman, A. Ferrara: ApJ **467**, 280 (1996)
61. E.E. Oran, J.T. Mariska, J.P. Boris: ApJ **254**, 349 (1982)
62. E.C. Ostriker, J.M. Stone, C.F. Gammie: ApJ **546**, 980 (2001)
63. P. Padoan, Å Nordlund: ApJ **526**, 279 (1999)
64. A. Parravano: ApJ **172**, 280 (1987)
65. E.N. Parker: ApJ **117**, 431 (1953)
66. T. Passot, E. Vázquez-Semadeni, A. Pouquet: ApJ **455**, 536 (1995)
67. T. Passot, E. Vázquez-Semadeni: submitted to Astron. Ap. (2002) (Paper II)
68. S.B. Pikel'ner: Sov. Astron. **11**, 737 (1968)
69. A. Rosen, J.N Bregman: ApJ **440**, 634 (1995)
70. P.V. Sasorov: Sov. Astron. Lett. **14**, 129 (1988)
71. F.J. Sánchez-Salcedo, A. Brandenburg: Mon. Not. Royal Astron. Soc. **322**, 67 (2001)
72. F.J. Sánchez-Salcedo, E. Vázquez-Semadeni, A. Gazol: ApJ submitted (2002) (Paper I)
73. J. Scalo, D. Chappell: MNRAS **310**, 1 (1999)
74. J. Schwarz, R. McCray, R.F. Stein: ApJ **175**, 673 (1972)
75. J.A. Sellwood, S.A. Balbus: ApJ **511**, 660 (1999)
76. F.H. Shu: *The Physics of Astrophysics. Vol. II: Gas Dynamics*, (University Science Books, Sausalito, 1992)
77. L. Spitzer, E.L. Fitzpatrick: ApJ **445**, 196 (1995)
78. E. Vázquez-Semadeni, T. Passot, A. Pouquet: ApJ **441**, 702 (1996)
79. E. Vázquez-Semadeni, T. Passot, A. Pouquet: ApJ **473**, 881 (1996)
80. E. Vázquez-Semadeni, A. Gazol, J. Scalo: ApJ **540**, 271 (2000) (Paper III)
81. K. Wada, M. Spaans, S. Kim: ApJ **540**, 797 (2000)
82. M.G. Wolfire,D. Hollenbach, C.F. McKee, A.G.G.M. Tielens, E.L.O. Bakes: ApJ **443**, 152 (1995)
83. Ya. B. Zel'dovich, S.B. Pikel'ner: Sov. Phys. JETP **29**, 170 (1969)

Developing Diagnostics of Molecular Clouds Using Numerical MHD Simulations

Eve C. Ostriker

Department of Astronomy, University of Maryland, College Park, MD 20742, USA

Abstract. An important aspect of astrophysical MHD turbulence research is developing diagnostics to connect simulations with the observable universe. Turbulent systems are by definition structurally complex in all fluid variables (density, velocity, and magnetic field), such that they must be described statistically. By developing and applying diagnostic tools to simulation data, it is possible to interpret empirical laws for the statistical properties of observed systems in terms of fundamental dynamical processes, and to identify and calibrate robust probes of physical parameters that cannot be measured directly. Using several different examples, I describe how structural diagnostic analyses have already yielded significant insights into the nature of turbulent molecular clouds. I review results from several different groups, and discuss directions for future diagnostics to enhance our understanding of cloud structure and constrain models of the evolutionary course that governs star formation.

1 Introduction

As the number, range, and depth of the papers in this volume witnesses, recent progress in modeling and understanding astrophysical MHD turbulence is impressive. Even with the intensive research of several groups over the past few years, however, many aspects of the fundamental turbulence phenomenon are not yet wholly understood – which makes for continuing excitement in this emerging discipline. In using the results of MHD simulations to interpret the dynamics of the interstellar medium, the technical challenges involved in numerically modeling and characterizing turbulence are compounded by astrophysical uncertainties in posing the numerical problem to be solved. For molecular clouds, open astrophysical questions include:

- What is the source (original, and potentially, maintaining) of turbulence?
- What is the mean magnetic field strength, and variation of mass-to-flux ratio, in molecular clouds?
- What is the range of sizes and masses of molecular clouds in the Milky Way and other spiral galaxies?
- How are clouds formed? how long do they live? how are they destroyed?

From the point of view of defining an idealized problem for an MHD simulation, these astrophysical questions translate to uncertainties in the input spectral form (in space and time) of the turbulent driving, the value and variation of the plasma β parameter, the importance of self-gravity, and the initial and boundary conditions for the simulation.

E. Falgarone and T. Passot (Eds.): LNP 614, pp. 252–270, 2003.
© Springer-Verlag Berlin Heidelberg 2003

The complexity of the turbulence phenomenon demands detail and variety in the analytical methods used to characterize its structure. For application to understanding astronomical systems – where the physical inputs are uncertain, and only projected distributions are available – the eventual aim is to develop a set of simple, robust diagnostics of MHD turbulence that have direct connections to observable quantities. Potentially, there are many different avenues for this sort of analysis, and extensive exploration is required to determine what directions are most productive. Because large-scale numerical simulations of turbulence under interstellar conditions are only now becoming computationally practical, the present tasks include first the "forward" process of characterizing MHD turbulence obtained from new simulations with a range of parameter values, and then using these results to select and calibrate diagnostics for the "reverse" process of discriminating systemic parameters from observables.

As examples of the process of developing diagnostics of molecular clouds' internal structure, kinematics, and magnetization from turbulent MHD simulations, my discussion here will focus on recent work analyzing and interpreting density and column density statistics (Sect. 2), properties and definitions of clumps (Sect. 3), linewidth-size relations (Sect. 4), and statistics of polarization maps (Sect. 5). I will update previous work (see also [44], [30], [31], [11]), present several new results, and make connections to the conclusions of other groups. Chapters in the same volume covering topics most directly related to those addressed here include those by Crutcher, Heiles, & Troland; Nordlund & Padoan; MacLow; Cho, Lazarian & Vishniac and Zweibel, Heitsch, & Fan.

2 Density and Column Density Statistics

In general, a turbulent velocity field leads to production of significant local density variations in a compressible medium (i.e. a medium with Mach number $\mathcal{M} \equiv v_{\rm rms}/c_{\rm s} \gg 1$, where $v_{\rm rms}$ is the turbulent velocity dispersion and c_s the sound speed). This is true regardless of the magnetic field strength, because in the case of a weak magnetic field ($v_{\rm A} \equiv B_0/\sqrt{4\pi\bar{\rho}} \sim c_s \ll v_{\rm rms}$, where $\boldsymbol{B}_0$ is the mean magnetic field and $\bar{\rho}$ is the mean density), magnetic pressure forces are weak compared to ram pressure, and in the case of strong magnetic fields ($v_{\rm A} \sim v_{\rm rms}$), compression is unhindered along the mean field direction (and indeed enhanced by forces associated with gradients in $B_\perp^2$, where $\boldsymbol{B}_\perp$ is the component perpendicular to $\boldsymbol{B}_0$).

If compression and rarefaction events are spatially and temporally independent, then for the case of an isothermal equation of state (approximately true under molecular cloud conditions), the resultant one-point density distribution function (often referred to as a "PDF" – probability distribution function) is expected to obey a lognormal form ([39], [29]). If $1 + \delta_i$ is the enhancement/decrement factor for density in the i^{th} compression/rarefaction event affecting a given fluid element, then the density after N events will be

$$\rho = \bar{\rho}\Pi_{i=1}^{N}(1 + \delta_i) \tag{1}$$

so that the logarithm of the density,

$$\log(\rho/\bar{\rho}) = \Sigma_{i=1}^{N} \log(1+\delta_i), \tag{2}$$

will be the sum of independent random variables; by the Central Limit Theorem, this implies that $\log(\rho/\bar{\rho})$ should obey a Gaussian distribution.

The results of numerical simulations ([46],[32], [39],[30],[31]) indeed bear out the expectation that a log-normal form for the volume density PDF prevails (at least away from the tails) under isothermal conditions. This result holds both for forced and decaying turbulence, and for simulations with varying mean magnetic fields. Figure 1 shows an example of the distributions of fractional volume and mass as a function of volume density for four forced-turbulence MHD simulations with $\mathcal{M} = 5$ (see [44] for details on the models), with comparisons to the lognormal functions with the same mean and dispersion.

As the example in Fig. 1 shows, the average mass compression factor is relatively insensitive to the mean magnetic field strength. Although the minimum value of the mass-weighted mean $\langle \log(\rho/\bar{\rho})\rangle_M$ increases (logarithmically) with the fast-magnetosonic Mach number $\mathcal{M}_{\rm F} \equiv v_{\rm rms}/(c_{\rm s}^2 + v_{\rm A}^2)^{1/2}$, the scatter from "cosmic variance" is large enough that there does not appear to be a unique relation between $\mathcal{M}_{\rm F}$ (or $\mathcal{M}$; cf. [32], [29]) and $\langle \log(\rho/\bar{\rho})\rangle$ [31]. Typical values of the mean compression factor for matter in the simulations, $\langle \rho/\bar{\rho}\rangle_M$, range over 3–6 for $\mathcal{M} = 5-9$, consistent with the compression factor needed to excite the CO molecule – rendering molecular clouds observable – when the volume-weighted density is only $n \sim 100\,{\rm cm}^{-3}$ (e.g. [42]).

More directly observable than the distribution of volume densities in a cloud is its distribution of column densities, corresponding to line-of-sight integrations of density, $\Sigma \equiv \int \rho {\rm d}s$. Heuristically, one might expect the column density at any projected position to be determined by a series of independent (in space and time) compressions and rarefactions, similarly to the process of events described in (1). The difference for column density is that each event affects only a fraction $f_i < 1$ of the line of sight, so that the column density is given by

$$\log(\Sigma/\bar{\Sigma}) = \Sigma_i \log(1+f_i\delta_i) \tag{3}$$

instead of (2). The factor f_i may be thought of as the ratio of the correlation length of a given compression/rarefaction event to the overall linear scale of the cloud along the line of sight. If the individual enhancement/decrement factors are independent random variables, then the resultant column density PDF should take on a log-normal form. Because each $1+f_i\delta_i$ is closer to unity than $1+\delta_i$, however, the mean and dispersion of $\log(\Sigma/\bar{\Sigma})$ are expected to be smaller than the corresponding quantities for $\log(\rho/\bar{\rho})$. These expectations are indeed borne out by analyses of column densities in simulations, as shown by [31]; distributions show a lognormal form (see also Fig. 2), and typical values of the mass-weighted mean column density are $\langle \Sigma/\bar{\Sigma}\rangle_M = 1.1-1.4$ for $\mathcal{M} = 5-9$.

From (3), note that if the factors f_i are small (corresponding to having the dominant correlation length small compared to the size of the numerical box/physical cloud), then $\delta\Sigma/\bar{\Sigma} \equiv \Sigma/\bar{\Sigma} - 1 \approx \Sigma_i f_i \delta_i$, implying a Gaussian

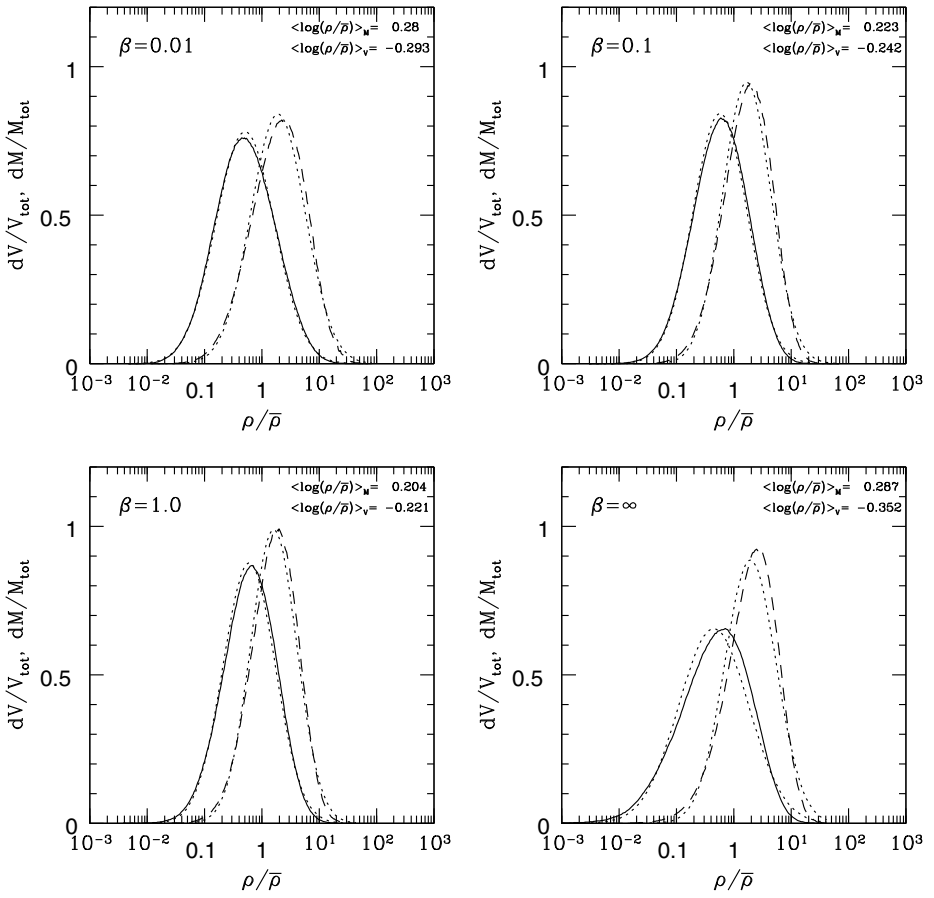


Fig. 1. Distributions of fractional volume (solid curves) and fractional mass (dashed curves) as a function of density for Mach-5 forced turbulence simulations with varying mean magnetic field strength characterized by $\beta \equiv c_s^2/v_A^2$, for $v_A = B_0/\sqrt{4\pi\bar{\rho}}$. Dotted curves show lognormal distributions for comparison. Each panel is also labeled with the mass- and volume-weighted mean compression magnitudes.

distribution for $\delta\Sigma/\bar{\Sigma}$ if the terms $f_i\delta_i$ are independent random variables. Numerical evidence on how column PDFs transition from Gaussian to lognormal form as f_i increases is presented in [47].

Preliminary comparisons between observed PDFs of column density in filamentary molecular clouds – obtained from stellar extinction data [2] – and simulated PDFs from turbulence models are very encouraging [31]. The data are consistent with lognormal distributions, with comparable width to those from simulations having turbulent Mach numbers and power spectra comparable to those in observed clouds. It remains to be seen how much more specific information about a cloud can be learned from its column density. Although there were earlier some hopes that column density PDFs could help distinguish the

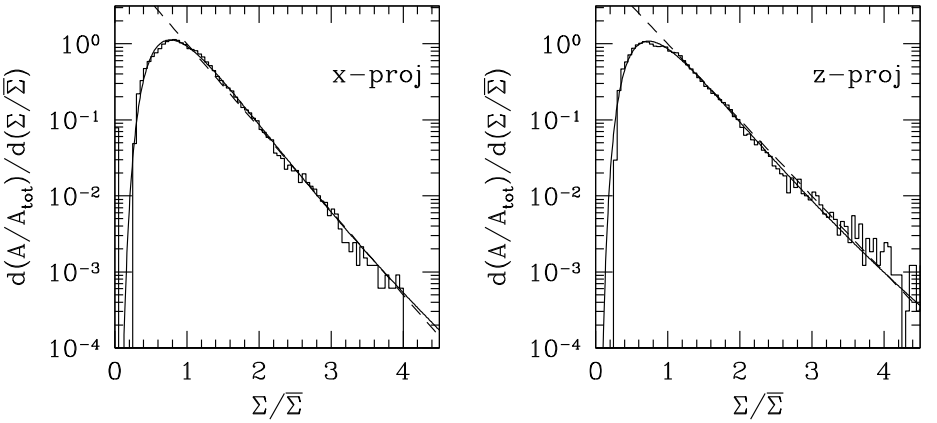


Fig. 2. Distributions of fractional area (histograms) as a function of column density for Mach-7 decaying turbulence, for two different projection directions, from $\beta = 0.01$ simulation (see [31] for details). Solid curves show the corresponding lognormal distribution fits; dashed lines indicate exponential fits on high-column side, with slopes -1.1 and -1 for the $\hat{x}$ and $\hat{z}$ projections.

mean magnetic field strength in clouds [33], models with matched Mach numbers "observed" from varying directions do not show strong or consistent correlations of the mean column density contrast with the value of $\beta \equiv c_{\rm s}^2(B_0^2/4\pi\bar{\rho})^{-1}$ [31]. In spite of this insensitivity to the mean magnetic field strength, one might still hope to constrain the distribution of volume densities from the distribution of a cloud's column densities. The mathematical degeneracy between δ_i (the volume compression/rarefaction increment) and f_i (the spatial coherence length of an event) evident in (3) indicates that there is no simple inversion method. Methods that combine the column PDF with the spatial (two-point) correlation function in column density maps (reflecting f_i, or more generally, the shape of the density power spectrum) may however be able to lift this degeneracy; this represents an important direction for future study.

Because the population of the tails of the density and column density PDFs may be more affected by intermittency than the population near the peak, and because the equation of state may depart from isothermality in more overdense, optically thick regions (cf. [41],[29]) observational departures from log-normality are more likely to occur there. Simulations (even with a uniform, isothermal equation of state) show a variety of behaviour in the tail distributions, although PDFs that are lognormal over more than three orders of magnitude are common from isothermal decaying-turbulence simulations. Figure 2 shows, for example, that for column densities less than $\Sigma/\bar{\Sigma} = 4$ (corresponding to some 98% of the projected area of the particular simulation), a lognormal function is an excellent approximation.

Over a limited range of column densities above the mean, lognormal PDFs can typically be fit by an exponential function as well (i.e. $\log[dA/d\Sigma] = C_1 + C_2\Sigma/\bar{\Sigma}$ for C_1 and C_2 constants). Fig. 2 shows examples of local fits of this kind

with slopes $C_2 = -1.1, -1$. Similarities between this local exponential form in simulations and in molecular-line observations [6], with emphasis on potential dependences of the slope on the largest density correlation scale, have been investigated by [8] (see also [31] for related discussion of resolution effects). It is not yet clear whether there is inherent physical significance in the local exponential form, or whether it is primarily a convenient mathematical approximation to a lognormal on the high-column-density side of the distribution where most of the observable matter is found.

3 Clumps in Turbulent Clouds

A longstanding unsolved problem in astrophysics is what determines the stellar initial mass function (IMF). Many different physical processes could potentially affect the IMF; an abbreviated list includes: (i) turbulent stresses in a large-scale cloud producing non-self-gravitating clumps with a range of masses/sizes, (ii) self-gravity in inhomogeneous clumps/cores leading to sub-fragmentation of collapsing condensations, (iii) dynamical instabilities in massive disks – formed by the initial collapse of rotating cores – fragmenting them into binary or multiple star/disk systems, (iv) outward momentum flux from stellar radiation and/or MHD winds truncating accretion onto forming stars from the outer parts of their parent cores. The relative importance to the final IMF of each of these (and other) processes remains to be determined, and major technical challenges are involved in attacking any of these questions via direct numerical simulations. The large range of scales (nine orders of magnitude!) involved in going from a cloud to a star points to the need for adaptive mesh refinement (AMR) schemes in order to follow even a portion of the overall process.

For the present, we can begin by assessing the clumpy structure produced by turbulence and self-gravity at moderate scales in uniform-grid MHD simulations of GMCs. This clumpy structure can be characterized in many ways, and applying varying methods is valuable for understanding and illustrating different aspects of the formal dynamical problem, as well as for interpreting observations.

One class of *non-hierarchical* structure-analysis methods is similar to the CLUMPFIND algorithm introduced by [48]. In this method (see [11] for details), one chooses a threshold volume density $\rho_{\rm th}$ (or column density $\Sigma_{\rm th}$), identifies the set of local maxima in the data cube (potentially first smoothing the data to reduce pixel noise), and then defines clumps by assigning matter at $\rho > \rho_{\rm th}$ to the nearest local maximum. The shape of a clump can be quantified by computing the ratios of principal axes in its moment of inertia tensor, and the importance of self-gravity in binding a condensation can be described, e.g., by computing the ratio of gravitational energy to the (weighted) sum of kinetic, thermal, and magnetic energies.

A full discussion of the results of applying this clump-finding method to a set of decaying-turbulence (with initial $\mathcal{M} = 14$), self-gravitating MHD simulations is given in [11]. Results from this analysis (taking $\rho_{\rm th}/\bar{\rho} = 10$) include:

- At any time, only the high-mass wing of the clump distribution is self-gravitating;

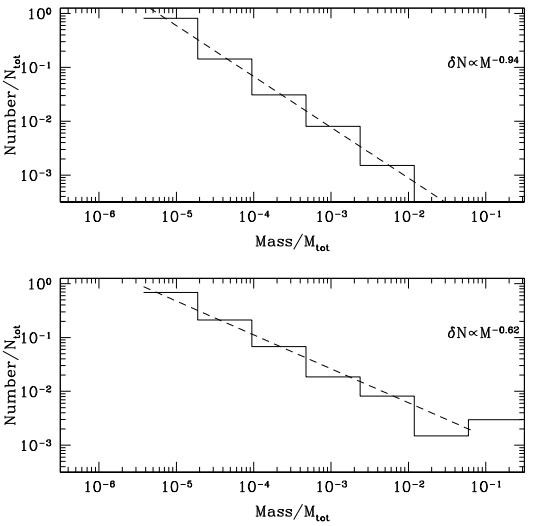


Fig. 3. Number of clumps as a function of mass for two different clump identification algorithms. Lower panel shows results of excluding clumps that lie within larger clumps; upper panel shows results of counting clumps at any given spatial scale regardless of whether they lie within a larger concentration.

- Characterizing the spectral shape of the high-mass wing as $\mathrm{d}N/\mathrm{d}M \propto M^{-x}$, the slope x is in the range 2–3, becoming shallower over time from mergers;
- The turnover in the mass spectrum is spatially well-resolved (at ~ 8 grid zones across) with $M_{\rm peak}/M_{\rm tot}$ of a few times 10^{-4};
- The minimum clump mass is typically a factor 10 below the peak value;
- Clump shapes are intrinsically triaxial, and project to having two-dimensional axis ratios $\sim 2:1$;
- The shapes of the mass functions of apparent clumps in column density maps are similar to those of true three-dimensional clumps, but shifted to larger masses by an order of magnitude.

An interesting point is that $M_{\rm peak}$, and also the minimum mass $M_{\rm min}$ of a clump for given $\rho_{\rm th}/\bar{\rho} = 10$, do not vary with the value of β. Also, although the Mach number declines by a factor 4–5 over the course of the simulations, the peak and minimum masses do not change significantly. This suggests that the mass function of clumps retains a "memory" of the dynamical history of a cloud, rather than being determined solely by the cloud's instantaneous turbulent Mach number and spectrum (as proposed in [34]).

The results for typical slopes of the high-mass end of the clump mass function are intriguingly similar to the value 2.35 for the Salpeter stellar IMF, which also appears to describe the core IMF in forming clusters [1]. Similar results have also been obtained by analysis of a variety of simulations by other groups (e.g. [18], [37],[4]). While the conclusions from these preliminary analyses are

promising, many questions still remain open. It not yet clear in general how the mass scales (the peak, minimum, and maximum) depend on the input parameters ($\mathcal{M}$, as well as the spectral shape, and potentially β) and on dynamical history. Uncertainties also remain in how three-dimensional clumps relate to clumps seen in projection, and whether the latter distribution can be used to deduce the former.

Another important set of questions is how the definition of a "clump" – both the specific clump identification algorithm with its chosen set of parameters (such as $\rho_{\rm th}$), and the overall category of algorithms in which a specific method lies – affects the results. A clear categorical distinction is between non-hierarchical identification algorithms like CLUMPFIND (in which every mass element is assigned to a single clump), and hierarchical algorithms, in which a given mass element may be counted as part of many clumps, at different levels of a hierarchy. As an astronomical analogy, it is clear that for many purposes it is valuable to count galaxies whether or not they are part of larger clusters or supercluster; counting "objects" is a function of the spatial scale under consideration.

As an example of how hierarchical considerations affect the mass spectrum, consider the distributions shown in Fig. 3. For this analysis, a "clump" is any cubic region at a given spatial scale in which the density exceeds the mass-weighted mean density for the ensemble of cubes at that scale (see [31] for further details on this "region of contrast" algorithm). For the lower panel in Fig. 3, only clumps that do not lie within other clumps are counted for the mass spectrum; for the upper panel, the spectrum counts clumps regardless of "overlap." As should be expected, allowing for clumps-within-clumps leads to a relatively steeper mass spectrum, with a slope $\mathrm{d}N/\mathrm{d}M \propto M^{-1.9}$ in this instance. The no-overlap spectrum is $\mathrm{d}N/\mathrm{d}M \propto M^{-1.6}$. This difference between the hierarchical and non-hierarchical mass spectra is very interesting because it is reminiscent of the difference between the steep stellar IMF, and shallower observationally-determined GMC mass functions (e.g. [43]). Because the density condensations produced by turbulence as "initial conditions" for collapse are hierarchically nested, this difference in slopes offers intriguing support for the idea that fragmentation during gravitationally-collapsing stages may play a crucial role in defining the stellar IMF. An important question for future AMR simulations to address is when and why gravity chooses an "inner" versus "outer" mass scale for final collapsed objects.

4 Linewidth-Size Relationships

The direct observables produced by spectral-line mapping of a molecular cloud are data cubes of intensity as a function of two plane-of-sky positions and the line-of-sight velocity (in radio astronomy, the intensity is described as a brightness temperature). In principal, one would like to extract the spatial distribution of velocity and emissivity from these data cubes. Because only projected data are available, and a turbulent cloud has no spatial symmetries to exploit, direct inversion is not possible. However, one may still hope to deduce statistical

properties of the turbulence from the full intensity data cube. Various complex techniques to do this are under development by several groups – including Principal Component Analysis [16], [7]; the Spectral Correlation Function [40],[35]; and Velocity Channel Analysis [22], [23]. Here, I will briefly discuss a simple technique to estimate a theoretically fundamental – and observationally much-investigated – property of turbulence, the variation of linewidth with physical size scale.

Averaged over volumes with the *same size* in all three directions, the mean linewidth simply reflects the underlying one-dimensional velocity power spectrum, since

$$\sigma_v^2(s) = \int\int\int_0^s \mathrm{d}^3x\, v^2(\boldsymbol{x}) = \frac{1}{(2\pi)^3}\int\int\int_{2\pi/s}^{\infty} \mathrm{d}^3k\, v^2(\boldsymbol{k}). \tag{4}$$

If the turbulence has a power-law spectrum, $v^2(\boldsymbol{k}) \propto |k|^{-\alpha}$, then $\sigma_v(s) \propto s^{(\alpha-3)/2}$, so for e.g. Kolmogorov or Burgers spectra with $\alpha = 11/3$ or 4, $\sigma_v \propto s^{1/3}$ or $s^{1/2}$.

In observations, however, any region of size s^2 in projection (on the plane of the sky) in general extends over a scale at least as large along the line of sight. If s is small compared to the overall scale of a cloud, L, then the observed linewidth from a region of projected area s^2 can have contributions from $L/s \gg 1$ volume elements of size s^3 along the line of sight. If power increases with scale ($\alpha > 3$), then the velocity centroids of the multiple s^3 volume elements *on average* differ, such that the linewidth over the area s^2 integrated over the whole line of sight will *on average* exceed $\sigma_v(s)$, reaching up to $\sigma_v(L)$. However, because of the statistical nature of the distribution (e.g. if it obeys Gaussian random statistics), for *some* projected positions the centroids of the multiple s^3 volume elements will differ very little, such that the line-of-sight integrated velocity dispersion will be close to $\sigma_v(s)$. From this argument, one expects that the *mean linewidth* would vary weakly with projected size, whereas the *lower envelope* of the linewidth distribution would vary more strongly with projected size, and in fact trace the underlying three-dimensional mean linewidth-size relation. Analysis of simulation data cubes indeed bears out this expectation [31], as shown for example in Fig. 4.

The foregoing discussion is helpful for interpreting well-known observational aspects of molecular cloud kinematic scalings. In particular, by relation to Fig. 4, one may understand why the mean linewidth-size relation for apparent clumps observed in molecular tracers with moderate $\rho_{\rm crit}/\bar{\rho}$ is relatively flat (e.g. [45],[5], [48]) compared to the relatively steeper "Larson's Law" ($\sigma_v \propto s^{1/2}$) linewidth-size relations (cf. [19], [43]) that apply to objects that are spatially "isolated" – either because $\rho_{\rm crit}/\bar{\rho}$ is large (for dense cores within clouds) or because of phase differences with their surroundings (for molecular clouds within the atomic ISM). For tracers with critical density near the (mass-weighted) mean density in a cloud, it is not unlikely for multiple structures that are separated and in relative motion along the line of sight to contribute to the linewidth at a given position on a map. For tracers with higher critical densities, the probability

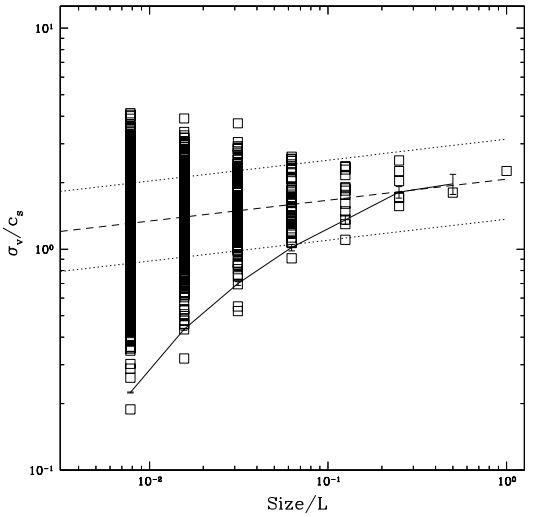


Fig. 4. Linewith-size relations for projected and 3D simulated data. Distribution (squares) shows total linewidth for apparent clumps in a map vs. linear dimension of projected clump (i.e. square root of its area); dashed line shows (linear) fitted mean linewidth-size distribution for projected clump map (slope is 0.2). Solid curve shows mean linewidth-size relation for 3D volumes from the same data cube. Data set is $\beta = 0.1$ model with Mach number 5 from [31], with threshold density $\rho/\bar{\rho} = 3$.

of chance projections of dense condensations is much lower because there are many fewer such condensations. Spectral correlation function analyses [35] also demonstrate that simulated spectral data cubes from higher-density tracers yield more spatial variation than do data cubes from lower-density tracers, for the same reason: the spectrum at a given point in a map from a low-density tracer samples more completely along a line of sight, and is thus more representative of the mean spectrum averaged over a whole cloud – compared to the spectrum from a high-density tracer.

This discussion of linewidth-size relations in 2D and 3D also serves to illustrate the point that coherent structures in position-velocity space in general differ from coherent structures in 3D position (physical) space; detailed analyses of simulation data cubes have shown this in a number of different ways (e.g. [31],[38],[4]). As a consequence, mass functions of density condensations, or other measures of structure in the density such as its Fourier power spectrum, cannot necessarily be obtained by treating the line-of-sight velocity as a surrogate for line-of-sight position. Instead, it is necessary to use statistical approaches to diagnose structure in the physical density distribution – just as statistical approaches are required for diagnosing structure in the velocity distribution. Depending on the relative power in velocity and density fluctuations, true density structure may be best discerned by integrating intensity over velocity and

then correcting for physical superposition of dense structures using information from two-point correlation functions in the column density (see [22] for related discussion).

5 Polarization as a Magnetic Field Diagnostic

It has long been recognized that polarization studies are important for diagnosing basic properties and structure of the ISM, because they provide relatively direct access to the elusive – but dynamically important – magnetic field (e.g. [13],[49]). *Provided* that dust grains preferentially align with short axes parallel to the local direction of the magnetic field ([10]; see e.g. recent review of [21] for thorough discussion of alignment mechanisms), an ordered $\boldsymbol{B}$-field will lead to observable polarization. For grains aligned with their short axes parallel to the local magnetic field, background stars are observed in optical/near-IR wavelengths with polarization parallel to the local magnetic field, and local dust emission is observed in far-IR/sub-mm wavelengths with polarization perpendicular to the local magnetic field. If the angle of $\boldsymbol{B}$ with respect to the plane of the sky is i, the local contribution to polarization is $\propto \cos^2 i$ times the difference between long- and short-axis grain crossections times the density of polarizing grains. Taking the simplest-possible assumption of a uniform ratio of polarizing-grain density to gas density (but see below), it is straightforward to create simulated polarization maps by integrating the radiative transfer equations for the Stokes parameters ([25],[26],[24]). For a medium with optical depth $\tau \ll 1$, the fractional polarization in (thermal) emission and in dust-absorbed starlight are related by $P_{\rm em} = P_{\rm abs}/\tau$ (e.g. [17]), so that polarization from emission is proportional to polarization from extinction divided by the column of intervening matter.

Maps and analyses of polarized extinction [31], and emission [36],[14] computed from MHD simulations have recently been presented by several groups. One question of interest is how the fractional polarization varies with the column $N \equiv \int n ds$ of absorbing or emitting material. For a uniform magnetic field, uniform polarization efficiency of grains, but spatially-nonuniform distribution of matter, $P_{\rm abs}$ would increase linearly with N, while $P_{\rm em}$ would be independent of N. For a spatially nonuniform field, the situation is much more complicated. In this case, the increase in $P_{\rm abs}$ or $P_{\rm em}\tau$ with N would (a) in general be shallower because variations in the magnetic field direction decorrelate the direction of grains, reducing the net contribution to polarization per unit length along the line of sight, and (b) no longer follow a linear relation if the amplitude of field fluctuations is large and/or if the number of effective correlation lengths of magnetic field orientation along the line of sight varies with column density (this can yield, e.g., $P_{\rm abs} \propto N^{1/2}$ [25]). For weaker mean magnetic fields B_0 and a given power spectrum of fluctuations δB, the directional decorrelation in polarization occurs at a physically smaller scale than for the strong-B_0 case, so that lower polarization is expected for a given column density. These expected trends are indeed evident in distributions of $P_{\rm abs}$ vs. column density from MHD simulations, as shown for example in Fig. 5.

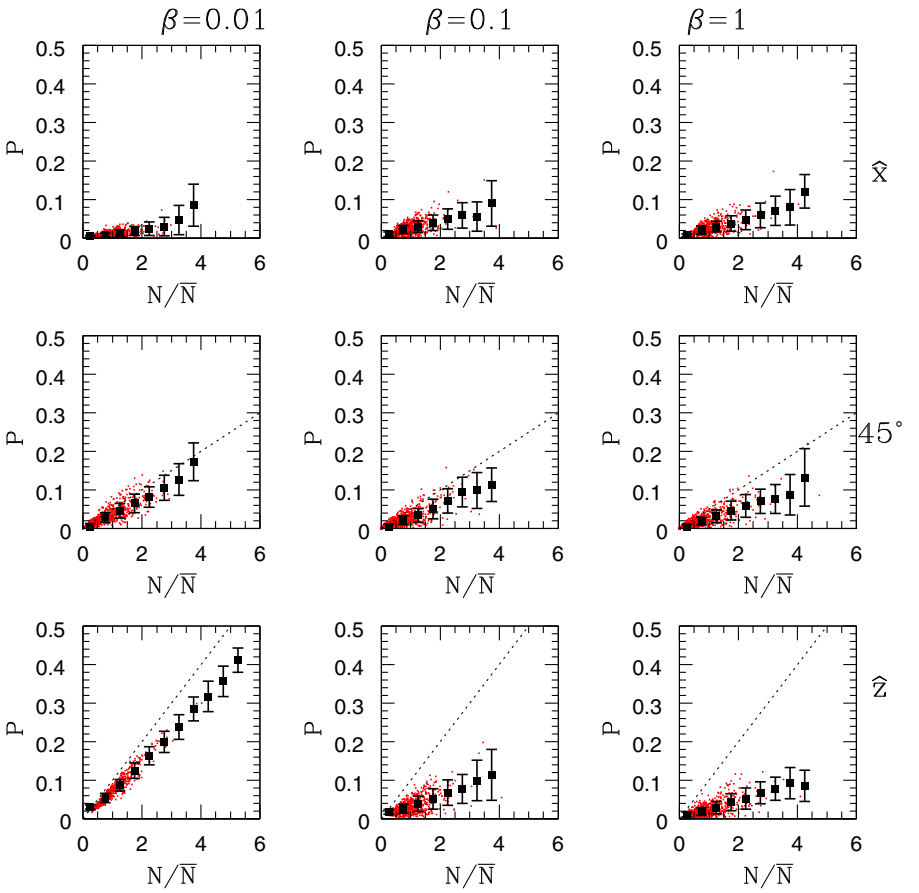


Fig. 5. Distributions of simulated polarized-extinction as a function of column density, for decaying-turbulence simulations with $\mathcal{M}$=7 from [31]. Scatter plots show $P_{\rm abs}$ for a randomly-chosen subset of positions on the projected map, for three different projection directions, and three different mean magnetic field strengths. Squares with 1-σ error bars show the mean $P-N$ relation for the full data sets. Dotted lines show what the $P-N$ relation would be for uniform magnetic fields. P is arbitrarily normalized to a value 0.1 for uniform density $N=\bar{N}$ for a projection perpendicular to $\hat{B}_0$

Interestingly, observed distributions (e.g.[12], [3]) of polarized-extinction vs. column (or A_V) in molecular clouds do not show the behavior evident in Fig. 5, which would correspond to a secular increase in $P_{\rm abs}$ up to A_V of 30 or more (for $\bar{N}$ corresponding to a typical GMC $A_V \sim 7.5$ [27]). Instead, the observed increase of $P_{\rm abs}$ with A_V flattens near A_V of unity, possibly indicating that grain alignment fails in the deep interiors of clouds (e.g. [20]). Additional support for this idea comes from comparison of observed distributions (e.g. [15]) of $P_{\rm em}$ vs. intensity (proportional to N) in dense cores with simulated distributions;

the simulations show insufficient decrease in $P_{\rm em}$ with column unless high-A_V regions have decreased polarizing efficiency [36].

Potentially, one of the most important applications for polarization studies is to use the variation in polarization directions to diagnose the strength of the magnetic field [9]. The basic physical idea behind the so-called "Chandrasekhar-Fermi" method is that weaker magnetic fields produce lower tension forces for a given displacement, so that for a given velocity field, a lower mean magnetic field strength results in larger distortions in the magnetic field direction. For the case of a single low-amplitude Alfvén wave in a uniform magnetic field with plane-of-sky component B_p, the magnetic field and velocity perturbations obey $\delta B/B_p = \delta v/v_{\rm A,p}$, where $v_{\rm A,p} = B_p(4\pi\bar{\rho})^{-1/2}$. For a wave with perturbation direction in the plane of the sky, the dispersion in polarization directions is $\delta\phi = \langle(\delta B/B_p)^2\rangle^{1/2}$ if the net polarization is either parallel or perpendicular to $\boldsymbol{B}$. If there is an Alfvén wave component with perturbations along the line of sight having the same amplitude as the component with perturbations in the plane of the sky, then $\delta v/v_{\rm A,p} \equiv \langle v_{\rm los}^2\rangle^{1/2}/v_{\rm A,p} = \delta\phi$, so that B_p is given by $B_{\rm CF} \equiv (4\pi\bar{\rho})^{1/2}\delta v/\delta\phi$. Thus, with measures of the mean density, the observed velocity dispersion, and the dispersion in polarization angles, the plane-of-sky field strength may in principle be estimated.

Since several of the idealizations described above do not hold in a real molecular cloud, it is desirable to test and/or recalibrate the Chandrasekhar-Fermi (hereafter "C-F") relation using MHD simulations, which allow for more complex dynamical structure. Figure 6 shows an example of such a test using simulated polarized-extinction data, comparing the true mean plane-of-sky field strength with the "one-wave/equipartition" C-F estimate given above. As reported in [31], when the dispersion in polarization angles is sufficiently small ($\delta\phi < 25^\circ$) – such that linear theory is adequate, a good estimate of the plane-of-sky field is $\langle B_p\rangle \sim 0.5B_{CF}$. Evidently, the presence of more than one wave along the line of sight reduces $\delta\phi$, so that B_{CF} tends to overestimate the true B_p. For larger dispersions in polarization angle, Fig. 6 shows that the measure $0.5B_{CF}$ "calibrated" at small $\delta\phi$ may either under- or over-estimate B_p. Note that even if the mean magnetic field is strong, $\delta\phi$ can be large – and linear theory inappropriate – for system orientations in which the mean magnetic field direction lies near the line of sight. Analyses of simulated polarized-emission maps, combined with kinematic measurements [14], or with synthetic radiative transfer spectral maps [36], yield similar results for testing of the C-F formula.

We note that the C-F method measures only the mean *plane-of-sky* magnetic field $\langle B_p\rangle$; Fig. 7(a) shows that even for small dispersion in polarization angles, there can be large variation in the *total* field strength compared to the "corrected" C-F estimate. In any individual cloud, the line-of-sight field can be estimated by the Zeeman effect (at least in principle; in practice this is difficult), with the two estimates combined to obtain the total three-dimensional field strength (e.g. [28]). Another potential caveat for observational application of the C-F method is that the principal correlation scale for the dispersion in magnetic field directions must be resolved in the plane of sky by the effective

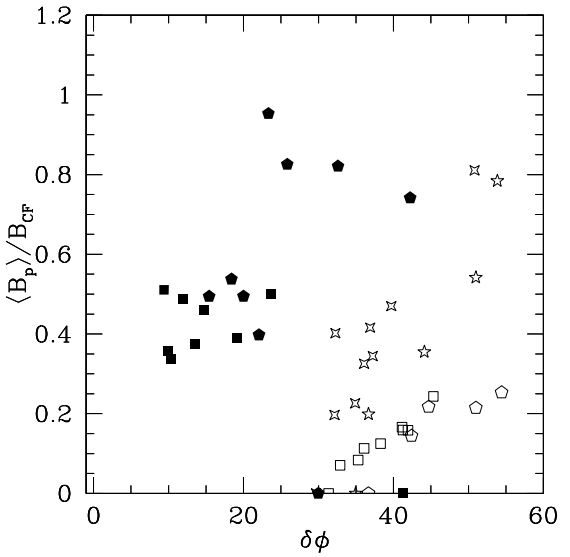


Fig. 6. Plane-of-sky component of the mean magnetic field, B_p, relative to Chandrasekhar-Fermi estimate B_{CF}, for $\mathcal{M}$=7 (four-sided symbols) and $\mathcal{M}$=9 (five-sided symbols) decaying turbulence MHD simulations [31]; $\delta\phi$ is the corresponding dispersion (in degrees) of directions in the polarized-extinction map. Solid, starred, and open symbols represent $\beta = 0.01,\ 0.1,\ 1.0$ models.

observational "beam". For polarized extinction, this poses no difficulties because background star observations have minimal beam thickness. For polarized emission from warm cores, however, large-beam averaging of spatial fluctuations tends to reduce $\delta\phi$ relative to its value for a "pencil-beam" observation, potentially resulting in an overestimate of B_p unless an appropriate correction factor is applied [14].

When variations in the polarization directions in a map are large, the fluctuations of the plane-of-sky magnetic field must be comparable to the mean value of B_p. In this case, it is not possible to distinguish from polarization studies alone whether (a) a cloud has a large mean magnetic field that is "hidden" along the line of sight (as is the case for the solid points in Fig. 6 having large $\delta\phi$), or (b) the mean magnetic field is simply weak (as is the case for the starred and open points in Fig. 6). Observed line-of-sight velocities combined with an assumption of equipartition between magnetic and kinetic energies can still yield an approximate measure of the rms magnetic field strength (see e.g. Fig. 7b) that is correct within a factor ~ 2, although an important caveat is that this estimate could be arbitrarily far off for cases with very strong mean magnetic fields ($v_{\rm A} \gg v_{rms}$) that happen to lie near the line of sight. [1] Fitting formulae that interpolate

[1] Also note that while B_{rms} is a dynamically-important quantity, it is not equivalent to the mean magnetic field strength $|\langle \boldsymbol{B} \rangle|$ that enters into the mass-to-flux ratio, which ultimately determines whether a cloud or core is super- or sub-critical.

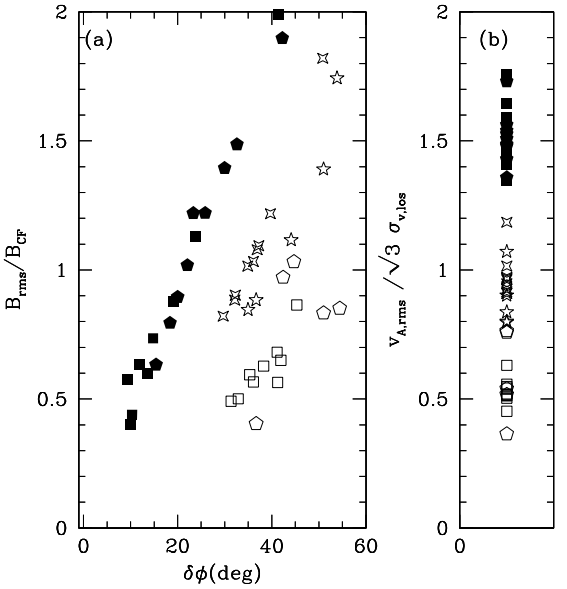


Fig. 7. (a) Total (rms) magnetic field strength relative to C-F estimate B_{CF} vs. dispersion (in degrees) of directions in the polarized-extinction map. (b) Total (rms) Alfvén speed compared to $\sqrt{3}$ times the line-of-sight velocity dispersion. Data and symbols for (a) and (b) are as in Fig. 6.

between the small-$\delta\phi$ and large-$\delta\phi$ limits (with the implicit assumption that $v_{\rm A}$ along the line of sight is not very large) have been provided by [14].

6 Summary

To understand the intrinsic nature of MHD turbulence and the relation between the turbulence observed in astronomical systems and the turbulence simulated numerically, it is crucial to develop structural diagnostics. These diagnostics may:

- Enhance conceptual understanding of the turbulent phenomenon;
- Enable determination of astronomical systems' properties that are either difficult to observe at all (e.g. $\boldsymbol{B}$) or only indirectly observable because of projection effects (e.g. $\boldsymbol{v}_k$, ρ_k);
- Provide a physical basis or interpretation for observed empirical "laws" (e.g. column density distributions, linewidth-size relations, mass functions of clouds, clumps, and stars);
- Indentify when additional physical ingredients may be needed in a computational model.

From the examples outlined in this paper, it is clear that significant advances along these lines have already been accomplished.

Conceptual and (approximate) quantitative understanding of what determines the PDFs of density and column density observed in molecular clouds have already been obtained from analyses to date (see Sect. 2). The development of lognormal statistics from Gaussian random processes under near-isothermal conditions appears to obtain robustly for a variety of conditions, with little sensitivity to the magnetic field strength. Because one-point density statistics are subject to "cosmic variance" (if low wavenumbers dominate the power spectrum, different realizations of a given power spectrum have significant variation), and because degeneracies prevent the direct inversion of column density statistics to obtain volume density statistics, there may not be a highly accurate way to determine the physically-important volume-averaged density in a cloud from more direct observables such as the observed velocity dispersion and observed distribution of extinction. Future work that combines one-point column density statistics with two-point correlation functions could potentially be valuable in constraining the spatial power spectrum of density fluctuations. Such analyses could also be useful in relating mass functions of clumps seen in projection with true three-dimensional clumps (see Sects. 3, 4).

Analysis of clumpy structure in model clouds (see Sect. 3) shows intriguing correspondence to observations: mass functions of self-gravitating structures are comparable to the Salpeter IMF, and there are significant differences between steeper (more "stellar-like") mass functions when subcondensations are not subsumed into larger structures, and shallower (more "cloud-like") mass functions when substructure is discounted. These findings support the concepts that clumping imposed by turbulence, as well as fragmentation during gravitational collapse, may both be important in establishing the IMF. Future work is required to determine whether there might be a relationship between Mach numbers (instantaneous and historical), the overall size and mass of a cloud, and the characteristic "peak" sizes and masses of self-gravitating condensations within a simulation. More coverage of model parameter space (allowing for different power spectra, forcing, etc.) will be important for deciding how sensitive the resulting clump mass functions may be to physical conditions.

Some of the longest-established empirical results about molecular clouds concern the correlations among physical scale and spectral linewidth, and analyses of simulations have proven valuable in interpreting how these empirical "laws" relate to the underlying properties of turbulent clouds and how they are observed (see Sect. 4). Steep ("Larson's law") dependence of linewidth on size probably reflects the true three-dimensional power spectrum, since the structures to which these steep laws apply are observed in tracers that render them well-separated from the background. On the other hand, the weak dependence of mean linewidth on 2D size of apparent moderate-density clumps within clouds may largely arise from projection effects – with the scale sampled by the velocity dispersion on average extending over much of the whole cloud's depth. Both simple methods using the lower envelope of the linewidth-size distribution – and more complicated methods using detailed spectral shapes and their spatial correlations, are

very promising for being able to distinguish the true three-dimensional power spectrum from molecular line data cubes.

Since magnetic fields are difficult to measure directly, there are particularly strong motivations to develop indirect diagnostics (see Sect. 5). Polarization either in absorption or emission is sensitive to the local direction of the magnetic field, and the variation in the local direction of the magnetic field is sensitive to the magnitude of the magnetic field and level of turbulence in a cloud. Thus, one might hope to combine observed measures of variation in the polarization direction with observed measures of turbulence via molecular linewidths to infer the mean magnetic field strength; this kind of indirect method was originally proposed by Chandrasekhar and Fermi. Testing the Chandrasekhar-Fermi method with simulation "data" shows that for low dispersion in the polarization angle, recalibration by a factor one-half indeed yields a good measure of the mean plane-of-sky magnetic field. When the polarization direction has large fluctuations, the Chandrasekhar-Fermi method loses accuracy, but simulations show that an assumption of magnetic/kinetic equipartition is usually correct within a factor ~ 2. Because underresolution tends to enhance the estimated field strength, and because high-column/high-density regions may not be efficient polarizers, there are some potential caveats in applying the Chandrasekhar-Fermi method for polarized-emission data. The promise shown by analyses to date, together with the potential to obtain large-scale maps of polarized absorption and emission in molecular clouds, marks this area as an important direction for further research.

These detailed results are exciting, and represent only a small sample of the progress that has taken place in this field to date. Perhaps the most fundamental advance, however, has been the movement to match our sophisticated concept of what a molecular cloud *is* – a complex structure with multiple-scale, large-amplitude, turbulent fluctuations in all fluid variables; with a commensurately sophisticated way to diagnose structure – by developing and testing analytical tools on detailed simulation data cubes that are self-consistent, time-dependent solutions of the MHD equations. Recent accomplishments have greatly advanced our field, but we are still very much in the era of discovery. Ongoing interplay between simulation, analysis, and observation will be essential for continued progress in building a comprehensive dynamical model of molecular clouds.

Acknowledgements

I am grateful to J. Stone and C. Gammie for permission to present results from collaborative work, and to the referee A. Lazarian for helpful comments. This research is supported by NASA grants NAG 53840 and NAG 59167.

References

1. André, P., Ward-Thompson, D., & Barsony, M. 1999, in Protostars and Planets IV, ed. V. Mannings, A. P. Boss & S. S. Russell (Tucson: University of Arizona Press), p. 59

2. Alves, J., Lada, C.J., Lada, E.A., Kenyon, S.J., & Phelps, R. 1998, ApJ, 506,292
3. Arce, H.G., Goodman, A.A., Bastien, P., Manset, N., & Sumner, M. 1998, ApJ, 499, L93
4. Ballesteros-Paredes, J., & Mac Low, M.-M. 2002, ApJ, 570, 734
5. Bertoldi, F., & McKee, C.F. 1992, ApJ, 395, 140
6. Blitz, L., & Williams, J. P. 1997, ApJ, 488, L145
7. Brunt, C.M., & Heyer, M.H. 2002, ApJ, 566, 276
8. Burkert, A., & Mac Low, M.-M. 2001, astro-ph/0109447
9. Chandrasekhar, S., & Fermi, E. 1953, ApJ, 118, 113
10. Davis, L., & Greenstein, J. 1951, ApJ, 114, 206
11. Gammie, C.F., Lin, Y.-T., Stone, J.M., & Ostriker, E.C., 2001, ApJ, in preparation
12. Goodman, A.A., Jones, T.J., Lada, E.A., & Myers, P.C. 1995, ApJ, 448, 748
13. Heiles, C. 1996, in Polarimetry of the Interstellar Medium, Eds. W.G. Roberge & D.C.B. Whittet (ASP Press:San Francisco), p. 457
14. Heitsch, F., Zweibel, E.G., Mac Low, M.-M., Li, P., & Norman, M.L. 2001, ApJ, 561, 800
15. Henning, Th., Wolf, S., Launhardt, R., & Waters, R. 2001, ApJ, 561, 871
16. Heyer, M., & Schloerb, P. 1997, ApJ, 475, 173
17. Hildebrand, R.H. 1988, QJRAS, 29, 327
18. Klessen, R.S. 2001, ApJ, 556, 837
19. Larson, R. B. 1981, MNRAS, 194, 809
20. Lazarian, A., Goodman, A.A., & Myers, P.C. 1997, ApJ, 490, 273
21. Lazarian, A. 2000, in Cosmic Evolution and Galaxy Formation, Eds. J. Franco, E. Terlevich, O. Lopez-Cruz, & I. Aretxaga (ASP:San Francisco), p. 69 (astro-ph/0003314)
22. Lazarian, A., & Pogosyan, D. 2000, ApJ, 537, 720
23. Lazarian, A., Pogosyan, D., Vázquez-Semadeni, E., & Pichardo, B., 2001, ApJ, 555, 130
24. Lee, H.M, & Draine, B.T. 1985, ApJ, 290, 211
25. Martin, P.G. 1974, ApJ, 187, 461
26. Martin, P.G. 1975, ApJ, 202, 393
27. McKee, C.F. 1999, in The Physics of Star Formation and Early Stellar Evolution, Eds. C. Lada, & N.Kylafis (Dordrecht:Kluwer), p.29
28. Myers, P.C., & Goodman, A.A. 1991, ApJ, 373, 509
29. Nordlund, Å., & Padoan, P. 1999, in Interstellar Turbulence, eds. J. Franco & A. Carramiñana, (Cambridge:CUP), p 218
30. Ostriker, E., Gammie, C.F., & Stone, J.M. 1999, ApJ, 513, 259
31. Ostriker, E.C., Stone, J.M., & Gammie, C.F. 2001, ApJ, 546, 980
32. Padoan, P., Jones, B. T., & Nordlund, Å. P. 1997, ApJ, 474, 730
33. Padoan, P., & Nordlund, Å. 1999, ApJ, 526, 279
34. Padoan, P., & Nordlund, Å. 2002, ApJ, 576, 870
35. Padoan, P., Rosolowsky, E.W., & Goodman, A.A. 2001, ApJ, 547, 862
36. Padoan, P., Goodman, A., Draine, B.T., Juvela, M., Nordlund, Å, & Rögnvaldsson, Ö.E. 2001, ApJ, 559, 1005
37. Padoan, P., Nordlund, Å, Rögnvaldsson, Ö.E., & Goodman, A. 2001, in "From Darkness to light: origin and evolution of young stellar clusters" ASP conference series, eds. Thierry Montmerle and Philippe André, p. 279
38. Pichardo, B., Vázquez-Semadeni, E., Gazol, A., Passot, T., & Ballesteros-Paredes, J. 2000, ApJ, 532, 353
39. Passot, T. & Vázquez-Semadeni, E. 1998, Phys. Rev. E, 58, 4501

40. Rosolowsky, E.W., Goodman, A.A., Wilner, D.J., & Williams, J.P. 1999, ApJ, 524, 887
41. Scalo, J, Vázquez-Semadeni, E., Chappell, D., & Passot, T., 1998, ApJ, 504, 835
42. Scoville, N.Z., & Sanders, D.B. 1987, in Interstellar Processes, ed. D.J. Hollenbach & H.A. Thronson, Jr. (Dordrecht:Reidel), p. 21
43. Solomon, P. M., Rivolo, A. R., Barrett, J., & Yahil, A. 1987, ApJ, 319, 730
44. Stone, J.M., Ostriker, E.C., & Gammie, C.F. 1998, ApJ, 508L, 99
45. Stutzki, J., & Gusten, R. 1990, ApJ, 356, 513
46. Vázquez-Semadeni, E., 1994, ApJ 423, 681
47. Vázquez-Semadeni, E., & García, N., 2001, ApJ, 557, 727
48. Williams, J. P., de Geus, E. J. & Blitz, L. 1994, ApJ, 428, 693
49. Zweibel, E.G. 1996, in Polarimetry of the Interstellar Medium, Eds. W.G. Roberge & D.C.B. Whittet (ASP Press:San Francisco), p. 486

Star Formation and the Initial Mass Function

Åke Nordlund[1] and Paolo Padoan[2]

[1] Astronomical Observatory / NBIfAFG and Theoretical Astrophysics Center
Juliane Maries Vej 30, DK-2100 Copenhagen Ø, Denmark
[2] Jet Propulsion Laboratory, 4800 Oak Grove Drive, MS 169-506
California Institute of Technology, Pasadena, CA 91109-8099, USA

Abstract. Supersonic turbulence fragments the interstellar medium into dense sheets, filaments, cores and large low-density voids, thanks to a complex network of highly radiative shocks. The turbulence is driven on large scales, predominantly by supernovae. While on scales of the order of the galactic disk thickness the magnetic energy is in approximate equipartition with the kinetic energy of the turbulence, on scales of a few pc the turbulent kinetic energy significantly exceeds the magnetic energy.

The scaling properties of supersonic turbulence are well described by a new analytical theory, which allows to predict the structure functions of the density and velocity distributions in star-forming clouds up to very high order.

The distribution of core masses depends primarily on the power spectrum of the turbulent flow, and on the jump conditions for isothermal shocks in a magnetized gas. For the predicted velocity power spectrum index $\beta = 1.74$, consistent with results of numerical experiments of supersonic turbulence as well as with Larson's velocity-size relation, one obtains by scaling arguments a power law mass distribution of dense cores with a slope equal to $3/(4-\beta) = 1.33$, consistent with the slope of the Salpeter stellar initial mass function (IMF). Results from numerical simulations confirm this scaling. The analytical model for the stellar IMF and its numerical estimate show that turbulent fragmentation may explain the origin of brown dwarfs, but only if the critical mass for collapse under dynamical conditions is an order of magnitude smaller than the Jeans' mass from a linear stability analysis.

The main conclusion is that the stellar IMF directly reflects the mass distribution of prestellar cores, due predominantly to the process of turbulent fragmentation.

1 Introduction

Turbulence in the interstellar medium (ISM) of the Milky Way – and more generally turbulence in the discs of other galaxies – is of crucial importance for both the structure and evolution of the galaxy. The importance of turbulence is both direct, through its influence on the pressure equilibrium and stratification, and indirect, through its influence on the star formation process.

The vertical pressure equilibrium and stratification of the ISM is determined by the level of turbulence, together with the temperature distribution of the medium (which is in turn probably tightly coupled to the turbulence), and it is likely that even the distributions of magnetic fields and cosmic ray particles, which also contribute to the pressure and stratification, are integral parts of the same process; it is unlikely that the near equipartition of the energy content of turbulence, magnetic fields, and cosmic ray particles is a mere coincidence.

E. Falgarone and T. Passot (Eds.): LNP 614, pp. 271–298, 2003.
© Springer-Verlag Berlin Heidelberg 2003

It has long been realized that turbulence in the interstellar medium, in particular in the cold, molecular cloud components, is highly supersonic [45,46,47,27,28]. More recently, it has been realized that the supersonic nature of the turbulence is a boon, rather than a nuisance, when trying to understand the properties of the ISM, the cold molecular clouds, and star formation [62,67,25]. It turns out that supersonic turbulence is in many respect similar to ordinary, subsonic turbulence, and that it thus has a number of generic, statistical properties. Much like ordinary turbulence, its decay time is of the order of the dynamical time, even in the MHD-case [53,55,54,83,67]. And much like ordinary turbulence, it is characterized by power law velocity power spectra and structure functions over an inertial range of scales [9,11,12].

An important difference between supersonic and subsonic turbulence is the distribution of density. A supersonic medium is, by definition, highly compressible; on average its gas pressure P_g is small relative to the dynamic pressure ρu^2. As a consequence, a supersonic medium is characterized by a wide distribution of densities. A turbulent and isothermal supersonic medium has a log-normal density probability distribution (PDF) [87,64,73], with a dispersion of linear density proportional to the Mach number [64,58,61]. Cold molecular clouds are indeed approximately isothermal, and are known to have a very intermittent density distribution, consistent with the properties of isothermal supersonic turbulence [63]. Deviations from isothermal conditions are in general of the type where compression leads to even lower temperatures (effective gas gamma less than unity) [78], resulting in a density PDF skewed towards greater probability at high densities. The PDF may be described as a skewed log-normal, with a high density asymptote that formally tends to a power law in the limit $T \to 0$ [78,58].

Effectively then, supersonic turbulence acts to fragment the ISM, causing local density enhancements also over a range of geometrical scales. Molecular clouds themselves represent relatively large scale density enhancements, probably caused by the random convergence of large-scale ISM velocity features [5,6,34]. Inside molecular clouds smaller scale turbulence leads to high contrast local density enhancements in corrugated shocks, intersections of shocks, and in knots at the intersection of filaments. Such small-scale density enhancements are 'up to grabs' by gravity; if their density is sufficiently high, relative to their temperature and the local magnetic field strength, they form pre-stellar cores, and eventually collapse to form stars. The decisive importance of turbulence in this process makes it possible to predict the distribution of masses of the pre-stellar cores, and hence the distribution of new borne stars, the initial mass function (IMF) [69,71].

The process of star formation is indeed crucial to understand. Only by understanding star formation, qualitatively and quantitatively, can we understand galaxy formation. We need to understand evolution effects to answer questions such as "Was star formation different in the Early Universe?". We need to understand environmental effects to answer questions such as "Do other galaxies have different 'Larson laws'?"

We also need to understand star formation to answer questions related to Gamma-Ray Bursts; e.g., "Are Very Massive Stars progenitors of Gamma-Ray Bursts?", and "What environment does the blast wave associated with Gamma-Ray Bursts encounter"?

Finally, we would like to understand star formation as such, because it is a neat problem – one that involves supersonic, selfgravitating MHD turbulence and thus was thought to be enormously difficult. With access to supercomputer modeling the problem has become tractable, and it has turned out *a posteriori* that it is even partly amenable to analytical theory.

In the subsequent sections of this tutorial star formation and turbulence in the interstellar medium is discussed in more detail. Section 2 discusses supernova driving of the ISM, Sect. 3 discusses properties of supersonic turbulence, Sect. 4 summarizes a new theory of supersonic turbulence, while Sect. 5 discusses star formation and the initial mass function. Conclusions are summarized in Sect. 6.

2 Supernova Driving of the Interstellar Medium

With turbulence being of such fundamental importance in determining the structure and star formation efficiency of the interstellar medium it is important to understand what its primary sources are, and what its overall energy budget is.

First, an estimate and lower limit of the energy input needed to sustain interstellar turbulence is given by Kolmogorov's scaling expression for the energy transfer rate in a turbulent cascade [42],

$$\epsilon \sim \rho U^3/L \ , \tag{1}$$

where U and L are velocity amplitudes and length scales, respectively. In Kolmogorov's classical theory this quantity is assumed to be invariant across the inertial range, and for our purposes this is adequate; subsequent enhancements of Kolmogorov's theory [79] and modifications for supersonic conditions [12] would not change the following estimates significantly.

Observationally, the velocity dispersion in the ISM adheres to Larson's scaling law,

$$U \sim 1\,\mathrm{km\,s^{-1}} \left(\frac{L}{\mathrm{pc}}\right)^{\alpha} \ , \tag{2}$$

with $\alpha \approx 0.4$ [45,46,28], which means that an estimate based on (1) only depends very weakly on the scale L on which the estimate is based. On scales $L \sim 1$ kpc, the turbulent velocity dispersion is of the order $U \sim 10\ \mathrm{km\,s^{-1}}$ [45], which leads to the estimate $\epsilon \sim 5\,10^{50}\ \mathrm{erg\,kpc^{-3}\,Myr}$, using an average ISM density $\sim 1.5\,10^{-24}\ \mathrm{g\,cm^{-3}}$ [15].

For comparison, the rate of energy input to the ISM from supernovae is of the order of $10^{53}\ \mathrm{erg\,kpc^{-3}\,Myr}$, based on a rate of one SN per 70 years in a galactic volume spanned by a radius of 15 kpc and a disk thickness of 200 pc [4]. Thus, less than one percent of the average supernova energy input is necessary to sustain the turbulent cascade of energy in the ISM.

Two questions come to mind: 1) Is there at all a turbulent cascade, and 2) is the energy from supernova at all available for feeding such a cascade?

The answer to the first question is definitely affirmative; whatever its source, the observed velocity field at large scales can do nothing but drive a cascade towards smaller scale, since there is no dissipation mechanism that operates at such large scales. As has recently been shown [11,12], it makes little difference whether turbulence is subsonic or supersonic; similar cascades arise in both cases, only details such as power law exponents differ. The predicted scaling of the velocity dispersion with size is consistent with the observed (Larson's law) scaling [45,46,28]. Most of the observed scatter around the expected scaling (e.g., Fig. 9 in [28]) is probably due to cloud-to-cloud variations –observations for a single cloud (Polaris [35,26]) define a remarkably well defined velocity scaling over more than three orders of magnitude in size [60]. A complementary piece of evidence for power law behavior comes from the observed relation between age difference and spatial separation [25].

The answer to the second question is less obvious, but in the end also affirmative. One might think that supernova energy input occurs at small scales, and hence cannot be a source at large scales for the turbulent cascade. However, as has been demonstrated by detailed numerical simulations [43,33,4], supernovae are indeed capable of sustaining a turbulent cascade with velocity dispersions consistent with observed values. The transfer of energy to large scales occurs through the expansion of supernova bubbles and super-bubbles; i.e., via the hot component of the ISM. The hot component coexists with cooler components (or rather a continuous distribution of temperatures), and the expansion of the hot component into channels and chimneys creates kinetic energy on large scales, available for cascading to smaller scales, also in the cooler components.

The numerical models demonstrate that supernova-driving of the interstellar medium is a viable and probably dominating mechanism, at least in the disc of our galaxy. The detailed numerical models are also broadly consistent with analytical and semi-analytical models of supernova feedback and turbulent self-regulation in galactic discs [20,84].

One may ask whether other sources of energy input could be significant. Winds from hot stars is one candidate that may contribute [89,72]; regions that create supernovae of type II are likely to also contain hot, early type stars. Jets from new-borne stars have been mentioned as an energy input candidate, but it is unlikely to be significant on scales above a few pc.

Larson's original paper [45] lists irregularities and asymmetries in the rotation curve at larger scales that fit into a general power law behavior, at $L \sim 2 - 10$ kpc, although with a break at $L \sim 1$ kpc. This could be taken as an indication that such irregularities, stemming for example from large-scale density waves, could also be a source of driving for the turbulent cascade. But the relation could also go the other way; irregularities on scales of several kpc could be the imprint of old super-bubbles, stretched in the direction of rotation by the differential rotation.

In other contexts, such as star-burst galaxies, the balance between contributions to the driving may be different; kinetic energy input from collisions or close interactions between galaxies may be an important energy input channel there, for example.

3 Turbulent Cascade of the Interstellar Medium

When regarded as an isolated phenomenon, molecular clouds have traditionally given rise to concerns about the source of their turbulence, their life times, and about their support against gravity [80,34].

3.1 Molecular Clouds as Part of a Turbulent Cascade

Molecular clouds are known to be significantly supersonic, with observed turbulent velocities of the order $\mathrm{km\,s^{-1}}$, while typical sound speeds at molecular cloud temperatures are $\sim$ 0.2–0.3 $\mathrm{km\,s^{-1}}$. The supersonic velocities were assumed to give rise to very rapid dissipation in shocks, and hence explaining how the observed velocities are sustained was regarded as a problem. A popular suggestion for a solution to the problem was that the observed velocities are essentially magneto-hydrodynamic waves, with very low dissipation rates [3,92].

But in light of the conclusions of the previous section there is really no reason to be concerned about how the turbulence of molecular clouds is sustained; supply of kinetic energy at molecular cloud size scales is available from the turbulent cascade; i.e., simply from larger scale motions. In fact, a molecular cloud is probably borne precisely because the larger scale velocity field happens to have a local maximum of convergence there [5,6,34].

3.2 Supersonic Turbulent Cascades

Figure 1 shows examples of power spectra of supersonic turbulence, for two values of the numerical resolution. The numerical simulations were performed with the same method as in [64,67,70,68]; a fifth order in space, third order in time staggered mesh method, using per-unit-mass variables. The turbulence is driven by a random external force that is applied in $1 \leq k \leq 2$ in Fourier space. Only the solenoidal components of the force are used. In order to ensure a smooth forcing the time derivatives of the Fourier components of the force are regenerated at time intervals of about one dynamical time and the force is computed from a time integral.

An inertial (power law scaling) range is present, and extends to progressively higher wavenumbers at higher numerical resolution. How is it possible that supersonic turbulence gives rise to a turbulent cascade much like that of incompressible turbulence? One clue comes from the dissipation rate of supersonic turbulence. Numerical experiments revealed it to be similar to that of incompressible turbulence, if expressed in terms of dynamical times $\tau_{\mathrm{dyn}} = \ell / v_{\mathrm{rms}}(\ell)$, even for MHD-turbulence [8,53,55,54,67,83]. In qualitative terms one reason why

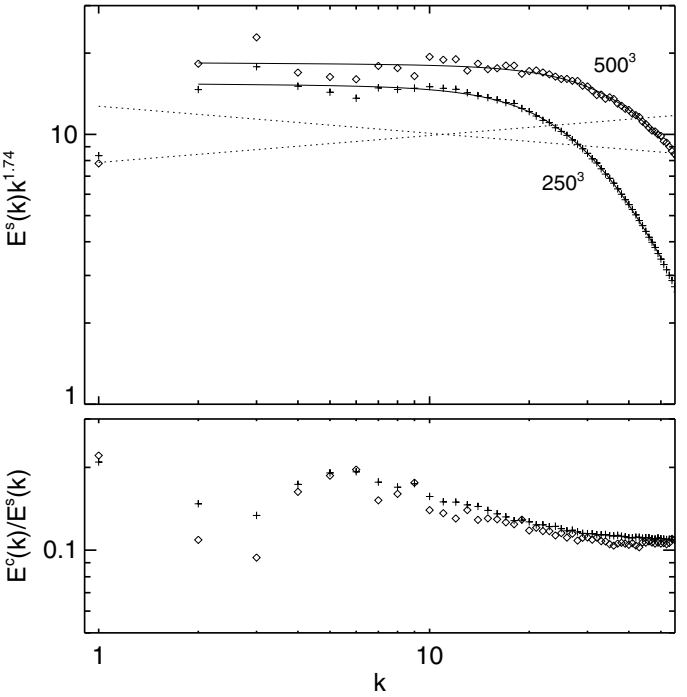


Fig. 1. The solenoidal power spectrum, compensated by $k^{1.74}$, and the ratio of compressional to solenoidal power in $\mathcal{M} \approx 10$ numerical experiments with resolution 250^3 (diamonds) and 500^3 (stars), using random solenoidal driving at $1 \leq k \leq 2$. Dashed lines show comparison slopes with spectral indices -1.64 and -1.84, respectively.

the dissipation rate is not as high as was naively expected is that shocks in three-dimensional supersonic turbulence are typically oblique rather than head-on, and that fragmentation decreases the efficiency of interaction [77,62].

The ratio of compressional to solenoidal kinetic energy is small in isotropic supersonic turbulence; typically $E_c/E_s \sim 0.1-0.2$ [11] (cf. Fig. 1). To appreciate why this is so it is helpful to consider the velocities on either side of a shock sheet formed by two interacting large scale streams. By definition, the gas upstream of the two stand-off shocks on either side of the sheet have no casual connection, and their orientations are therefore random with respect to each other.

In terms of a coordinate system with one axis perpendicular to the plane of the sheet, two of the three velocity components are parallel to the sheet and hence incompressible, and only one is perpendicular (compressional). The compressional component is the one that gives rise to the shock, with its associated stagnant region inside the shock sheet. It follows that one should indeed expect the compressional component to, on the average, carry less than one third of the kinetic energy, consistent with what is found in the numerical experiments.

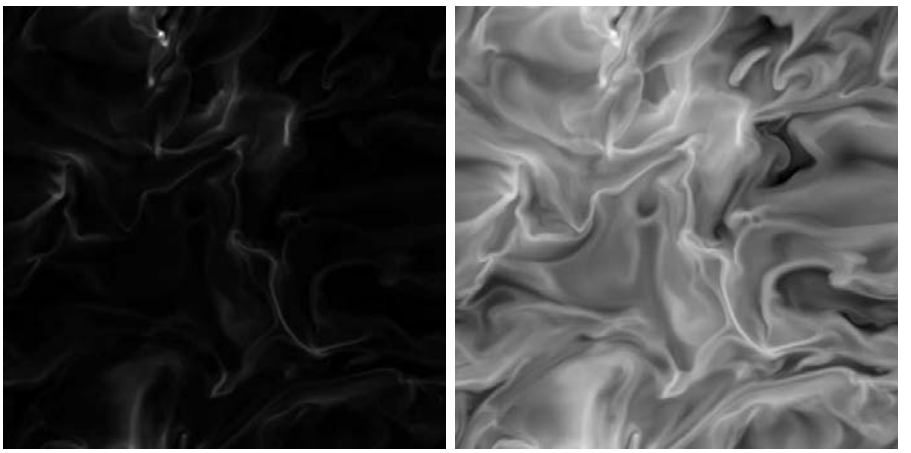


Fig. 2. Density in an arbitrary cut through a numerical experiment with $\mathcal{M} \approx 10$ turbulence at a resolution of 500x500x500. a) Linear scaling (printed with gamma=0.7), normalized to the local maximum of the density. b) Logarithmic scaling, normalized to the local maximum and minimum of the density.

More generally, one may think of three-dimensional supersonic turbulence as an ensemble of shock sheets, and their associated intersections in filaments and knots. The typical history of a trace particle in such a flow is that it participates in a series of oblique shocks where, in each shock, the particle looses some of its kinetic energy.

If the system consisted of an ensemble of stationary, plane-parallel shock sheets a fluid parcel would first hit one sheet, where its perpendicular kinetic energy would be essentially lost. It would then slide along the sheet until it hit the filamentary intersection with another sheet, and finally slide along the filament until it ended up in the stagnant region of a knot-like intersection of filaments.

In a more general picture shock sheets are neither stationary nor plane-parallel, which allows trace particles to participate in a more extended series of shocks. As its kinetic energy is gradually reduced, so is the scale over which new shocks are likely to be produced.

Figure 2 illustrates the structure of the density field in supersonic turbulence, modeled at a resolution of 500^3. The left hand side panel shows linear density. Due to the large density contrast, only a few shock sheets and filaments are visible. The right hand side panel shows logarithmic density, and illustrates the general presence of intermittent density structure over a range of scales and density levels.

The history of a fluid parcel in the real ISM might be as follows: It achieves initial, large kinetic energy by either being part of the ejecta from a supernova or, more likely, by being hit by the ejecta from a supernova. It becomes further compressed as the stream to which it belongs collides with other streams. As density increases cooling becomes significant, and the temperature decreases.

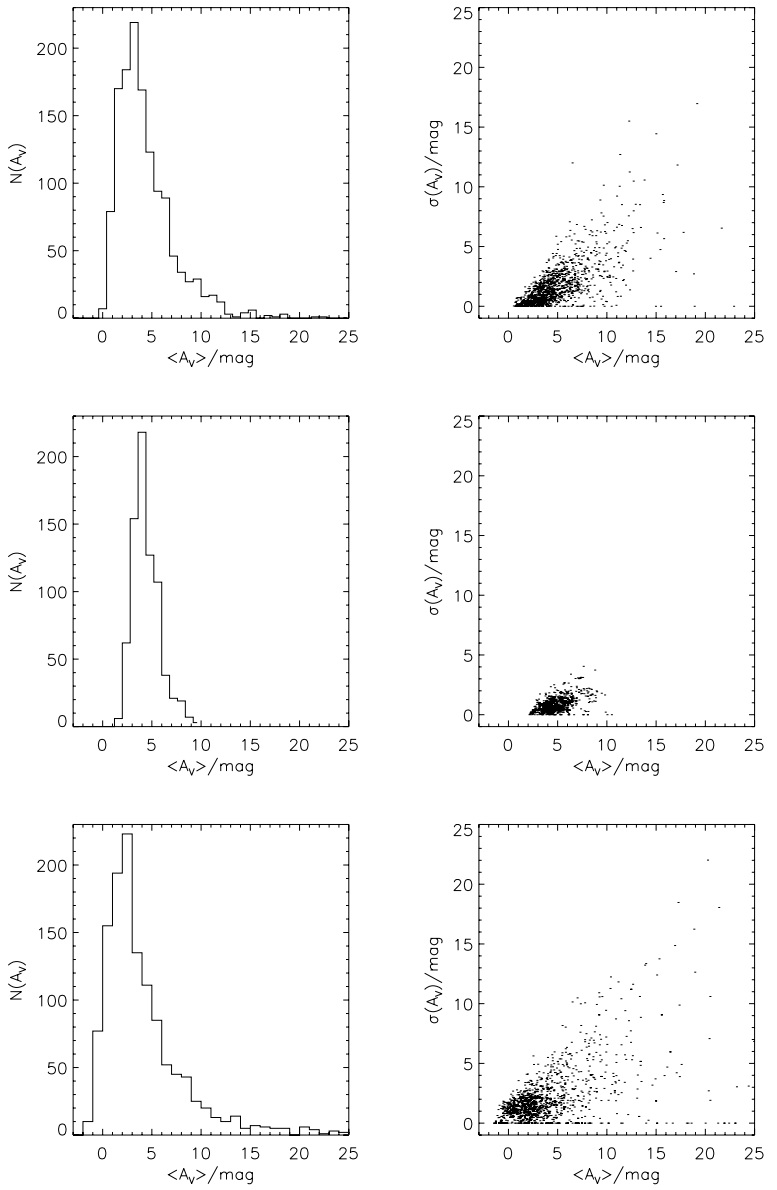


Fig. 3. Histograms of extinction (left panels) and plots of dispersion of extinction in cells versus the mean cell extinction (right panels). The top panels show the result from a super–Alfvénic model while the middle panels are from an equipartition model [67]. The bottom panels are observational data for the cloud IC5146 [44].

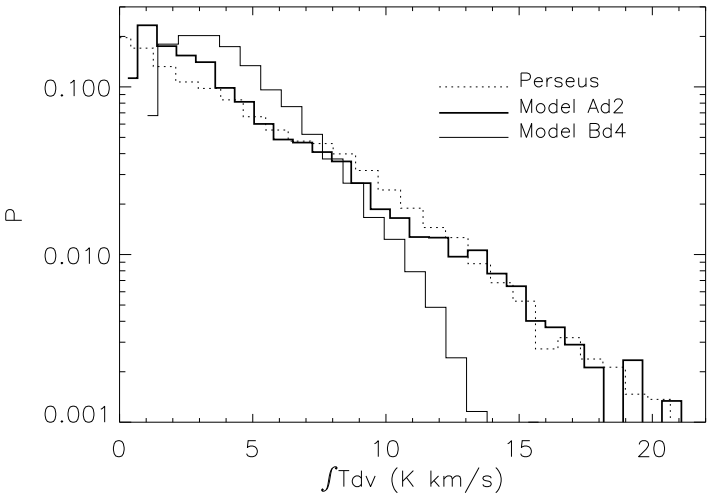


Fig. 4. Histograms of integrated antenna temperature of synthetic CO spectra [67], for a super–Alfvénic model (thick line), for an equipartition model (thin line), and for the Perseus cloud (dotted line).

The parcel may eventually end up as part of a molecular cloud. Inside the cloud, the process repeats itself, on successively smaller scales, creating in the end a shock core massive enough to form a star. More likely, though, the fluid parcel ends up in a structure too small to collapse by self-gravity, where it survives until being hit by the blast wave from another supernova, or the wind from a new-borne massive star.

3.3 Super-Alfvénic Conditions

The proposal that magneto-hydrodynamic waves are main contributors to the velocity field in molecular clouds is now obsolete for several independent reasons. First, with the velocity field of molecular clouds part of a turbulent cascade there is no longer a problem to sustain the motions. Second, with the demonstrations that MHD-turbulence decays more or less as rapidly as hydrodynamic turbulence [8,53,55,54,67,83], the presumed 'advantage' of MHD-turbulence has gone away. Third, with observational and theoretical evidence that star formation takes place on a time scale not much longer than a crossing time [25] the required life times of molecular clouds are much shorter than was assumed in earlier work.

The notion that the velocity field in molecular clouds is essentially Alfvén waves lead to the assumption that the kinetic and magnetic energy in molecular clouds are in near equipartition. Although inverting observations to obtain the magnetic field strength, density and velocity in the same structures is notoriously difficult (cf. the discussion in Sect. 4.1 of [31]), equipartition remained a popular null-hypothesis.

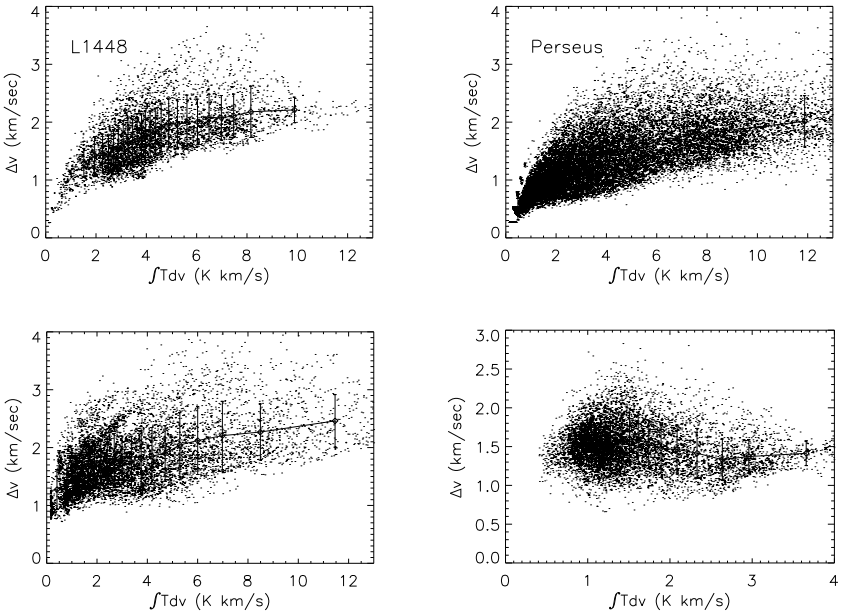


Fig. 5. Scatter plots of J=1-0 ^{13}CO line width versus velocity integrated antenna temperature for two observed regions (two upper panels), for a super-Alfvenic model (lower left), and for an equipartition model (lower right) [37,67,66]. The diamond symbols show the mean value of the equivalent width in each interval of integrated antenna temperature and the "error bars" show the one σ distribution around the mean.

With access to numerical simulations it is possible to use the more robust 'forward analysis' method, where synthetic diagnostics computed from the results of numerical simulations are compared directly with the corresponding observational diagnostics. Comparisons of extinction statistics, synthetic molecular lines, the antenna temperature – line width relation, and the statistical upper envelope relation between density and magnetic field strength (cf. Figs. 3–6) all lead to the same conclusion; models with equipartition between magnetic and kinetic energy are inconsistent with the observations while models where the kinetic energy dominates over the magnetic energy (super-Alfvénic models) are consistent with the observations [63,65,66,67].

A direct illustration of the consistency of super-Alfvénic conditions with Zeeman observations of magnetic field strength is given in Fig. 7. Note that several cores with B in excess of 100 μG are found, even though the average B in the simulation is only 2.4 μG. This is a good example of the power of forward comparisons in situations with strong intermittency; it would be very difficult to recover the mean field strength, or the mean magnetic energy, directly from the observations, which sample only the very small fraction of the cloud volume filled by the densest regions. Further illustrations are given in Figs. 8–9, which show comparisons of velocity statistics with observations.

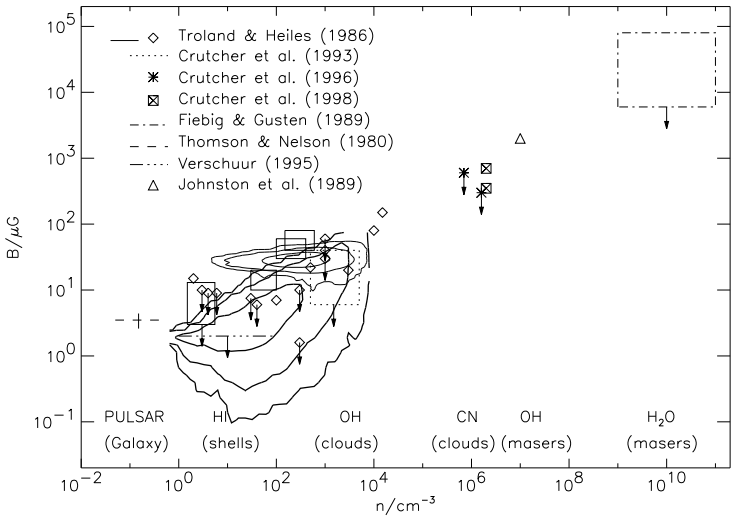


Fig. 6. The $B - n$ relation: observations and theoretical models. The thick contour lines are from the a super–Alfvénic model and the thin contour lines are from an equipartition model [67].

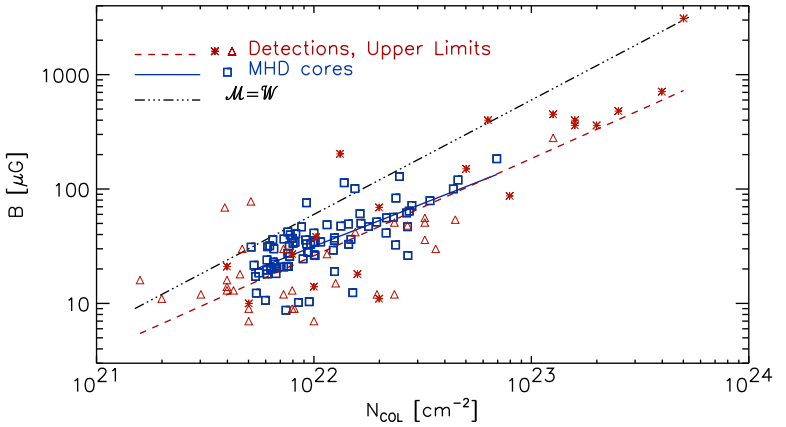


Fig. 7. Magnetic field strength versus H_2 column density. Asterisks represent Zeeman splitting measurements [16] (detections and upper limits) with a least squares fit (dashed line). Squares represent cores from a super-Alfvénic numerical experiment (average B = 2.4 μG) [70], with a least squares fit (solid line). The dotted-dashed line marks equality between magnetic and gravitational energies.

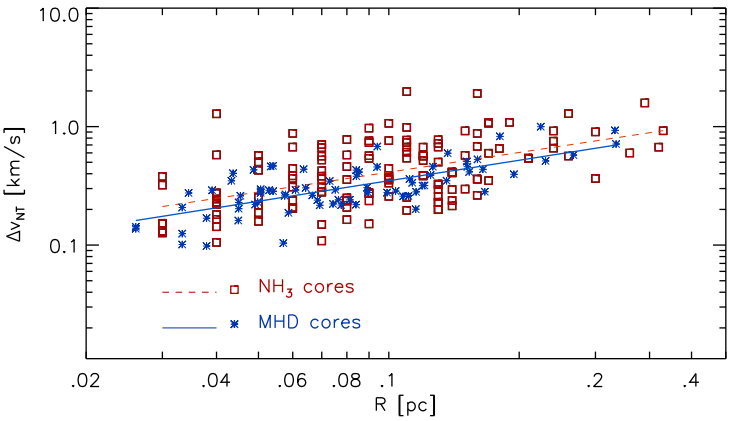


Fig. 8. Non-thermal line width of observed (squares, with dashed line least squares fit) and modeled (stars, with solid line least squares fit) NH_3 cores versus their size [40,70].

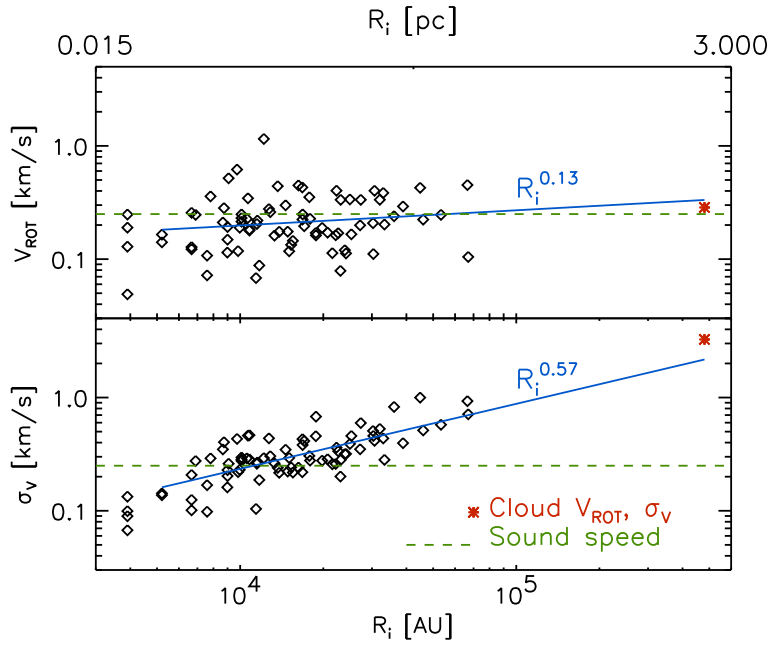


Fig. 9. Rotational velocity (upper panel) and velocity dispersion (lower panel) for numerical cores.

Figure 8 is a comparison of the correlation of non-thermal line width with size in NH_3 cores from the compilation by Jijina, Myers & Adams [40] and in cores selected from a simulation of supersonic and super-Alfvénic turbulence [70].

The least squares fit to the observational data yields the power law exponent 0.56 ± 0.22 and the fit to the data the exponent 0.57 ± 0.15. Figure 9 shows the correlation of rotational velocity (upper panel) and internal velocity dispersion (lower panel) with size, for the numerical cores only. The rotational velocities are very low and of the order of the sound speed, as found in the observational data. In both panels of Fig. 9 the asterisk symbols that correspond to a size of almost 3 pc provide the values of rotational velocity and velocity dispersion computed over the whole simulated volume. Although the least squares fits are computed only for the cores, the values for the whole system are consistent with the fits.

3.4 The Magnetic Flux Problem

A "magnetic flux problem" is often mentioned in this context. It is argued that since the average density of molecular clouds is at least 100 times larger than the average density in the galactic disk, and assuming magnetic flux freezing, the average magnetic field strength of molecular clouds should be much larger than the average galactic values of a few μG. As demonstrated by Fig. 7 there is in fact no real problem – observations of core magnetic fields are completely consistent with predictions from models with average magnetic field strengths of a few μG – there is at most a conceptual / perceived problem.

This conceptual problem has a straightforward solution, which, ironically, is most easily demonstrated by the equipartition model. It has been shown in many numerical works that supersonic turbulence, even with equipartition of kinetic and magnetic energy (the traditional model for molecular clouds) generates a complex density field, with very large contrast sheetlike and filamentary density structures. These density enhancements do *not* correspond to significant variations of the magnetic field strength in equipartition models, since in them strong compression can occur only along magnetic field lines. To the extent that turbulence on large scales (disk thickness) has approximate equipartition of kinetic and magnetic energy, molecular clouds can still easily form, as a consequence of compressions *along* magnetic field lines.

Once a cloud is formed by large scale equipartition turbulence, it has a mean magnetic field strength close to the galactic value, and its internal dynamics is super-Alfvénic, because of the much increased density. Equipartition on the large scale, therefore, is not a problem for the origin of super-Alfvénic clouds.

The argument applies recursively; should the super-Alfvénic cloud by chance create a region with local equipartition, further increase of the density is still possible, by inflow of mass along magnetic field lines. One sees the statistical outcome of this in the B-n relation (Fig. 6); for any given density there is a wide distribution of magnetic field strengths, up to an upper envelope given by approximate equipartition.

Inflow along magnetic field lines is also likely to occur in the phase when gravitation has taken over, after local cores are formed along filaments and in corrugated sheets. In that situation the magnetic field has already been com-

pressed, and is oriented predominantly along the same filaments and sheets that also contain abundant mass at high density.

3.5 Gravitationally Bound and Unbound Clouds

The scenario where star formation takes place in essentially a crossing time [25] also alleviates earlier concerns about how to support molecular clouds against gravitational collapse. The gravitational binding energy of molecular clouds is often comparable to their turbulent kinetic energy, and hence quite a bit larger than their thermal energy [46,28]. This raised the question of the support of the clouds against gravitational collapse. Could the turbulent velocities be translated into a turbulent pressure that was able to support the clouds against collapse [13,14,88,41]? Or did the solution lie in the observed, strong fragmentation of the medium [77,62]?

In the 'turbulent fragmentation & star formation in a crossing time' paradigm it is natural to find some molecular clouds with roughly virial mass, as well as some with substantially less than virial mass, while there are essentially none with much larger than virial mass (cf. Fig. 7 of [28]). Turbulent fragmentation creates clouds, initially without regard to gravity. Some of the clouds that are produced are gravitationally unbound (but may contain sub-structures that are gravitationally bound). Some other clouds are massive enough to be gravitationally bound (at least until their first supernovae blow out a major part of the cloud gas). Clouds that are created with a mass larger than virial start to collapse, which increases the velocity dispersion of their sub-structures until they appear to be essentially virial.

The latter case represents the most direct and simple mechanism by which turbulence prevents global collapse of molecular clouds; i.e., through fragmentation rather than through "turbulent pressure". The mechanism may be illustrated by considering that extreme intermittency caused by strongly supersonic turbulence and cooling could create conditions where individual density maxima move in essentially ballistic orbits relative to one another [77]. Even under less extreme conditions intermittency may cause individual density maxima to collapse, while the cloud as such does not [62,64,41,36].

3.6 Power Laws and Equipartition

As mentioned above, even highly supersonic turbulence is characterized by power laws [11,12]. However, because of the strong intermittency of density and its correlation with the velocity field, the spectrum of kinetic energy is not the same as the power spectrum of velocity.

It is appropriate to define the spectrum of kinetic energy as the power spectrum of $\rho^{1/2}\boldsymbol{u}$, since the sum of squares of its Fourier components is equal to the kinetic energy. Empirically, from numerical simulations, one finds that the spectrum of kinetic energy is quite a bit more shallow than the power spectrum of velocity. The latter has a power exponent consistent with the theoretical expectation $\beta = 1.74$ [11]. The former has a power exponent $\beta_{\rm k.e.} \approx 1.1$.

The power spectrum of the magnetic field is approximately parallel to that of velocity in the high-k part of the inertial range, and hence the spectrum of magnetic energy is *steeper* than the spectrum of kinetic energy. This may appear strange, at first. Why would the magnetic field have a power spectrum similar to that of velocity, when magnetic energy, $\frac{1}{2}B^2$, is measured in the same units as kinetic energy $\frac{1}{2}\rho u^2$ and not in the units of velocity power u^2? A possible explanation is that B^2 is weighted more towards the kinetic energy of the bulk of the volume. Assuming a log-normal PDF of density, with a dispersion of linear density [58]

$$\sigma_\rho = b\mathcal{M} \ , \tag{3}$$

where $b \approx \frac{1}{2}$, the most common density is

$$\rho_0 = \frac{\langle\rho\rangle}{(1+b^2\mathcal{M}^2)^{\frac{1}{2}}} \ , \tag{4}$$

which is smaller than the average density $\langle\rho\rangle$. Thus, if the magnetic energy is in equipartition with the kinetic energy at those densities, rather than at the higher densities towards which the average kinetic energy is weighted, this would explain both why the spectrum of $\boldsymbol{B}$ is similar to that of $\boldsymbol{u}$ and why the average magnetic energy is below equipartition with the kinetic energy.

With a difference in the power law exponents a gap develops from the (observed) equipartition at large scales (~ 100 pc). A power law index difference of 0.65 implies $E_{mag}/E_{kin} < 0.1$ at pc scales, which is consistent with the forward analysis of numerical simulations [67].

Note that the discussion above is complementary to the one at the end of Sect. 3.3 – both views are helpful for understanding why small-scale ISM motions are super-Alfvénic.

4 A New Analytical Theory of Supersonic Turbulence

Due to the complexity of the Navier-Stokes equations, mathematical work on turbulence is often inspired by experimental and observational measurements. Since geophysical and laboratory flows are predominantly incompressible, turbulence studies have been limited almost entirely to incompressible flows (or to infinitely compressible ones, described by the Burgers equation). Little attention has been paid in the past to highly compressible, or super-sonic turbulence.

Turbulent flows are traditionally described statistically by the structure functions of their velocity field [29]. The structure functions are defined as

$$S_p(\ell) = \langle|\boldsymbol{u}(\boldsymbol{x}+\boldsymbol{\ell}) - \boldsymbol{u}(\boldsymbol{x}))|^p\rangle \propto l^{\zeta(p)} \ , \tag{5}$$

where $\boldsymbol{u}$ is the component of the velocity field perpendicular (transversal structure functions) or parallel (longitudinal structure functions) to the vector $\boldsymbol{\ell}$. In the inertial interval the structure functions obey scaling laws and the exponent $\zeta(p)$ can be determined. The power spectrum of the velocity is the

Fourier transform of the second order structure function, and may be expressed as $E(k) \propto k^{-\beta} \propto k^{-1-\zeta(2)}$.

One may think that the study of high-order structure functions is interesting only for testing models of intermittency in turbulent flows, and not very useful in the context of ISM turbulence and star formation. Actually, the intermittent nature of turbulence is crucial in modeling the process of star formation driven by turbulent fragmentation. Stars are formed in the densest regions of turbulent flows. These regions contain only a few percent of the total mass and fill an almost insignificant fraction of the total volume of a star forming cloud. High order moments defining the tails of statistical distributions of velocity and density are therefore very important in the process of star formation. Furthermore, low-order density structure functions, which are obviously important to describe basic properties of turbulent fragmentation, can be shown to depend on velocity structure functions of very high order [12].

The scaling of the velocity structure functions in incompressible turbulence is best described by the She-Leveque formula [79],

$$\frac{\zeta(p)}{\zeta(3)} = p/9 + 2\left[1 - \left(\frac{2}{3}\right)^{p/3}\right] \ . \tag{6}$$

The scaling exponents are computed relative to the third order, $\zeta(p)/\zeta(3)$, because according to the concept of extended self-similarity [7,19] the relative exponents are universal and better defined than the absolute ones.

Boldyrev [9] has proposed an extension of the She-Leveque's formalism [79] to the case of supersonic turbulence. Based on the physical interpretation of (6) by Dubrulle [19], a fundamental parameter in the derivation of the velocity structure functions is the Hausdorff dimension of the support of the most singular dissipative structures in the turbulent flow. In incompressible turbulence the most dissipative structures are organized in filaments along coherent vortex tubes with Hausdorff dimension $D = 1$, while in supersonic turbulence dissipation occurs predominantly in sheet-like shocks, with Hausdorff dimension $D = 2$. The new velocity structure function scaling proposed by Boldyrev [9] for supersonic turbulence is

$$\frac{\zeta(p)}{\zeta(3)} = p/9 + 1 - \left(\frac{1}{3}\right)^{p/3} \ . \tag{7}$$

This velocity scaling has been found to provide a very accurate prediction for numerical simulations of supersonic and super-Alfvénic turbulence [11], and has been used to infer the structure of the density distribution in turbulent clouds [12].

5 Star Formation and the Initial Mass Function

At least three unrelated ways of explaining the process of star formation and the origin of the stellar initial mass function (IMF) may be found in the literature: i) Ambipolar drift contraction of sub-critical cores [80,1]; ii) opacity-limited

gravitational fragmentation [39,30,90,85,48,74,82,81,91]; and iii) turbulent fragmentation [2,46,77,21,87,62,68,69].

The first type of models rely on the assumption that both protostellar cores and their parent clouds are long-lived systems in near equilibrium, supported against their gravitational collapse by magnetic field pressure. As discussed above, this assumption has been proven incorrect based on observational data and is inconsistent with the turbulent nature of star-forming clouds [67,25,34,68]. Furthermore, these type of models do not address the problem of the formation of massive stars or brown dwarfs, and have traditionally focused more on the evolution of individual protostars, without providing a self-consistent picture for the origin of the initial conditions.

The second type of models is also inconsistent with the properties of star-forming clouds, because it applies the concept of gravitational fragmentation to the large scale, in the attempt of modeling the formation of a whole stellar population. The concept of gravitational instability is based on a comparison between the gas thermal and gravitational energies to define the smallest unstable mass, or Jeans' mass. However, star-forming clouds, as any region of the cold ISM above a scale of approximately 0.1 pc, contain a kinetic energy of turbulence that is much larger (typically 100 times larger) than their thermal energy, making the comparison of thermal and gravitational energies irrelevant on the large scale. Attempts to redefine the Jeans' mass [18,2] assuming that turbulence can provide pressure support against the gravitational collapse are flawed, because they miss the basic point that supersonic turbulence is actually fragmenting the gas. The main effect of the large kinetic energy of turbulence, relative to the thermal energy, is that the gas density and velocity fields in star-forming regions are highly non-linear, against the assumption of the gravitational instability model. In other words, clouds are already fragmented by turbulence, quite independent of their self-gravity.

The third type of models, which we refer to as *turbulent fragmentation* models, focus on the importance of the observed supersonic turbulence in molecular clouds and are therefore consistent with the large scale dynamics of star-forming regions. The idea of star formation driven by supersonic turbulence was proposed twenty years ago by Larson [46], but has become popular only in the last few years, thanks to the progress of numerical simulations of supersonic magneto-hydrodynamic (MHD) turbulence.

According to the model of turbulent fragmentation, protostellar cores are formed from gas compressed by shocks in the supersonic turbulent flow [68]. While scale-free turbulence generates a power law mass distribution down to very small masses, only cores with a gravitational binding energy in excess of their magnetic and thermal energy can collapse. The shape of the stellar IMF is then a power law for large masses, since the majority of large cores are larger than their Jeans' mass. At smaller masses, the IMF flattens and then turns around according to the probability of small cores to be dense enough to collapse, which is determined by the PDF of gas density.

5.1 The Initial Mass Function

The mass distribution of dense cores formed in a supersonic turbulent flow follows from a number of properties of such flows [69]: i) The power spectrum of the turbulence is a power law; ii) the dynamics on scales covered by the power law is approximately selfsimilar; iii) the typical size of a dense core scales as the thickness of the postshock gas; iv) the relevant shock jump conditions are those of MHD shocks.

These properties are to a considerable extent already verified by numerical simulations, which also produce corresponding numerical IMFs [70], but for the purpose of deriving a theoretical IMF one may just adopt them as assumptions [69].

The second assumption, about approximate selfsimilarity, is a crucial one. Since velocity amplitudes do depend on scale (by the first assumption and by Larson's relation) it is not an exact property. Nevertheless, since these flows are supersonic they follow essentially inertial paths in a large fraction of space (upstream of shocks). The thickness and density of the downstream, shocked gas does depend on the Mach number, and hence on the scale, but the filling factor of the shocked gas is quite small and does not disturb the overall selfsimilarity much.

The distribution of cores that form in the shocked, downstream gas may be regarded as a distribution over linear sizes L of the upstream flows out of which they formed. By the assumption of approximate selfsimilarity the number of such regions per unit $\log L$ scales as L^{-3} (e.g., under this assumption scales ten times as small are a thousand times more abundant). The upstream (Alfvénic) Mach number is denoted $\mathcal{M}(L)$ and is assumed to scale as L^{α}, where α is related to the power spectrum index $\beta = 1 + 2\alpha$.

The typical mass of the cores that form in the shocked gas scales as $\lambda^3\rho_1$, where (from the MHD shock jump conditions) $\lambda \sim L/\mathcal{M}(L)$ is the thickness of the postshock gas, $\rho_1 \sim \rho_0\mathcal{M}(L)$ is its density, where ρ_0 is the upstream mass density (similar to the mean density). The typical core mass is thus

$$m(L) \sim \rho_0 L^3/\mathcal{M}(L)^2 \sim L^{3-2\alpha} \ , \tag{8}$$

which leads to the following expression for the mass distribution of dense cores:

$$N(m) \sim L(m)^3 \sim m^{-3/(3-2\alpha)} \sim m^{-3/(4-\beta)} \ . \tag{9}$$

If the power spectral index β is consistent with the observed velocity dispersion-size Larson relation [46] and with the numerical and analytical results [9,11], then $\beta \approx 1.74$ and the mass distribution is

$$N(m)\,\mathrm{d}\log m \propto m^{-1.33}\mathrm{d}\log m \ , \tag{10}$$

which is almost identical to the Salpeter stellar IMF [76].

If L_0 is defined as the scale of a molecular cloud, with average mass density ρ_0 and Alfvénic Mach number $\mathcal{M}_0$, the mass of the largest cores formed by

turbulent fragmentation is estimated to be

$$m_{max} \approx \frac{\rho_0 L_0^3}{\mathcal{M}_0^2}. \tag{11}$$

In MCs with mass $M_0 \approx \rho_0 L_0^3 \approx 10^4$ $M_\odot$ and Mach number $\mathcal{M}_0 \approx 10$, $m_{max} \approx 100$ $M_\odot$.

While the majority of massive cores are larger than their Jeans' mass, m_J, the probability that small cores are dense enough to collapse is determined by the PDF of the density of the cores, which is approximately Log-Normal. Even small cores have a finite chance to be dense enough to collapse. If $p(m_J)\, dm_J$ is the Jeans' mass distribution obtained from the PDF of gas density [64], the fraction of cores of mass m with gravitational energy in excess of their thermal energy is given by the integral of $p(m_J)$ from 0 to m. The mass distribution of collapsing cores is therefore

$$N(m)\, \mathrm{d}\log m \propto m^{-3/(4-\beta)} \left[\int_0^m p(m_J) \mathrm{d}m_J\right] \mathrm{d}\log m\ . \tag{12}$$

Due to a scaling mistake the Jeans' mass given in [69] (equation 21 of that paper) is smaller than the classical formula by about an order of magnitude. Nevertheless, the position of the IMF maximum is approximately correctly predicted. This leaves two possibilities: Either 1) the low-mass IMF is due to other processes, such as gravitational fragmentation or ejection of stellar embryos from multiple systems [75], or 2) the classical Jeans' mass expression overestimates the mass needed for collapse in a dynamic medium.

The classical Jeans mass arises from considering weak perturbations of a homogeneous medium, which is a situation very different from the one under which the turbulent clumps form. The context of the Bonnor-Ebert stability analysis, with a truncated spherically symmetric clump, confined by a surrounding constant pressure medium, is more akin to the turbulent case, although it is an idealized equilibrium situation, with no external velocity field. The estimate of the critical mass based on the Bonnor-Ebert analysis is about a factor of three smaller than the classical Jeans' mass, while its scaling behavior is the same as for the Jeans' mass.

What we can see in our numerical experiments with both turbulence and selfgravity is that even smaller clumps, similar to the ones that were expected to collapse from the incorrectly normalized Jeans mass criterion actually do collapse, under the influence of external velocity fields. For the purpose of the discussion we therefore explore the consequences of *assuming* that the effective critical mass under dynamic conditions is an order of magnitude smaller than the classical Jeans' mass.

The resulting mass distribution is plotted in Fig. 10, for $\beta = 1.8$. In the top panel the mass distribution is shown for three different values of the largest turbulent scale L_0, assuming Larson type relations [46] to rescale the average gas density, $\langle n \rangle$, and the rms Mach number, $\mathcal{M}$, as a function of size, L_0. The mass distribution is a power law, determined by the power spectrum of turbulence, for

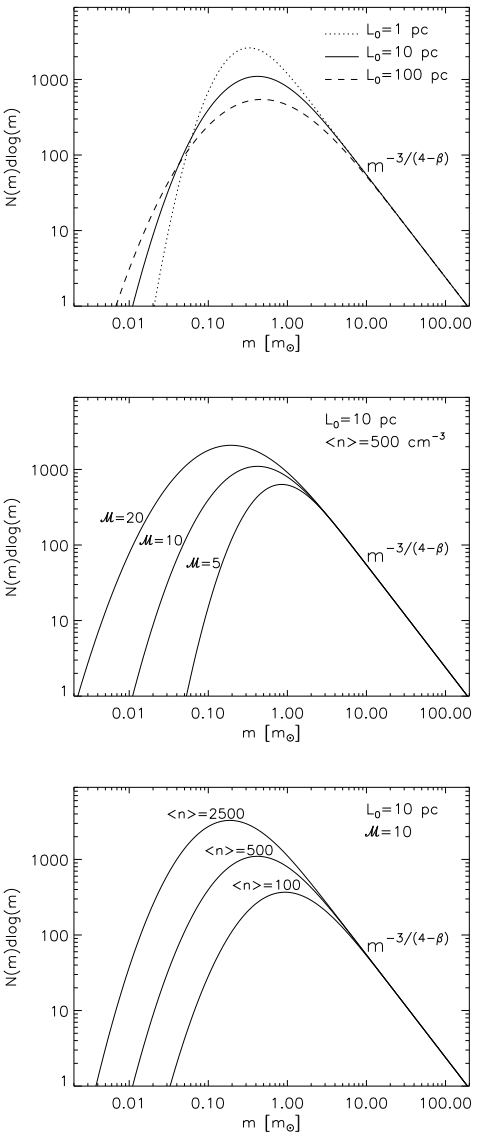


Fig. 10. Mass distribution of gravitationally unstable cores from (12). Top panel: Mass distribution as a function of the largest turbulent scale L_0, assuming Larson type relations (for rescaling $\langle n\rangle$ and $\mathcal{M}$ with L_0), $T_0 = 10$ K and $\beta = 1.8$. Middle panel: Mass distribution as a function of the rms Mach number of the flow, assuming $\langle n\rangle = 500$ cm^{-3}, $T_0 = 10$ K and $\beta = 1.8$. Bottom panel: Mass distribution as a function of $\langle n\rangle$, assuming $\mathcal{M} = 10$, $T_0 = 10$ K and $\beta = 1.8$.

masses larger than approximately 1 $m_\odot$. At smaller masses the mass distribution flattens, reaches a maximum at a fraction of a solar mass, and then decreases with decreasing stellar mass. The mass distribution peaks at approximately 0.4 $m_\odot$ for the values $\mathcal{M} = 10$, $\langle n \rangle = 500$ cm^{-3}, $T_0 = 10$ K and $\beta = 1.8$, typical of nearby molecular clouds. Collapsing sub-stellar masses are found, thanks to the intermittent density distribution in the turbulent flow. This may provide a natural explanation for the origin of brown dwarfs, but only if the critical mass for collapse under dynamical conditions is indeed an order of magnitude smaller than the formal Jeans' mass.

Note that the power law shape of the IMF for mass values larger than about 1 $m_\odot$ is not affected by the average physical properties of the system. On the other hand the abundance of brown dwarfs is very sensitive to the average gas density and the rms Mach number of the flow. The middle and bottom panels of Fig. 10 show the dependence of the mass distribution on the rms Mach number of the flow and on the average gas density respectively. One can see in the middle panel that for an average gas density of $\langle n \rangle = 500$ cm^{-3} and an rms Mach number $\mathcal{M} = 5$, typical of a molecular cloud complex such as Taurus, brown dwarfs are very rare, while for the same average gas density and an rms Mach number $\mathcal{M} = 10$, typical of a molecular cloud complex such as Orion (the density may be even larger), brown dwarfs would be very abundant (even more abundant if the IMF were plotted in units of linear mass interval).

The thermal Jeans' mass is a more strict condition for collapse than the magnetic critical mass. The magnetic critical mass depends on the core morphology in relation to the magnetic field geometry and strength. The latter correlates with the gas density with a very large scatter [67]. It is possible therefore that magnetic pressure support against the gravitational collapse limits the efficiency of star formation, while its effect on the shape of the mass distribution is of secondary importance.

Observations show that the stellar IMF is a power law above 1–2 $m_\odot$, with exponent around the Salpeter value $x = 1.35$, roughly independent of environment [22,24], gradually flattens at smaller masses, and peaks at approximately 0.2–0.6 $m_\odot$ [38,17,51,49,52,50]. The shape of the IMF below 1–2 $m_\odot$, and particularly the relative abundance of brown dwarfs, may depend on the physical environment [50]. Our theoretical IMF is consistent with all of these observational results.

It has been argued that only a small fraction of the mass of each collapsing core may end up into the final star, due to mass loss in protostellar winds, with a major effect on the stellar IMF. However, stellar winds could be important for the origin of the stellar IMF only if the ratio of initial core mass to final stellar mass were comparable to the total mass range for stars ($\sim 10^4$, from ~ 100 $M_\odot$ to ~ 0.01 $M_\odot$), as pointed out by Elmegreen [23]. This is highly unlikely, because i) the correct slope and mass range of the IMF is already achieved by turbulent fragmentation alone and ii) observational results indicate that the mass distribution of prestellar cores is indistinguishable from the stellar

IMF [56,57,86,59], as predicted in earlier work on turbulent fragmentation and the origin of the stellar IMF [62].

5.2 Mass Distribution of Prestellar Cores in Numerical Simulations

The mass distribution of prestellar cores may be measured directly in numerical simulations of supersonic turbulence. With a mesh of 250^3 computational cells, and assuming a size of the simulated region of a few pc, it is not possible to follow numerically the gravitational collapse of individual protostellar cores. However, dense cores at the verge of collapse can be selected in numerical simulations by an appropriate clumpfind algorithm. Such an algorithm should scan all density levels and recognize when a large core is fragmented into smaller and denser ones, in which case the large core should not be counted.

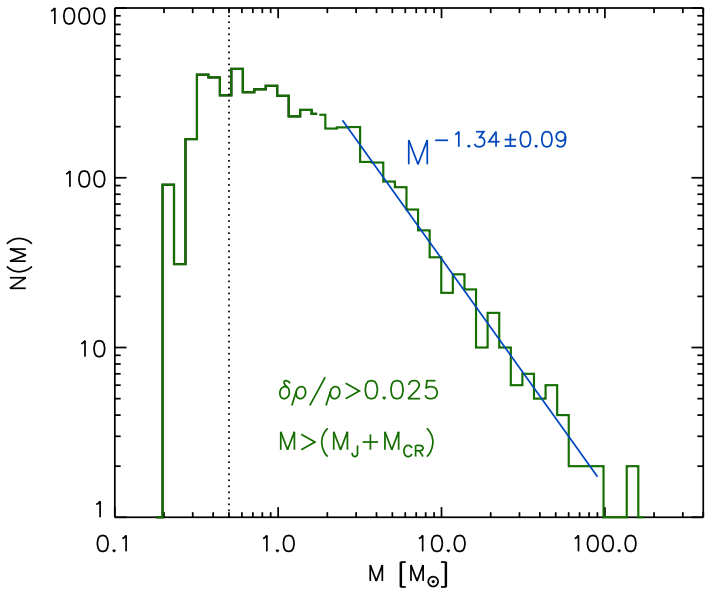


Fig. 11. Mass distribution of collapsing cores in the range 0.2–100 $M_\odot$, derived from a 128^3 numerical simulation.

A mass distribution of collapsing cores, derived from the density distribution in a numerical simulation is shown in Fig. 11. The computational box with 128^3 cells has been scaled to two scale ranges, suitable for sampling cores in the intervals 0.2–2 $M_\odot$ and 2–100 $M_\odot$, respectively. The mass distribution above 1 $M_\odot$ is a power law consistent with our analytical result and with the observations. Below 1 $M_\odot$ the histogram flattens and then turns around at approximately 0.3 $M_\odot$, also consistent with the analytical theory and the observations. The

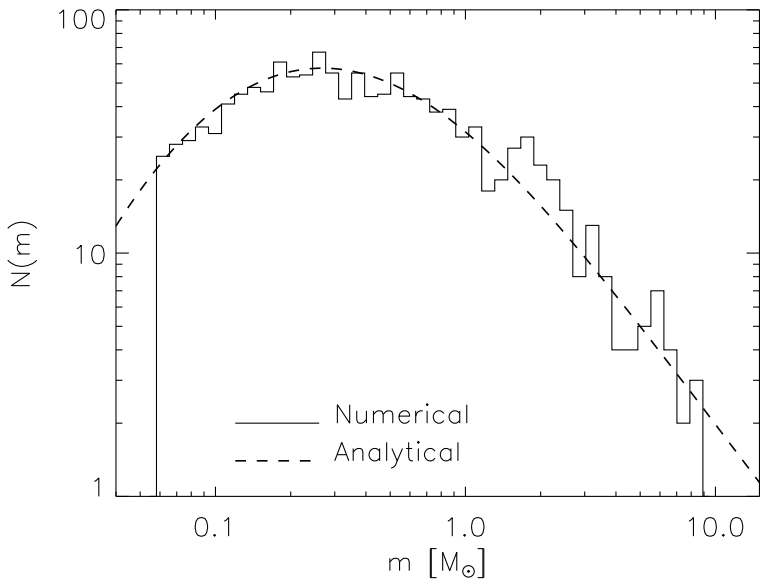


Fig. 12. Solid line: Mass distribution of collapsing cores, derived from the density distribution of two snapshots of a 250^3 simulation with rms Mach number $\mathcal{M} \approx 10$. The simulation is scaled to physical units assuming $\langle n \rangle = 500$ cm^{-3}, $T_0 = 10$ K, and a size of 10 pc. The fractional mass in collapsing cores is 5% of the total mass. Dashed line: Analytical mass distribution, $N(m)$, computed for $\langle n \rangle = 500$ cm^{-3}, $T_0 = 10$ K, $\beta = 1.8$, and assuming that the critical mass for collapse under dynamic conditions is an order of magnitude smaller than the classical Jeans' mass.

cut-off at ~ 0.2 M$_\odot$ is due to the finite numerical resolution; the grid size, rms Mach number, and mean density together impose a limitation on the mass of collapsing cores.

Stretching the mass interval of sampled cores further into the brown dwarf regime requires larger numerical resolution. Figure 12 shows the mass distribution of collapsing cores derived from two snapshots of a 250^3 simulation. The average gas density has been scaled to 500 cm^{-3} and the size of the computational box to 10 pc. These values have been chosen to be able to select cores in a range of masses from a sub-stellar mass to approximately 10 M$_\odot$. With this particular values of average gas density and size of the computational box, the smallest mass that can be achieved numerically is 0.057 M$_\odot$. Brown dwarfs masses (< 0.08 M$_\odot$) are therefore included. With an even larger numerical mesh, or assuming a larger average density (and a smaller size), even smaller masses would be selected. Turbulent fragmentation may thus provide a natural explanation for the origin of brown dwarfs.

Observed star-forming clouds appear very filamentary, as the projected density field of supersonic turbulent flows. We have performed accurate comparisons

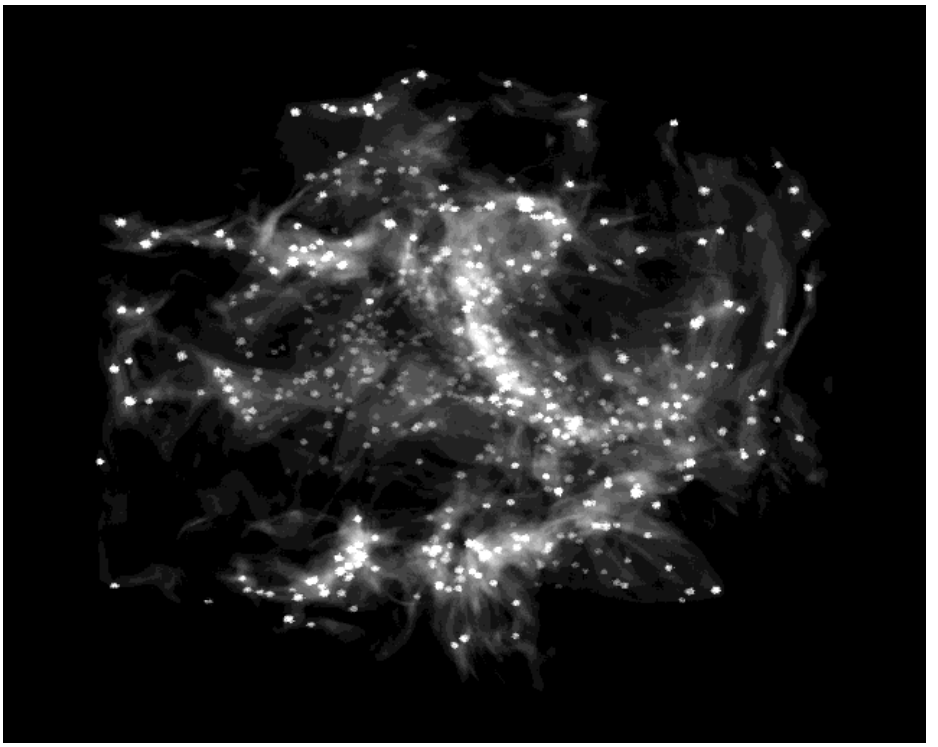


Fig. 13. Voxel projection of the density field of a snapshot of a 250^3 numerical simulation of supersonic and super-Alfvénic turbulence. Collapsing cores are highlighted as bright spheres, with brightness and size varying as a function of the core mass. The brightness also depends on the column density of gas between each core and the observer, in order to mimic the effect of dust extinction. Slightly extended patches of bright emission are "unresolved" stellar clusters. The fraction of mass in collapsing cores is about 5% in this simulation.

of statistical properties of turbulent flows with observational data, by computing synthetic spectral maps of molecular transitions [65,66,68]. The synthetic spectral maps are obtained by computing the non-LTE radiative transfer problem using the density and velocity fields of the MHD simulations. We have shown that fundamental statistical properties of supersonic turbulence are unambiguously found in the observational data of star-forming clouds [68].

In star-forming clouds, prestellar cores and young stars tend to concentrate in the densest filaments and cores. Since filaments and cores of the same nature are found in the numerical simulations, it is interesting to visualize the position of the collapsing cores selected numerically, relative to the gas density distribution. In Fig. 13 a voxel projection of the density field is shown, where all the numerically selected cores have been highlighted as bright spheres. The size and brightness is a function of the core mass. The brightness is also a function of the optical depth of the gas between the observer and the stars, to mimic the effect of dust

extinction. Figure 13 shows beautiful filamentary structure in both the gas and the stellar distribution, very reminiscent of observed star-forming regions. It is quite amazing that a numerical simulation of randomly driven supersonic and super-Alfvénic turbulence with periodic boundary conditions is able to produce at the same time i) density structures morphologically and statistically consistent with the observations; ii) prestellar cores correlated with the gas distribution in a way qualitatively similar to the observations and with a value of magnetic field strength typically observed; iii) a mass distribution of the same prestellar cores that is consistent with the observed stellar mass distribution; and iv) a star formation efficiency consistent with that in observed molecular clouds of only a few percent per large-scale dynamical time.

6 Conclusions

The main conclusion from the preceding sections is that the statistics of star formation is primarily controlled by supersonic turbulence, rather than by gravity. Star formation takes place in cold molecular clouds, which are part of a turbulent cascade in the interstellar medium. The ultimate energy input to the cascade comes from supernovae, with a possibly significant contribution from local variations of the galactic rotation curve (density waves). The clouds owe their existence to random convergence of the interstellar medium velocity field, which creates local density enhancements over a range of scales. The internal, supersonic and super-Alfvénic velocity field in molecular clouds is responsible for their fragmentation, thus preventing global collapse but triggering local collapse at the many local density maxima whose mass exceeds the local Jeans' mass. Such prestellar cores are formed as sheet corrugations and filamentary density enhancements, and are taken over by self-gravity only after they have been shaped by the turbulence.

The velocity field of the cascade is dominated by power in solenoidal (shearing) motions, even though it is supersonic and super-Alfvénic. Its spectrum of kinetic energy is less steep than its velocity and magnetic field power spectra, which explains how conditions can be super-Alfvénic on small (molecular cloud) scales, even though there is rough equipartition between magnetic and kinetic energy density on large (disk thickness) scales.

A Salpeter like IMF is the result of the near-self-similar, power law nature of turbulence in molecular clouds, in combination with density jump amplitudes determined by MHD-shock jump conditions.

Star formation (at least in our galaxy) bites its own tail; it is driven by supernovae and at the same time the birth of massive stars gives rise to new supernovae that re-enforce the driving. External sources of turbulence, such as kinetic energy from galaxy collisions and merging may be the primary driving agent in star-burst galaxies.

Different physical conditions (primarily higher temperatures and lower metal abundances) in the Early Universe would lead to higher mass at the low-mass cut-off, and a much weaker magnetic field would lead to a steeper IMF slope.

Acknowledgments

We are grateful to the referee for suggestions that led to improvements of the presentation. The work of ÅN was supported by a grant from the Danish Natural Science Research Council, and in part by the Danish National Research Foundation, through its establishment of the Theoretical Astrophysics Center. The work of PP was in part performed while PP held a National Research Council Associateship Award at the Jet Propulsion Laboratory, California Institute of Technology.

References

1. Adams, F. C. and Fatuzzo, M.: Astrophys. J. **464**, 256 (1996)
2. Arny, T.: Astrophys. J. **169**, 289 (1971)
3. Arons, J. and Max, C. E.: Astrophys. J. **196**, L77 (1975)
4. de Avillez, M. A.: MNRAS **315**, 479 (2000)
5. Ballesteros-Paredes, J., Vázquez-Semadeni, E., and Scalo, J.: Astrophys. J. **515**, 286 (1999)
6. Ballesteros-Paredes, J., Hartmann, L., and Vázquez-Semadeni, E.: Astrophys. J. **527**, 285 (1999)
7. Benzi, R., Ciliberto, S., Tripiccione, R., Baudet, C., Massaioli, F., and Succi, S.: Phys. Rev. E **48**, 29 (1993)
8. Biskamp, D., and Muller, W. C.: Physics of Plasmas, **7** (12), 4889 (2000)
9. Boldyrev S.: Astrophys. J. **569**, (2002)
10. Boldyrev S.: private communication (2002)
11. Boldyrev, S., Nordlund, Å., and Padoan, P.: Astrophys. J. **573**, (2002)
12. Boldyrev, S., Nordlund, Å., and Padoan, P.: Phys. Rev. Lett., **89**, 031102 (2002)
13. Bonazzola, S., Heyvaerts, J., Falgarone, E., Perault, M., and Puget, J. L.: Astron. Astrophys. **172**, 293 (1987)
14. Bonazzola, S., Perault, M., Puget, J. L., Heyvaerts, J., Falgarone, E., and Panis, J. F.: Journal of Fluid Mechanics **245**, 1 (1992)
15. Boulares, A. and Cox, D. P.: Astrophys. J. **365**, 544 (1990)
16. Bourke, T. L., Myers, P. C., Robinson, G., and Hyland, A. R.: Astrophys. J. **554**, 916 (2001)
17. Bouvier, J., Stauffer, J. R., Martin, E. L., Barrado y Navascues, D., Wallace, B., and Bejar, V. J. S.: Astron. Astrophys. **336**, 490 (1998)
18. Chandrasekhar, S.: Proc. R. Soc. London, Ser. A, **246**, 301 (1958)
19. Dubrulle, B.: Phys. Rev. Lett. **73**, 959 (1994)
20. Efstathiou, G.: MNRAS **317**, 697 (2000)
21. Elmegreen, B. G.: Astrophys. J. **419**, L29 (1993)
22. Elmegreen, B. G.: in M. Livio (ed.), *Unsolved Problems in Stellar Evolution*, Cambridge University Press, p. 59 (1998)
23. Elmegreen, B. G.: Astrophys. J. **522**, 915 (1999)
24. Elmegreen, B. G.: in T. Montmerle and Ph. André (eds.) *From Darkness to Light*, ASP Conference Series (2002)
25. Elmegreen, B. G.: Astrophys. J. **530**, 277 (2000)
26. Falgarone, E., Panis, J.-F., Heithausen, A., Pérault, M., Stutzki, J., Puget, J.-L., and Bensch, F.: Astron. Astrophys. **331**, 669 (1998)
27. Falgarone, E. and T. G. Phillips: Astrophys. J. **359**, 344 (1990)

28. Falgarone, E., J.-L. Puget, and M. Perault: Astron. Astrophys. **257**, 715 (1992)
29. Frisch, U.: *Turbulence*, Cambridge University Press (1995)
30. Gaustad, J. E.: Astrophys. J. **138**, 1050 (1963)
31. Goodman, A., and Heiles, C.: Astrophys. J. **424**, 208 (1994)
32. Goodman, A. A., Barranco, J. A., Wilner, D. J., and Heyer, M. H.: Astrophysical Journal **504**, 223 (1998)
33. Gudiksen, B. V.: PhD thesis, Niels Bohr Institute for Astronomy, Physics, and Geophysics, University of Copenhagen (1999)
34. Hartmann, L., Ballesteros-Paredes, J., and Bergin, E. A.: Astrophys. J. **562**, 852 (2001)
35. Heithausen, A., Bensch, F., Stutzki, J., Falgarone, E., and Panis, J. F.: Astron. Astrophys. **331**, L65 (1998)
36. Heitsch, F., Mac Low, M., and Klessen, R. S.: Astrophys. J. **547**, 280 (2001)
37. Heyer, M. H., Carpenter, J. M., and Ladd, E. F.: Astrophysical Journal **463**, 630 (1996)
38. Hillenbrand, L. A.: Astronomical Journal **113**, 1733 (1997)
39. Hoyle, F.: Astrophys. J. **118**, 513 (1953)
40. Jijina, J., Myers, P. C., and Adams, F. C.: Astrophysical Journal Supplement Series **125**, 161 (1999)
41. Klessen, R. S., Heitsch, F., and Mac Low, M.: Astrophys. J. **535**, 887 (2000)
42. Kolmogorov, A. N.: Dokl. Akad. Nauk. SSSR, **30**, 301 (1941)
43. Korpi, M. J., Brandenburg, A., Shukurov, A., Tuominen, I., and Nordlund, Å.: Astrophys. J. **514**, L99 (1999)
44. Lada, C. J., Lada, E. A., Clemens, D. P., Bally, J.: Astrophys. J., **429**, 694 (1994)
45. Larson, R. B.: MNRAS **186**, 479 (1979)
46. Larson, R. B.: MNRAS **194**, 809 (1981)
47. Larson, R. B.: MNRAS **200**, 159 (1982)
48. Low, C. and Lynden-Bell, D.: MNRAS **176**, 367 (1976)
49. Luhman, K. L.: Astrophys. J. **525**, 466 (1999)
50. Luhman, K. L.: Astrophys. J. **544**, 1044 (2000)
51. Luhman, K. L. and Rieke, G. H.: Astrophys. J. **525**, 440 (1999)
52. Luhman, K. L., Rieke, G. H., Young, E. T., Cotera, A. S., Chen, H., Rieke, M. J., Schneider, G., and Thompson, R. I.: Astrophys. J. **540**, 1016 (2000)
53. Mac Low, M.-M.: in J. Franco, A. Carramiñana (eds.), *Interstellar Turbulence*, Cambridge University Press (1999)
54. Mac Low, M.-M.: Astrophys. J. **524**, 169 (1999)
55. Mac Low, M., Smith, M. D., Klessen, R. S., and Burkert, A.: jtitleThe Decay of Supersonic and Super-Alfvenic Turbulence in Star-Forming Clouds Astrophysics and Space Science **261**, 195 (1998)
56. Motte, F., Andre, P., and Neri, R.: Astron. Astrophys. **336**, 150 (1998)
57. Motte, F., André, P., Ward-Thompson, D., and Bontemps, S.: Astron. Astrophys. **372**, L41 (2001)
58. Nordlund, Å. and Padoan, P.: in J. Franco, A. Carramiñana (eds.), *Interstellar Turbulence*, Cambridge University Press, p. 218 (1999)
59. Onishi, T., Mizuno, A., Kawamura, A., and Fukui, Y.: in T. Nakamoto (ed.) Proceedings of *Star Formation 1999*, Nobeyama Radio Observatory, p. 153 (1999)
60. Ossenkopf, V., and Mac Low, M.-M.: Astron. Astrophys. **390**, 307 (2002)
61. Ostriker, E. C., Stone, J. M., and Gammie, C. F.: Astrophys. J. **546**, 980 (2001)
62. Padoan, P.: MNRAS **277**, 377 (1995)
63. Padoan, P., Jones, B. J. T., and Nordlund, Å.: Astrophys. J. **474**, 730 (1997)

64. Padoan, P., Nordlund, Å., and Jones, B. J. T.: MNRAS **288**, 145 (1997)
65. Padoan, P., Juvela, M., Bally, J., and Nordlund, A.: Astrophys. J. **504**, 300 (1998)
66. Padoan, P., Bally, J., Billawala, Y., Juvela, M., and Nordlund, Å.: Astrophys. J. **525**, 318 (1999)
67. Padoan, P. and Nordlund, Å.: Astrophys. J. **526**, 279 (1999)
68. Padoan, P., Juvela, M., Goodman, A. A., and Nordlund, Å.: Astrophys. J. **553**, 227 (2001)
69. Padoan, P., and Nordlund, Å.: Astrophys. J., **576**, 870 (2002)
70. Padoan, P., Nordlund, Å., Rögnvaldsson, Ö. E., and Goodman, A.: in T. Montmerle and Ph. André (eds.) *From Darkness to Light*, ASP Conference Series (2002)
71. Padoan, P., Nordlund, Å., Rögnvaldsson, Ö. E., and Goodman, A.: submitted to Astrophys. J., astro-ph/0011229 (2002)
72. Passot, T., Vazquez-Semadeni, E., and Pouquet, A.: Astrophys. J. **455**, 536 (1995)
73. Passot, T., and Vazquez-Semadeni, E.: Phys. Rev. E **58**, 4501 (1998)
74. Rees, M. J.: MNRAS **176**, 483 (1976)
75. Reipurth, B. and Clarke, C.: Astron. J., **122**, 432 (2001)
76. Salpeter, E. E.: Astrophys. J. **121**, 161 (1955)
77. Scalo, J. M. and Pumphrey, W. A.: Astrophys. J. **258**, L29 (1982)
78. Scalo, J., Vazquez-Semadeni, E., Chappell, D., and Passot, T.: Astrophys. J. **504**, 835 (1998)
79. She, Z.-S., and Lévêque, E.: Phys. Rev. Lett. **72** 336 (1994)
80. Shu, F. H., Adams, F. C., and Lizano, S.: Annual Review of Astronomy and Astrophysics **25**, 23 (1987)
81. Silk, J.: Astrophys. J. **214**, 152 (1977)
82. Silk, J.: Astrophys. J. **214**, 718 (1977)
83. Stone, J. M., Ostriker, E. C., Gammie, C. F.: Astophys. J. Lett. 508, L99 (1998)
84. Struck, C., and Smith, D. C.: Astrophys. J. **527**, 673 (1999)
85. Suchkov, A. A., and Shchekinov, I. A.: Soviet Astronomy **19**, 403 (1976)
86. Testi, L., and Sargent, A. I.: Astrophys. J. **508**, L91 (1998)
87. Vazquez-Semadeni, E.: Astrophys. J. **423**, 681 (1994)
88. Vazquez-Semadeni, E. and Gazol, A.: Astron. Astrophys. **303**, 204 (1995)
89. Vazquez-Semadeni, E., Passot, T., and Pouquet, A.: Astrophys. J. **441**, 702 (1995)
90. Yoneyama, T.: Publications of the Astronomical Society of Japan **24**, 87 (1972)
91. Yoshii, Y. and Saio, H.: Astrophys. J. **295**, 521 (1985)
92. Zweibel, E. G. and Josafatsson, K.: Astrophys. J. **270**, 511 (1983)

The Structure and Dynamics of Filamentary Molecular Clouds

Jason D. Fiege[1,2]

[1] Herzberg Institute of Astrophysics, National Research Council of Canada, 5071 West Saanich Road,Victoria, B.C. Canada V9E 2E7
[2] Canadian Institute for Theoretical Astrophysics, McLennan Labs, University of Toronto, 60 St. George Street, Toronto, Ontario, Canada M5S 3H8

Abstract. We summarize models of self-gravitating, filamentary molecular clouds that are threaded by helical magnetic fields. We review the observational evidence suggesting that some clouds may exist in this regime. An analytic limit is derived that demonstrates why the Fiege & Pudritz model always tends to an $\sim r^{-(1.9\pm0.1)}$ density profile at large radius, in agreement with observations. An equation is also derived for the radial compression or expansion of filamentary clouds that are out of equilibrium. We use this analysis to show that filamentary clouds are extremely stable with respect to radial collapse or expansion, which allows them to survive long enough to fragment gravitationally along their axes. Fragmenting, helically magnetized filaments have a unique kinematic signature, which is discussed and compared with recent observations. We discuss models of prolate cores that form within filaments. Polarimetry models of filaments are discussed and compared with the data.

1 Introduction

Molecular clouds are exotic astrophysical objects whose internal structure and dynamics are remarkably rich. They range from wispy, weakly self-gravitating cirrus clouds with no star formation, to strongly self-gravitating giant molecular clouds (GMCs), whose massive cores are the birthplaces of stellar clusters and groups, and the seat of most of the star formation in our Galaxy. Dark clouds are intermediate in mass between these extremes and represent a regime where star formation is relatively isolated, and proceeds at a modest pace compared to the GMCs.

The typical temperature of a molecular cloud is about $15K$. At this temperature, the one-dimensional thermal velocity dispersion is $0.23\ km\ s^{-1}$, assuming a mean molecular weight of 2.3. It was shown by Bonnor [6] & Ebert [17] that the maximum, critical mass $M_{B.E.}$ of a self-gravitating isothermal sphere surrounded by an external pressure P_S is given by

$$M_{B.E.} = 1.182\frac{\sigma^4}{G^{3/2}P_S^{1/2}} = 8.2\left(\frac{\sigma}{0.23\ km\ s^{-1}}\right)^4\left(\frac{P_S/k_B}{10^4\ K\ cm^{-3}}\right)^{-1/2} M_\odot, \quad (1)$$

where σ is the velocity dispersion of the gas and G is the gravitational constant. Therefore, thermal pressure can only support a few solar masses of material if the external pressure is on the order of $P_S/k_B \approx few \times 10^4\ K\ cm^{-3}$ [7] (see

E. Falgarone and T. Passot (Eds.): LNP 614, pp. 299–328, 2003.
© Springer-Verlag Berlin Heidelberg 2003

Sect. 2), where k_B is the Boltzmann constant. This is far less massive than some molecular clouds, which can exceed $10^6 M_\odot$ [53]. Molecular clouds would be in free-fall collapse and star-formation would proceed at an extraordinary rate if other means of support did not exist. It has long been known that spectral lines originating in molecular clouds are strongly non-thermal, and that clouds contain dynamically important magnetic fields. The non-thermal motions within molecular clouds have usually been attributed to either moderate amplitude Alfvén waves [2] or full-blown MHD turbulence (see [63] for review). Either of these viewpoints is consistent with the observational fact that the ordered component of the magnetic field, the random component, and the kinetic motions of the gas [45,46] usually have comparable energy densities. In either case, it appears that most molecular clouds exist in a physically interesting regime of approximate equipartition (to within a factor of a few) between kinetic, gravitational, and magnetic energy densities, which provides a unique laboratory for the study of astrophysical MHD physics.

Numerical simulations of MHD turbulence generally find that the turbulence dissipates in a few eddy turnover times. Thus, many simulations assume an ongoing injection of energy to mimic the internal energy provided by bipolar outflows, stellar winds, supernovae, supershells, and Galactic shear. Outflows and stellar winds from regions of active star formation are likely to be the most important mechanisms for maintaining the turbulence (e.g. [63]) in molecular clouds. However, the energy derived from star formation cannot be the sole mechanism, since there are many examples of molecular clouds that have strongly non-thermal linewidths but contain no sign of star formation. One must consider the possibility that either the turbulence is not damping as rapidly as the simulations suggest, or that other mechanisms exist for re-energizing the turbulence from external sources.

Molecular clouds and the substructure within them are often organized into slender filaments, which exist on many different scales. On the largest scales are the enormous Northern and Southern filaments in the Orion complex, which are respectively estimated to be $90pc$ and $300pc$ in length [35], each containing a few times 10^4 $M_\odot$ of gas. However, most filaments are somewhat smaller, such as the Taurus filaments and the streamers of rho Ophiuchus, which are typically 5-10 pc in length and contain $1 - few \times 100 M_\odot$ of material. There are also many examples of isolated globular filaments, which are comprised of strings of dense Bok globules that are usually periodically spaced and often connected by a more diffuse filament of gas [54].

Dense internal structures within molecular clouds are also filamentary. The "integral-shaped" filament of Orion A is the best example of this type of filament [5,4,30], which is about $13pc$ long and contains about $5000 M_\odot$ [5]. This filament is fragmented into a string of cores that are actively forming stars. The massive Kleinmann-Low (KL) core, which is an active region of massive star formation resides in the OMC-1 region of the filament. Recent sub-millimetre intensity maps show the existence of embedded filaments in other molecular clouds as

well [42,31,36], where material is often organized into filaments or long chains of cores that may have formed from the fragmentation of a filament.

Fiege and Pudritz [20] (hereafter FP1) have developed models of self-gravitating filaments, in which a helical magnetic field threads the cloud. Their models are pressure-truncated, meaning that they are bounded by a non-self-gravitating external medium whose pressure defines the radius of the self-gravitating filament. Alternatively, many simulations have shown that filaments may arise naturally, as a consequence of MHD turbulence, where they are often attributed to the intersection of shocks (see review [63] and references therein). This may account for some of the diffuse filaments seen in the ISM or in some molecular clouds that are weakly self-gravitating. However, we argue in this paper that the Fiege & Pudritz (hereafter FP) models may apply in relatively quiescent regions where strong self-gravity (and possibly a toroidal field) holds filaments together for a relatively long time.

The layout of the paper is as follows. We discuss the role of self-gravity in molecular clouds in Sect. 2 in terms of the classic analysis of Bonnor [6] and Ebert [17]. We briefly compare this analysis to the findings of turbulence simulations in Sect. 3. Section 4 discusses the structure and stability of the Ostriker [49] model of unmagnetized, isothermal, self-gravitating filamentary clouds. Section 5 presents a general virial analysis of self-gravitating filamentary clouds, including the effects of magnetic fields. We present detailed models of the structure of such filaments in Sect. 6. The radial dynamics of filamentary clouds are discussed in Sect. 7. In particular, we demonstrate the stability of magnetized, self-gravitating clouds against collapse or re-expansion in Sect. 7.2. In Sect. 8, we comment on the large-scale topology of the magnetic field in regions containing filaments. Section 9 reviews the fragmentation properties of filaments, and Sect. 10 discusses the structure of the cores that might form as a result of fragmentation. We review evidence in favour of the FP model from sub-millimetre polarimetry in Sect. 11. Our final discussion is provided in Sect. 12.

2 Self-gravity in Pressure-Truncated Isothermal Molecular Clouds

Molecular clouds and the cores within them are self-gravitating. There is a wealth of observational data demonstrating approximate virial equilibrium between the gravitational binding energy of molecular clouds and their internal kinetic energy [53], which is usually dominated by turbulence, except in quiescent cores, where thermal support becomes important [24,25,11]. It is relatively straightforward to derive an observational criterion for the importance of self-gravity from the classical analysis of Bonnor [6] and Ebert [17]. As we briefly mentioned in Sect. 1, Bonnor and Ebert demonstrated that a pressure bounded self-gravitating sphere has a critical mass for gravitational instability defined by (1). The radius of the critical cloud is given by

$$R_{B.E.} = 0.411 \frac{G M_{B.E.}}{\sigma^2}. \tag{2}$$

Combining (1) and (2), it is easy to show that there exists a critical surface density, averaged over the projected area of the cloud, that depends only on the external pressure:

$$\Sigma_{crit} = 1.60\sqrt{\frac{P_S}{G}} = 1.3 \times 10^{-2} \left(\frac{P_S/k_B}{10^{4.5}\ K\ cm^{-3}} \right)^{1/2} g\ cm^{-2}. \quad (3)$$

We have normalized the above expression to an external pressure of $P_S/k_B = 10^{4.5}\ K\ cm^{-3}$, which is typical of the time-averaged bounding pressure exerted on molecular clouds by the interstellar medium, including non-thermal contributions. This normalization was estimated as follows. Boulares and Cox [7] showed that the total pressure of the interstellar medium is $\approx (2.8 \pm 0.4) \times 10^4\ K\ cm^{-3}$ at the Galactic midplane. This estimate includes thermal pressure, magnetic fields, cosmic ray pressure, and turbulent pressure. Molecular clouds are essentially transparent to cosmic rays. Although the contribution of cosmic rays to the ISM pressure is important to the large-scale vertical stratification of the ISM, their contribution to the surface pressure on molecular clouds can be neglected since there is no significant interaction at the surface. Therefore, their contribution of $\approx 0.7 \times 10^4\ K\ cm^{-3}$ should be subtracted, resulting in a surface pressure of $\approx 2.1 \times 10^4\ K\ cm^{-3}$ at the midplane of the Galaxy. However, many molecular clouds are associated with HI clouds or complexes, which are typically overpressured by a factor of a few, and in some cases up to an order of magnitude, compared to other regions of the ISM [13]. Thus, we adopt $10^{4.5 \pm 0.5}\ K\ cm^{-3}$ as representative of the possible range of pressures bounding molecular clouds in most regions. Expressed as a critical number surface density of hydrogen nuclei, (3) becomes

$$N_{crit} = 7.73 \times 10^{21} \left(\frac{P_S/k_B}{10^{4.5}\ K\ cm^{-3}} \right)^{1/2} cm^{-2}. \quad (4)$$

Assuming that $A_v \approx N/(2 \times 10^{21} cm^{-2})$ [57], we see that a critical Bonnor-Ebert cloud has an optical depth of $A_v \approx 3.9$.

It can be shown from (1) and (2) that the mean (volume averaged) pressure within a critical Bonnor-Ebert sphere is higher than the external pressure by only a factor of 2.46. Therefore, strongly self-gravitating clumps within clouds should characteristically have an excess extinction of $\sim \sqrt{2.46} \approx 1.57$ times that of the cloud itself, which gives a total average extinction of ~ 10 for a critical spherical clump surrounded by the mean pressure within a critical cloud. Higher surface densities and extinctions are possible for cores residing in central, high pressure regions of their surrounding molecular clouds.

Note that all of the numerical estimates given in this section typically vary by a factor of a few when a magnetic field is present. Generally, a purely poloidal magnetic field offers support against self-gravity, and therefore serves to increase the critical mass M_{mag} beyond the Bonnor-Ebert mass $M_{B.E.}$:

$$M_{mag} \approx 1.18 \left[1 - \left(\frac{0.17}{G^{1/2} dM/d\Phi_B} \right)^2 \right]^{-3/2} M_{B.E.}, \quad (5)$$

where $dM/d\Phi_b$ is the mass to magnetic flux ratio evaluated for a flux tube passing through the centre of the cloud [59]. This formula is an approximation to the numerical models calculated by Tomisaka, Ikeuchi, & Nakamura [59]. Note that M_{mag} given by the above formula obviously does not apply in the unmagnetized limit, for which M_{mag} must tend to $M_{B.E.}$. However, it is accurate to within $\sim 10\%$ for the range of magnetized models studied by these authors.

An important point is that unmagnetized spheroidal self-gravitating clouds, can always be made to collapse by increasing the external pressure beyond P_{crit}, defined by

$$P_{crit} = 1.398 \frac{\sigma^8}{G^3 M^2} \tag{6}$$

for a given fixed mass M, which is obtained by solving (1) for P_S. This instability occurs because the gravitational energy scales as R^{-1}, and therefore can be made to dominate over the internal kinetic energy of an isothermal cloud, which remains constant as the cloud contracts under the influence of the external pressure. Note that the magnetic critical mass M_{mag} diverges as the term in square brackets in (5) tends to zero, which implies a critical mass to flux ratio

$$\left.\frac{dM}{d\Phi_B}\right|_{crit} = \frac{0.17}{\sqrt{G}}. \tag{7}$$

The Bonnor-Ebert instability only applies to magnetically super-critical clouds, where the mass to flux ratio exceeds this value. Pressure-driven collapse is not possible for magnetically sub-critical clouds, where the mass to flux ratio is less than $dM/d\Phi_B|_{crit}$.

In contrast to spheroidal clouds, both magnetized and unmagnetized pressure truncated filaments can always can withstand an arbitrarily large external pressure without collapsing, provided that a critical mass per unit length is not exceeded. The reasons behind this strikingly different behaviour are discussed in Sect. 4 and Sect. 5.1 below.

3 Turbulence Simulations

Existing state of the art simulations produce reasonably self-gravitating structures. For example, the simulations of Padoan et al. [51] typically have a characteristic optical depth of $A_{v,0} = 3$, and have a dynamic range of about 30 in surface density, as traced by simulated sub-millimetre intensity (See their Fig. 4 for example. Note that there are a few isolated points where the density contrast is higher.). The discussion in Sect. 2 suggests that clumps with these characteristics should be self-gravitating.

MHD turbulence simulations are usually dominated by filamentary structures at somewhat lower densities than the cores (e.g. [3,51,28]) These filaments are transient objects that are often attributed to the intersection of shocks. They are usually quite uniform in appearance, and do not generally appear to be fragmenting gravitationally into strings of cores or clumps. Although self-gravitating

filaments are always gravitationally unstable to fragmentation, the filaments that form in simulations usually do not survive long enough for fragmentation to occur before dissipating.

This may represent an important difference between existing turbulence simulations and some observed filaments, which appear to be strongly fragmented, often into a periodic system of cores or clumps [16,30]. Periodic structure is especially common in globular filaments, which are isolated strings of Bok globules [54]. Magnetostatic filaments, like those of the FP model and others [12,56,49,47,48,12,26] differ from the filaments that form in simulations in that that they survive long enough for gravitational fragmentation to occur. In fact, a number of authors have performed linear stability analyses, which predict fragmentation with approximately the correct periodicity [12,47,48,26,21,14].

We provide an estimate for the timescale required for fragmentation to occur in Sect. 9.2. While some filaments are undoubtedly transient objects formed at the intersection of shocks, perhaps others reside in relatively calm regions where they have time to fragment and develop periodic substructure, rather than re-expanding and dissolving into the surrounding turbulence. The Fiege & Pudritz model might apply to these filaments, whose internal structure, dynamics, stability, and cores are the subject of the remainder of this paper.

4 The Ostriker Model and Pressure-Truncated Hydrostatic Filaments

In Sect. 3, we argued that some filamentary molecular clouds might live much longer than the transient filaments that form at the intersections of shocks in many simulations of MHD turbulence. We now turn to a more detailed study of these filaments, which we approximate as being in quasi-static radial equilibrium.

The simplest equilibrium model of a filamentary cloud is that of a self-gravitating, unmagnetized, isothermal filament. Ostriker [49] studied such clouds, and found an exact solution for the density structure [1]:

$$\rho = \frac{\rho_c}{\left(1 + r^2/8r_0^2\right)^2}, \tag{8}$$

where r_0 is the core radius of the filament. The core radius is defined by

$$r_0 = \frac{\sigma}{\sqrt{4\pi G \rho_c}}, \tag{9}$$

where σ is the 1-dimensional velocity dispersion of the gas, assumed to be constant and possibly including both thermal and non-thermal contributions, G is the gravitational constant, and ρ_c is the central density along the axis of the filament. In the absence of external pressure, the Ostriker solution extends to infinite radius.

[1] This solution is actually contained within a class of magnetized models discovered by Stodòlkiewicz [56].

The mass per unit length, which we will refer to as the line mass for brevity, of the Ostriker solution is exactly

$$m_{vir} = \frac{2\sigma^2}{G}. \tag{10}$$

This quantity is denoted m_{vir}, since it also represents the virial line mass in the virial equation (14) below. An interesting property of the Ostriker solution is that no equilibrium solution exists when $m \neq m_{vir}$. If $m > m_{vir}$, the entire filament must collapse radially in about a free-fall time. On the other hand, the filament will re-expand into the surrounding gas if $m < m_{vir}$. It appears that there is a fine tuning problem because it would be difficult to assemble an Ostriker filament with precisely the correct line mass required for equilibrium.

Fortunately, there is a simple solution to this problem. Let us now assume that the filament is bounded by an external pressure P_S at a radius R_S. Then, by integrating (8), it is easy to show that the line mass is given by

$$m(R_S) = \frac{m_{vir}}{1 + 8r_0^2/{R_S}^2}. \tag{11}$$

This can be re-written in terms of the external pressure:

$$m(P_S) = m_{vir}\left[1 - \left(\frac{P_S}{\sigma^2\rho_c}\right)^{1/2}\right], \tag{12}$$

where we have made use of (8) and the fact that $P = \sigma^2\rho$. Finally, we solve (12) for the central density:

$$\rho_c = \frac{P_S/\sigma^2}{(1 - m/m_{vir})^2}. \tag{13}$$

Thus, we find that a solution exists for ρ_c for any choice of $m < m_{vir}$ and P_S. The filament is always able to find an equilibrium configuration as long as the critical line mass m_{vir} is not exceeded. Even an arbitrarily large surface pressure P_S cannot cause such a filament to collapse radially. This behaviour, which was first noted by McCrea [39], is completely different from the Bonnor-Ebert spheres discussed in Sect. 2, which can withstand a maximum external pressure P_{crit}, given by (6). The radial stability of filaments depends only on m/m_{vir}. Although this result was derived for purely hydrostatic filaments, it also holds true for magnetized filaments (discussed in Sect. 5.1 and Sect. 6.2 below), for which the magnetic field provides a correction to m_{vir}.

5 Virial Analysis of Magnetized Filamentary Molecular Clouds

5.1 The Virial Equation Applied to Magnetized Filaments

FP1 derived a useful form of the MHD virial equation for self-gravitating, magnetized, filamentary molecular clouds. The magnetic field is axisymmetric, with

both poloidal and toroidal components allowed, so that the field is helical in general. The virial equation for filaments is most concisely written as

$$\frac{P_S}{\langle P \rangle} = 1 - \frac{m}{m_{vir}} \left(1 - \frac{\mathcal{M}}{|\mathcal{W}|} \right), \tag{14}$$

where P_S, m, and m_{vir} are defined in Sect. 4, and the remaining terms are defined as follows.
Average Pressure:

$$\langle P \rangle = \frac{\int_{\mathcal{V}} P d\mathcal{V}}{\mathcal{V}} \tag{15}$$

Gravitational Binding Energy Per Unit Length:

$$\mathcal{W} = -m^2 G \tag{16}$$

Magnetic Energy Term:

$$\mathcal{M} = \frac{1}{4\pi} \int B_z^2 d\mathcal{V} - \left(\frac{B_{z,S}^2 + B_{\phi,S}^2}{4\pi} \right) \mathcal{V} \tag{17}$$

In these equations, $\mathcal{V} \equiv \pi {R_S}^2$ is the cross-sectional area of the filament, and subscript S refers to quantities evaluated at the cylindrical surface of the cloud.

The gravitational energy term in (14) has the interesting property that it does not depend on the radius of the cloud, but only on the line mass. This scaling is very different from spheroidal clouds, where the gravitational energy W scales as $W \propto r^{-1}$. This difference accounts for the unusual stability properties of pressure-bounded filaments that were discussed in Sect. 4. The virial equation given above demonstrates that the critical line mass is modified for magnetized filaments. We observe that a solution to (14) exists, with $\langle P \rangle > P_S$, as long as $m < m_{crit}$, where m_{crit} is defined by

$$m_{crit} \equiv m_{vir} (1 - \mathcal{M}/|\mathcal{W}|)^{-1}. \tag{18}$$

Thus, m_{crit} replaces m_{vir} as the critical line mass of a magnetized filament. It is not obvious from (17) that m_{crit} remains invariant when a filament is compressed radially or expands due to a changing external pressure. However, simple scaling arguments presented in Sect. 6.2 demonstrate that m_{crit} is invariant, and that magnetized filaments are always stable in the sense of Bonnor [6] and Ebert [17] whenever $m < m_{crit}$.

The magnetic energy term in (14) includes both a volume term, and a surface term. The volume term involves only B_z and provides a positive contribution to $\mathcal{M}$ (assuming that $B_{z,S}^2 < \mathcal{V}^{-1} \int B_z^2 d\mathcal{V}$). Thus, the poloidal field provides support to the filament. On the other hand, the toroidal field component enters only through the surface term, and always has a net confining effect, which is given explicitly by (19) below. The sign of $\mathcal{M}/|\mathcal{W}|$ determines whether the poloidal field or the toroidal field component dominates the structure of the

cloud. When $\mathcal{M}/|\mathcal{W}| > 0$, the poloidal field is dominant and there is net magnetic support. On the other hand, the toroidal field dominates when $\mathcal{M}/|\mathcal{W}| < 0$, and the cloud is squeezed radially. [2]

The force per unit volume due to the toroidal field is given by

$$\mathcal{F} = -\frac{1}{r^2}\frac{d}{dr}\left(\frac{r^2 B_\phi^2}{8\pi}\right). \tag{19}$$

The force is radially inward as long as B_ϕ does not fall off more steeply than r^{-1}. The total poloidal current $I(r)$ inside of radius r obeys

$$I(r) \propto rB_\phi. \tag{20}$$

In general, I is a monotonically increasing or constant function of r unless there are reversals in the current density $J = (2\pi r)^{-1} dI/dr$ within the filament. Thus, the toroidal field cannot fall off more rapidly than r^{-1} unless there are such current reversals. It is noteworthy that the confinement due to the toroidal field would vanish at any radius where the total current interior to that radius vanishes. This has important consequences for the overall topology of the magnetic field threading filaments, which we discuss in Sect. 8.

5.2 Observational Considerations

FP1 compiled published data on filamentary clouds, which they analyzed using the virial equation 14. They inferred that the magnetic field threading many filaments appears to be toroidally dominated. The basic point of their analysis is that the quantities $P_S/\langle P\rangle$ and m/m_{vir} are relatively straightforward to estimate observationally, which gives the remaining term $\mathcal{M}/|\mathcal{W}|$ by (14). FP1 found evidence that most of the clouds in their sample have $\mathcal{M}/|\mathcal{W}| < 0$, as illustrated in Fig. 1, which suggests that some filamentary molecular clouds might be dominated by dynamically important toroidal magnetic fields.

A physical interpretation of this finding is that many filaments appear to have an internal velocity dispersion that is too high for confinement by self-gravity. Such clouds must either be out of equilibrium and expand radially, or there must be an alternative source of binding energy. Since there is no observational evidence for global expansion, it is reasonable to consider magnetically confined models.

We show in the next few sections that this choice is quite fruitful in accounting for the observed properties of filamentary clouds, globular filaments, and cores associated with filaments.

[2] Note that it is possible, in principle, to create magnetically confined filaments using only B_z, if $B_{z,S}$ is greater than the mean B_z within the filament, so that the magnetic pressure acts inward. Such models of filaments are implausible, but a related model for cores [15] is discussed in Sect. 57.

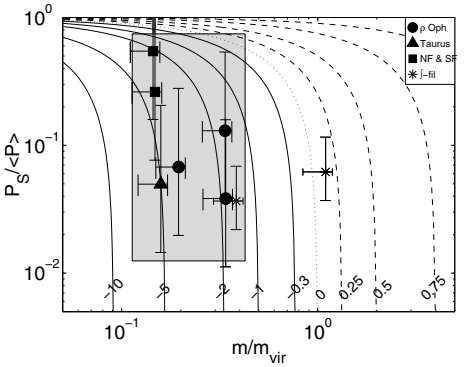


Fig. 1. (From FP1) Virial analysis of filamentary clouds. Curves are shown for various values of the virial parameter $\mathcal{M}/|\mathcal{W}|$. Positive values (dashed curves) indicate that the poloidal field is dominant, while negative values (solid curves) indicate that the toroidal field is dominant. The dotted line represents solutions with $\mathcal{M}/|\mathcal{W}|$=0. The $\int$-shaped filament appears twice, because we have used two independent data sets in our analysis.

6 The Radial Structure of Filaments Threaded by Helical Magnetic Fields

Observations have shown that the density profile is $\sim r^{-2}$ in all three filamentary clouds where accurate measurements have been taken [1,33,30]. The Ostriker [49] model and other magnetized models [56,47,48] predict steep density profiles of $\sim r^{-4}$ when $r > r_0$, which are not allowed by the observations. In this section, we show that the FP models have an $r^{-(1.9\pm0.1)}$ density structure, which is in much better agreement with the data. We also demonstrate why this occurs by calculating a simple analytic limit of their model.

6.1 Basic Equations

Detailed magnetostatic models are a step beyond the virial analysis discussed in Sect. 5. Besides the standard equations of self-gravitating magnetostatics and an equation of state, it is necessary to specify the mass loading of the field lines. This requires two flux to mass ratios Γ_z and Γ_ϕ, for the poloidal and toroidal field components respectively. These flux to mass ratios are defined by

$$\Gamma_z = \frac{B_z}{\rho} \tag{21}$$

$$\Gamma_\phi = \frac{B_\phi}{r\rho}, \tag{22}$$

and are illustrated in Fig. 2 [3]. We assume that Γ_z and Γ_ϕ are constant within the filament. This has the advantage that it eliminates the possibility of interchange

[3] Technically, Γ_z is the ratio of the flux to the line mass contained within a flux tube.

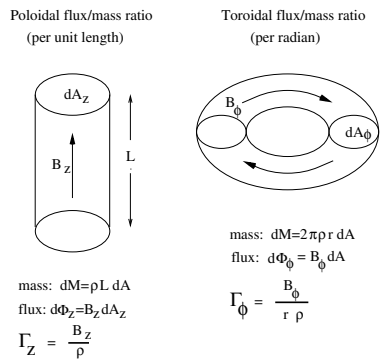


Fig. 2. (From FP1) A schematic illustration of the poloidal and toroidal flux to mass ratios introduced in (21) and (22).

instabilities, since all flux tubes then have equivalent mass loading. It is notable that constant Γ_ϕ arises naturally if a uniform filament (i.e. a uniform precursor of a filamentary cloud) with constant Γ_z and length L is twisted uniformly through an angle ϕ. Then

$$\frac{B_\phi}{B_z} = \frac{r\phi}{L}, \tag{23}$$

which leads to the result

$$\Gamma_\phi = \left(\frac{\phi}{L}\right)\Gamma_z. \tag{24}$$

We usually write Γ_z and Γ_ϕ in their dimensionless forms:

$$\begin{aligned}\tilde{\Gamma}_z &= \sqrt{\frac{\rho_c}{\sigma_c^2}}\left(\frac{B_z}{\rho}\right)\\ \tilde{\Gamma}_\phi &= \frac{1}{\sqrt{4\pi G}}\left(\frac{B_\phi}{r\rho}\right).\end{aligned} \tag{25}$$

FP1 drop the tildes for brevity, but we will find it useful to retain them.

The equations of self-gravitating magnetohydrostatics are Poisson's equation

$$\frac{1}{r}\frac{d}{dr}\left(r\frac{d}{dr}\Phi\right) = 4\pi G\rho, \tag{26}$$

and the equation of magnetohydrostatic equilibrium

$$\frac{d}{dr}\left(P + \frac{B_z^2}{8\pi}\right) + \rho\frac{d}{dr}\Phi + \frac{1}{r^2}\frac{d}{dr}\left(\frac{r^2B_\phi^2}{8\pi}\right) = 0. \tag{27}$$

We assume an equation of state of the form

$$P = \sigma^2\rho, \tag{28}$$

where σ is assumed to be constant. We will refer to this equation of state as "isothermal", by which we mean that the total one-dimensional velocity dispersion σ, which may include non-thermal contributions, is constant. The system of equations is closed with the additional assumption that the flux to mass ratios given in (21) and (22) are constant. In general, these equations must be solved numerically, as was done in FP1.

6.2 Scaling Arguments and the Bonnor-Ebert Stability of Magnetized Filaments

In this section, we briefly return to the discussion of the stability of magnetized filaments that we began in Sect. 5.1. Suppose that we were to re-write (26) to 28 entirely in terms of dimensionless quantities:

$$\tilde{r} \equiv r/r_0, \tag{29}$$

$$\tilde{\rho} \equiv \rho/\rho_c, \tag{30}$$

$$\tilde{P} \equiv P/(\sigma^2\rho_c), \tag{31}$$

$$\tilde{\mathbf{B}} \equiv \mathbf{B}/(\sigma\rho_c^{1/2}), \tag{32}$$

$$\tilde{\Phi} \equiv \Phi/\sigma^2, \tag{33}$$

where r_0, ρ_c, and σ are as defined in Sect. 4. We suppose that a solution to these dimensionless equations exists, which satisfies the usual boundary conditions for a self-gravitating magnetized filament. Specifically, $\tilde{\rho} = \tilde{P} = 1$, $d\tilde{\Phi}/d\tilde{r} = 0$, and $\tilde{B_\phi} = 0$ at $\tilde{r} = 0$, and $\tilde{\rho} = \tilde{P}s$ at the surface of the filament at $\tilde{r} = \tilde{R}_S$. The dimensionless line mass defined as $\tilde{m} \equiv 4\pi G m/\sigma^2$, can be calculated from the assumed solution as

$$\tilde{m}(\tilde{R}_S) = \int_0^{\tilde{R}_S} 2\pi\tilde{\rho}(\tilde{r})d\tilde{r}. \tag{34}$$

Now suppose that we wish to model a real filament whose measured line mass is $m = m_{vir}\tilde{m}$, which uniquely determines $\tilde{m}$, and hence $\tilde{R}_S$ by inverting (34). Such a solution exists as long as $m/m_{crit} < 1$, where m_{crit} is the critical line mass given by (18). Note that m/m_{crit} is a dimensionless quantity, which does not depend on the external pressure. Therefore, a magnetized isothermal filament that is initially sub-critical will remain subcritical regardless of how much the external pressure is increased.

Since $\tilde{R}_S$ is known, our solution determines the dimensionless surface pressure $\tilde{P}_S$ at this radius, which does not depend on the dimensional external pressure. Equation 31 therefore shows that the only effect of varying the external pressure is to change the central density, thus re-scaling the solution by (29) to (33):

$$\rho_c(P_S) = \frac{P_S}{\sigma^2\tilde{P}_S}. \tag{35}$$

Thus we have found a prescription which would allow a filament to adjust its structure to even an arbitrarily large external pressure. This verifies that magnetized filaments, like the truncated unmagnetized Ostriker filament, are always stable in the sense of Bonnor [6] and Ebert [17] whenever $m < m_{crit}$.

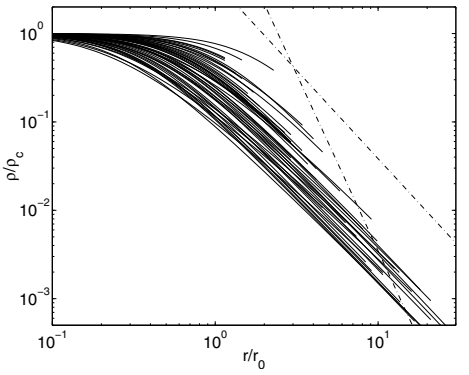


Fig. 3. (From FP1) A sample of 50 solutions to show the density structure of the FP model. The straight lines represent r^{-2} and r^{-4} power laws.

6.3 Solutions

Figure 3 shows a sample of 50 solutions to illustrate general trends in their properties. These solutions were chosen so that their virial parameters $P_S/\langle P\rangle$ and m/m_{vir} reside within the gray box in Fig. 1. Thus, these solutions are all in reasonable virial agreement with the observations.

The most striking feature of Fig. 3 is that nearly all of the solutions appear to approach approximately the same density profile of $\rho \propto r^{-(1.9\pm0.1)}$, which is much more shallow that the r^{-4} structure of the Ostriker solution.. It turns out that the dominance of the toroidal field in the outer envelope of the filament is responsible for this behaviour, as the example provided in Sect. 6.4 shows.

6.4 The r^{-2} Density Profile: An Analytical Limit

In this section, we examine the strongly magnetized limit of the FP model, in which self-gravity is neglected. This simplified model demonstrates why the FP model always tends to an approximately $\rho \propto r^{-(1.9\pm0.1)}$ density profile. As usual, $\tilde{\Gamma}_z$ and $\tilde{\Gamma}_\phi$ are assumed to be constant throughout this section.

It turns out that an exact analytical solution to (26) to (28) exists in this limit:

$$\frac{\rho(r)}{\rho_c} = \frac{4\pi}{\tilde{\Gamma}_\phi^2 (r/r_0)^2 + \tilde{\Gamma}_z^2} W\left[\frac{\tilde{\Gamma}_\phi^2 (r/r_0)^2 + \tilde{\Gamma}_z^2}{4\pi} e^{\tilde{\Gamma}_z^2/(4\pi)}\right], \tag{36}$$

where $W(x)$ is the Lambert W function defined by the relation

$$W(x) e^{W(x)} = x. \tag{37}$$

It is a property of the Lambert W function that $W(x) \to \ln(x)$ as $x \to \infty$. Thus, we find that

$$\frac{\rho(r)}{\rho_c} \to \frac{8\pi}{\tilde{\Gamma}_\phi^2} \frac{\ln(r/r_0)}{(r/r_0)^2}, \tag{38}$$

as $r \rightarrow \infty$, which is in good agreement with the $r^{-(1.9\pm0.1)}$ density profiles found by FP1. It is clear from (36) that the terms involving $\tilde{\Gamma}_z$ become insignificant when $r/r_0 \gg \tilde{\Gamma}_z/\tilde{\Gamma}_\phi \approx \mathcal{O}[1]$. Thus the density in the outer regions is controlled by the toroidal field alone. Note that the line mass diverges logarithmically:

$$m \rightarrow \frac{8\pi^2 \rho_c r_0^2}{\tilde{\Gamma}_\phi^2} \left[\ln\left(\frac{r}{r_0}\right) \right]^2 . \tag{39}$$

This analytical model is most appropriate for non-self-gravitating filaments, such as HI filaments, which were not studied by FP1. How well does this simple model approximate the FP models, where self-gravity was included? It should do reasonably well since $-5 \leq \mathcal{M}/|\mathcal{W}| < -1$ for most of their models (See Fig. 1 and (16) and (17)). We can make this argument more precise as follows. Since we have an analytic solution for the density, we can directly calculate the actual gravitational force term in (27), and compare to the magnetic force term. Giving only the asymptotic limits here, we find:

$$F_{grav} \equiv -\rho \frac{d\Phi}{dr} \rightarrow -\frac{32\pi^2}{\tilde{\Gamma}_\phi^4} \frac{\sigma^2 \rho_c}{r_0} \left[\ln\left(\frac{r}{r_0}\right) \right]^3 \left(\frac{r}{r_0}\right)^{-3} , \tag{40}$$

and

$$F_{mag} \equiv -\frac{1}{r^2}\frac{d}{dr}\left(\frac{r^2 B_\phi^2}{8\pi}\right) - \frac{d}{dr}\left(\frac{B_z^2}{8\pi}\right) \rightarrow -\frac{16\pi}{\tilde{\Gamma}_\phi^2} \frac{\sigma^2 \rho_c}{r_0} \ln\left(\frac{r}{r_0}\right) \left(\frac{r}{r_0}\right)^{-3} , \tag{41}$$

so that the ratio is given by

$$\frac{F_{grav}}{F_{mag}} \rightarrow \frac{2\pi}{\tilde{\Gamma}_\phi^2} \left[\ln(r/r_0)\right]^2 , \tag{42}$$

which diverges with radius, but only logarithmically. Therefore, gravity must always dominate over the magnetic field at large r/r_0. However, F_{grav}/F_{mag} remains less than unity until $r/r_0 > 53.3$, using the full expression for the ratio, rather than the limit, and typical values of $\tilde{\Gamma}_\phi = 15$ and $\tilde{\Gamma}_z$=5 [20]. If $\tilde{\Gamma}_z$ is increased to 7.5, the ratio becomes equal to unity when $r/r_0 = 15.3$. Note that both of these values exceed the truncation radius for most of the models shown in Fig. 3. Therefore, most filaments should have magnetically dominated outer envelopes within their finite truncation radii.

7 The Dynamics and Radial Stability of Filamentary Molecular Clouds

7.1 Radial Equation of Motion

The virial analysis of Section 5 applies only to static filamentary clouds. In this section, we extend our analysis to study the radial stability of filamentary

clouds and their evolution when out of radial equilibrium. We begin with the full virial equation, including all surface terms, and allow for the gas to be out of equilibrium radially:

$$\frac{1}{2}\ddot{\mathcal{I}} = 2\int P\ d\mathcal{V} - 2P_S\mathcal{V} - m^2G + \frac{1}{4\pi}\int B_z^2\ d\mathcal{V} - \frac{1}{4\pi}\left(B_{z,S}^2 + B_{\phi,S}^2\right)\mathcal{V} \quad (43)$$

where $\mathcal{V}$ is the cross-sectional area, $\mathcal{I}$ is the moment of inertia per unit length defined by

$$I = \frac{1}{2}amR_S{}^2 \quad (44)$$

and the dots represent a time derivative. In this equation, R_S is the radius of the filament and a is a numerical factor of order unity, which accounts for the deviation of the filament from uniformity. We assume that $m < m_{crit}$, so that there exists an equilibrium state (see Sect. 5.1). We assume a polytropic equation of state

$$\frac{\langle P\rangle}{P_0} = \left(\frac{\langle\rho\rangle}{\rho_0}\right)^{\gamma}, \quad (45)$$

in which ρ_0 and P_0 are fiducial values, which we set equal to the mean volume weighted density and pressure of the gas in its equilibrium state. We assume that the velocity dispersion σ within the filament is spatially constant but allowed to vary with time. We denote the central density, pressure, virial line mass, and cross-sectional area of the filament in its equilibrium state as ρ_c, P_c, m_{vir}, and $\mathcal{V}_0$ respectively. These quantities serve as constant reference values in the following analysis, which applies to non-equilibrium states.

Following Section 6.1, we assume that the poloidal and toroidal flux to mass ratios defined by (21) and (22) are constant throughout the cloud. Flux freezing guarantees that both quantities are also conserved when a filament contracts or expands radially. We simplify our analysis by assuming a radial velocity field of the form $v_r \propto r$. This restricted velocity field results in a simple rescaling of the radial structure of a filament, so that a in (44) is constant. It can also be shown from the induction equation of perfect MHD that the poloidal current I associated with the toroidal field, given by (20), is conserved when the velocity field is of this form. Thus, the toroidal field at the surface of the filament scales as $B_{\phi,S} \propto {R_S}^{-1} \propto \mathcal{V}^{-1/2}$:

$$B_{\phi,S} = B_{\phi S,0}\left(\frac{\mathcal{V}}{\mathcal{V}_0}\right)^{-1/2}, \quad (46)$$

where $B_{\phi S,0}$ is the equilibrium value of the toroidal field at the surface of the filament. This scaling relation is used to compute the toroidal component of the magnetic surface term in (43). We further assume that the poloidal field at the surface of the filament is constant: $B_{z,S} = B_{zS,0}$, where $B_{zS,0}$ is the equilibrium strength of the poloidal field bounding the filament in equilibrium.

After some straightforward algebraic manipulations, and with the help of (44), (45), and (46), (43) can be re-written in the following form, where we have

preserved the order of the terms for easy comparison:

$$\frac{1}{2}\tau^2\frac{d^2}{dt^2}\left(\frac{\mathcal{V}}{\mathcal{V}_0}\right) = 2\left(\frac{\mathcal{V}}{\mathcal{V}_0}\right)^{1-\gamma} - 2\left(\frac{P_S}{P_0}\right)\left(\frac{\mathcal{V}}{\mathcal{V}_0}\right) \\ -2\frac{m_{vir}}{m} + \frac{\tilde{\Gamma}_z^2}{4\pi}\left(\frac{\rho_0}{\rho_c}\right)\left(\frac{\langle\rho^2\rangle}{\langle\rho\rangle^2}\right)\left(\frac{\mathcal{V}}{\mathcal{V}_0}\right)^{-1} \\ -\frac{\tilde{\Gamma}_z^2}{4\pi}\left(\frac{\rho_0}{\rho_c}\right)\left(\frac{P_S}{P_0}\right)\left(\frac{\mathcal{V}}{\mathcal{V}_0}\right) - \frac{2\tilde{\Gamma}_\phi^2}{\pi}\frac{m}{m_{vir}}, \tag{47}$$

where the timescale τ is defined by

$$\tau \equiv \left(\frac{am}{\pi P_0}\right)^{1/2} = a^{1/2}\frac{R_{s0}}{\sigma}. \tag{48}$$

Note that all quantities in (47), except for $\mathcal{V}$, are constants.

7.2 The Radial Stability of Filamentary Molecular Clouds

We now examine the radial stability of polytropic filamentary molecular clouds by perturbing (47) about equilibrium. This differs from the Bonnor-Ebert stability analysis in Sect. 4 and Sect. 6.2, which involves only quasi-static radial perturbations. We assume that

$$\frac{\mathcal{V}}{\mathcal{V}_0} = 1 + \frac{\delta\mathcal{V}}{\mathcal{V}_0}, \tag{49}$$

where $\delta\mathcal{V} << \mathcal{V}_0$. To first order in $\delta\mathcal{V}/\mathcal{V}_0$, fluctuations about the equilibrium state obey the linear equation

$$\frac{1}{4}\tau^2\frac{d^2}{dt^2}\left(\frac{\delta\mathcal{V}}{\mathcal{V}_0}\right) + \left\{\gamma - 1 + \frac{P_S}{P_0} + \frac{\tilde{\Gamma}_z^2}{8\pi}\left(\frac{\rho_0}{\rho_c}\right)\left[\frac{\langle\rho^2\rangle}{\langle\rho\rangle^2} + \frac{P_S}{P_0}\right]\right\}\left(\frac{\delta\mathcal{V}}{\mathcal{V}_0}\right) = 0. \tag{50}$$

The cloud is stable against radial perturbations when the term in curly braces ({}) is greater than zero, and unstable to radial collapse or re-expansion into the surrounding gas otherwise.

Equation 50 demonstrates that all isothermal ($\gamma = 1$), pressure-bounded, equilibrium filaments are stable against radial perturbations. Likewise, all equilibrium filaments that are supported by Alfvén waves, which respond to compression and rarefaction as a $\gamma = 3/2$ polytrope [40], are stable. In fact, radial instability is only possible for equilibria with polytropic index $\gamma < 1$, and even these become stable if they are sufficiently magnetized or bounded by a sufficiently high external pressure. The fact that the trapped poloidal field helps to stabilize filaments is relatively intuitive, since B_z and the corresponding magnetic pressure increase when the filament is compressed and decrease when the filament expands, providing a net restoring force. Note that neither the gravitational energy term nor the toroidal field surface term enter into (50), since these terms do not scale with $\mathcal{V}$ in (47), and hence provide no restoring force.

We have shown that most filamentary clouds have equilibrium states (as long as $m < m_{vir}$) that are extremely stable to radial perturbations. Once a cloud attains such an equilibrium state, it can survive without collapsing nor re-expanding until longitudinal instabilities eventually break the filament into a string of clumps or possibly self-gravitating cores (see Sect. 9). Since (47) is dissipationless, stable solutions that begin out of equilibrium will tend to oscillate about equilibrium. In reality, such a filament would tend to settle into equilibrium as oscillations damp due to dissipation in shocks that form during the collapse, or other mechanisms involving the dissipation of turbulence or Alfvén waves.

Filamentary clouds might form when some damping process operates on a radially collapsing filament of gas that is initially out of equilibrium. This initial non-equilibrium could be magnetic in nature [9,8]. If some process in the turbulent molecular gas or the ISM generates a sufficient amount of vorticity, field lines could be sufficiently wound up to trigger a magnetically-driven collapse. This would tend to favour the development of filaments since the toroidal field produced by such winding would tend to propagate along poloidal field lines as torsional Alfvén waves. A similar mechanism was proposed by Habé et al. [27], who studied cloud collapse triggered by torsional Alfvén waves. In this model, the toroidal field of an impinging external torsional wave squeezes the cloud. The extra magnetic stress adds to the external pressure, effectively driving the cloud past the stability threshold given by (6).

8 The Global Topology of the Magnetic Field in Filamentary Clouds

The electrical current is not usually considered in MHD, where it plays the role of a secondary, and usually unnecessary, quantity that can be derived from the magnetic field. However, it is useful in the present discussion to consider the distribution of current in relation to the overall topology of the field. If the clouds contain a significant toroidal field and are in steady or quasi-steady state, then it is obvious that there must be a return current present outside of the filament, equal in magnitude to the total current associated with the toroidal field, in order to close the circuit. There are two distinct possibilities for the topology of the return current. This return current might flow entirely in the vicinity of an isolated filament, or a filament might be just one element residing within a much larger circuit, which could possibly contain other filaments. We address each of these possibilities in turn.

8.1 Isolated Filaments

It is evident in (14) and (17) that the toroidal field contributes to the net confinement of a filamentary cloud if and only if B_ϕ has a finite value at the surface of the cloud. If clouds are magnetically confined, then (20) shows that the return current must be external to the region of confinement. If the current returns as

a thin current sheet flowing along the surface of the filament or interior to the filament, then the toroidal field at the surface would be nullified, and so would its confining effects. Thus, the magnetic confinement required by the FP model requires the return current to be diffuse and extended throughout the surrounding gas, as in the case of protostellar jets (Ouyed & Pudritz 1997),

8.2 Current Networks and Open Circuits

If a helically magnetized filament resides within a larger circuit of currents, the return current will naturally take a more extended path than in the case of an isolated filament, and there will be no problem with the return current nullifying the magnetic confinement. This possibility is intriguing if magnetic stresses are partially responsible for the formation of filaments, in which case filaments might preferentially contract to equilibrium along current pathways in the ISM or within a more extended molecular cloud. The equations of Sect. 7.1 would apply to such a collapse. Note that a similar idea was first discussed by Carlqvist [9,8].

We conclude this section by returning to a more standard MHD viewpoint. Regardless of whether a helically magnetized filament is isolated or part of a larger network, twisting motions in the turbulent ISM or molecular cloud must be responsible for winding up the toroidal component of the field. We suggest that magnetic stresses resulting from such winding could play a role in the assembly of some filaments.

9 Fragmentation of Filaments

It is well-known that many filamentary clouds have roughly periodic density enhancements along their lengths, which might be the result of fragmentation from a more uniform filamentary cloud. Periodically spaced condensations have been noted by Dutrey et al. [16] and Johnstone & Bally [30] in the Orion A integral-shaped filament. Periodicity is especially evident in many of the relatively quiescent globular filaments shown in the Schneider & Elmegreen catalogue of dark globular filaments [54].

Self-gravitating filaments are always gravitationally unstable to fragmentation, in which a longitudinally uniform filament breaks into a string of clumps that are roughly periodic in separation. Chandrasekhar & Fermi [12] first demonstrated that uniform, incompressible filaments threaded by a purely poloidal magnetic field are unstable to fragmentation driven by self-gravity. A number of other more recent authors have performed similar calculations to analyze the stability of more complex magnetized filaments. For example, Nagasawa [47] considered the fragmentation of compressible, pressure-bounded filaments threaded by a poloidal field and embedded in a hot external medium with finite pressure but zero density. Nakamura, Hanawa, and Nakano [48] determined the stability properties of unbounded filaments threaded by a helical field. Gehman et al. [26] analyzed the fragmentation of magnetized logatropic filaments, which

are characterized by an equation of state that is softer than isothermal (see also McLaughlin & Pudritz [41]). FP2 analyzed the stability of their pressure-truncated, helically magnetized model, which they assumed to be embedded in a magnetized external medium with a finite density characteristic of the ISM or the molecular gas surrounding typical filaments.

It is worth stressing that the FP model differs from the equilibria used in the other analyses mentioned above because their model is characterized by an $r^{-(1.9\pm0.1)}$ density profile, in agreement with the observations [1,33,30]. The underlying equilibrium in the Nagasawa [47] and Nakamura et al. [48] calculations have steep r^{-4} density profiles, like the Ostriker [49] solution discussed in Sect. 4.

9.1 Equations and Boundary Conditions

The following eigensystem is a linear system of equations that applies to small perturbations about a stationary equilibrium state in self-gravitating MHD [20]:
Momentum Equation:

$$\begin{aligned} -\omega^2\rho_0\mathbf{v}_1 = {} & \gamma\frac{P_0}{\rho_0}\nabla\left[\nabla\cdot(\rho_0\mathbf{v}_1)\right] \\ & + \frac{1}{4\pi}\left\{(\nabla\times\mathbf{B_0})\times\left[\nabla\times\left(\rho_0\mathbf{v}_1\times\frac{\mathbf{B_0}}{\rho_0}\right)\right]\right. \\ & \left. + \left[\nabla\times\nabla\times\left(\rho_0\mathbf{v}_1\times\frac{\mathbf{B_0}}{\rho_0}\right)\right]\times\mathbf{B_0}\right\} \\ & - \rho_0\nabla\cdot(i\omega\Phi_1) + \nabla\Phi_0\nabla\cdot(\rho_0\mathbf{v}_1) \end{aligned} \tag{51}$$

Poisson's Equation + Mass Conservation:

$$0 = \nabla^2(i\omega\Phi_1) + 4\pi G\nabla\cdot(\rho_0\mathbf{v}_1). \tag{52}$$

In these equations, all quantities with subscript "0" refer to the underlying equilibrium state, while all quantities with subscript "1" refer to the perturbed quantities, which are all assumed to be of the form

$$f_1(r,z,\phi,t) = f_1(r)e^{i(\omega t+m\phi+k_z z)}, \tag{53}$$

where f_1 represents any of the perturbed quantities in (51), and $m = 0$ for axisymmetric perturbations. Note that ω is the angular frequency of the perturbation (growthrate of an unstable perturbation) and k_z is the wave number.

These equations were solved by FP1 for the coupled system comprised of a self-gravitating filamentary molecular cloud embedded in a non-self-gravitating external medium, as described above. Self-gravity was effectively turned off in the external medium by setting $G = 0$ outside of the filament ($r > R_S$). The filament was assumed to be isothermal ($\gamma = 1$), but γ was allowed to take on a different value in the external medium ($r > R_S$). The usual boundary conditions of self-gravitating MHD connect the material inside of the filament to the external medium. These are outlined in FP2 and explained in detail in many standard textbooks and references. The eigensystem given in (51) and (52), with appropriate boundary conditions, were solved numerically by FP2 using a finite difference method.

9.2 Results of the Stability Analysis

The main findings of the stability analysis in FP2 are as follows.

1) Filaments threaded by poloidal or helical magnetic fields, and unmagnetized filaments, are always unstable to fragmentation.

2) Decreasing the line mass increases the timescale for fragmentation, but the effect is not large. For example, the growth rate decreases by only a factor of ~ 1.5 when the line mass decreases from the critical value m_{vir} to $m_{vir}/10$.

3) Poloidal fields also help to stabilize filaments, but the effect is moderate. For a filament with a mass per unit length of $m_{vir}/5$, the growth rate can be decreased by at most a factor of ≈ 1.9 by increasing the poloidal field. Past this point, the effect saturates and no further stabilization is possible.

4) Toroidal fields can dramatically increase the stability of filamentary clouds against axisymmetric modes of gravitational fragmentation. When condensations begin to form along the filament, toroidal flux loops must be squeezed together. However, they repel each other, thus resisting the compression and slowing the development of the instability. Models become increasingly stable as $\tilde{\Gamma}_\phi/\tilde{\Gamma}_z$ increases from zero, until the most stable range is reached at $\tilde{\Gamma}_\phi/\tilde{\Gamma}_z \approx 2$. For higher values of this ratio, magnetically driven axisymmetric sausage instabilities are triggered, which become increasingly unstable as $\tilde{\Gamma}_\phi/\tilde{\Gamma}_z$ increases past 2. This is illustrated in Fig. 4 below. For these models, the instability develops on a timescale given by

$$\tau_{frag} \approx 1.8 \left(\frac{-\tilde{\omega}^2_{max}}{0.01} \right)^{-1/2} \left(\frac{n_c}{10^4 \ cm^{-3}} \right)^{-1/2} \ Myr. \tag{54}$$

5) The Schneider & Elmegreen Catalogue of Dark Globular Filaments [54] shows that the periodicity of cores in globular filaments is usually on the order of 3 times the filament's diameter, with a considerable amount of scatter. This is consistent with the wavelengths of the fastest growing modes of instability of models with $\tilde{\Gamma}_\phi/\tilde{\Gamma}_z \approx 2$.

Simulations usually do not show the obvious periodic fragmentation seen in globular filaments, the Orion integral-shaped filament, and other filaments. Filamentary structures that form in simulations generally do not survive long enough for gravitational instabilities to break them up, before re-expanding into the surrounding turbulence. This difference between existing turbulence simulations and observations of dense fragmenting regions suggests that some real filaments may be stable against radial perturbations (collapse, re-expansion, etc.) on timescales longer than τ_{frag}. Note that this radial stability is predicted by the analyses in Sect. 6.2 and Sect. 7.2, and is a key characteristic of the FP model.

9.3 Dynamical Consequences of Fragmentation

The dynamics accompanying the fragmentation of a filament threaded by a purely poloidal magnetic field are relatively simple. Gas essentially flows along

the filament into the forming condensations. These may bulge slightly outward in the linear phase of evolution as the gas accumulates. Examples of such modes are shown in Nagasawa et al. [47], Nakamura et al. [48], and FP2. The dynamics become somewhat more complex when the field is helical. In this case, the gas experiences torque due to the magnetic field when it begins to flow parallel to the axis of a filament toward a condensation, which causes it to rotate about the axis. The direction of rotation is in the sense that would tend to unwind the helix as gas flows toward the condensation from both sides. Therefore, the direction of rotation changes sign on opposite sides of a forming clump. A filament breaking into multiple clumps would have a striking kinematic signature dominated by flow parallel to the filament axis toward the condensations, infall onto the condensations, and rotational motions that alternate direction with the same wavelength as the core spacing. This kinematic pattern is an important prediction of the FP model, and is a characteristic feature of helical field models in general.

The first evidence for this kinematic signature was recently found by Falgarone, Pety, & Phillips [18], who observed the dense starless core L1512 and some of the nearby filaments in ^{12}CO. The filaments, which appear to converge on the core, are only ~ 1 pc long and ~ 0.1 pc in diameter, and are thus smaller than the filaments studied by FP1. Two of these filaments were found to have strong transverse velocity gradients that alternate periodically with a wavelength of $\approx 0.28\ pc$ along the length of the filament. Falgarone et al. [18] interpreted their data as periodically alternating rotation, as predicted by FP2. This structure was found to persist for two full wavelengths in one of the filaments, and a sinusoidal curve was found to reasonably fit the data. They determined that the wavelength is about 3 times the diameter of the filaments. This allowed them to infer from Fig. 4 that $\tilde{\Gamma}_\phi/\tilde{\Gamma}_z \approx 2$. This value represents the most stable range of filaments among an ensemble of all models that were studied by FP2.

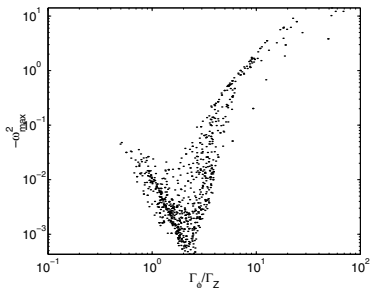


Fig. 4. (From FP2) The growth rates for a random sample of models that agree with the constraints in FP1. Note the highly stable regime near $\tilde{\Gamma}_\phi/\tilde{\Gamma}_z \approx 2$. Modes with $\tilde{\Gamma}_\phi/\tilde{\Gamma}_z \lesssim 2$ are gravity-driven, while modes with $\tilde{\Gamma}_\phi/\tilde{\Gamma}_z \gtrsim 2$ are driven mainly by MHD sausage mode instabilities. Note that this figure is reprinted from FP2, where tildes were assumed on Γ_z and Γ_ϕ.

Falgarone et al. [18] also demonstrated that the two filaments that they studied have virial properties consistent with a magnetically confined filament. The line mass of the filament was found to be 1.6 $M_\odot\ pc^{-1}$, and the authors calculated the virial line mass m_{vir} (see (10)) to be $\sim$ 7.5 $M_\odot\ pc^{-1}$. Thus, $m/m_{vir} \approx 0.2$, which is in the range of values found by FP1. Falgarone et al. [18] also estimated the external pressure, including turbulent ram pressure of the surrounding ISM, to be $\lesssim 10^4\ K\ cm^{-3}$, and the internal pressure to be in the range $(4.6 - 7) \times 10^4\ K\ cm^{-3}$. Thus, $P_S/\langle P\rangle \lesssim (0.14 - 0.22)$, which is also typical of the values found by FP1. It is clear that $P_S/\langle P\rangle < 1 - m/m_{vir}$. This indicates that the pressure within the filaments is too high for the gravitational binding energy to hold them together. The authors ruled out radial expansion and attributed the extra binding energy to a toroidal field. Solving for $\mathcal{M}/|\mathcal{W}|$ in (14), we find that $\mathcal{M}/|\mathcal{W}| \lesssim -2.7\ to\ -3.0$. Therefore, the virial theorem suggests that this filament is magnetically confined by the toroidal component of a helical field.

The combination of virial and kinematic evidence makes this system of filaments an intriguing source for future study. A detailed polarization map might help to check the authors' interpretation of these filaments as candidates for the FP model. Polarimetry is discussed in Sect. 11 below.

10 The Shapes of Cores and Bok Globules

10.1 Prolate, Oblate, or Triaxial?

There is some evidence that cores may be predominantly elongated objects that are approximately prolate in shape. The evidence comes from two different lines of reasoning. The first is a statistical analysis of the shapes of cores, which was first done by Myers et al. [44]. Their analysis showed that the mean projected axis ratio $\langle p\rangle$, defined as the ratio of the minor axis to the major axis, is in the range 0.5-0.6 for cores, which corresponds to an intrinsic axis ratio of 0.4-0.5. This finding was confirmed by Ryden [52], who performed a much more elaborate statistical analysis on several large data sets of cores and Bok globules. Her analysis ruled out the hypothesis of oblate cores at the 99% confidence level for some data sets. As discussed by Ryden [52], a lack of nearly circular objects (in projection) is the telltale sign that a population cannot be comprised of randomly oriented oblate spheroids; indeed the Myers et al. [44] data set does not contain any cores with $\langle p\rangle > 0.89$. Interestingly, this finding was contradicted by Jones, Basu, & Dubinski [32], who found that cores are intrinsically triaxial, but closer to oblate than prolate.

Statistical arguments to determine the shapes of cores all suffer from the same underlying limitation. The statistical determination of intrinsic axis ratios from projected axis ratios implicitly assumes that cores are oriented randomly in any given sample. This is probably not generally true, since many cores and Bok globules are located within filaments (e.g. [54,16,44,36]), or at least appear to

be distributed in approximately linear chains (e.g. [42,31]). In either case, their orientations may not be independent.

The second argument for the existence of prolate cores is based on the frequent association of elongated cores with filaments. Of the 16 cores studied by Myers et al. [44], 6 were observed to be embedded within long filamentary structures, and with their long axes aligned with the filament. Ryden [52] argued that these filamentary structures, with typical axis ratios of $\langle p \rangle = 0.1$, must truly be filaments and not edge-on sheets, because sheets would need to be viewed from a highly improbable, almost exactly edge-on orientation to be seen with a projected axis ratio this low. The alignment of elongated, embedded cores with their parent filaments strongly suggests that they are intrinsically prolate. There are many other examples of such apparently prolate, embedded cores in the Schneider & Elmegreen catalogue of Dark Globular Filaments [54].

10.2 Theoretical Models

It is quite difficult to construct an equilibrium hydrostatic or magnetostatic model of a prolate core. Isolated hydrostatic cores relax to spherical equilibria (Bonnor-Ebert [6,17] spheres), while cores threaded by poloidal fields relax to a flattened, oblate shape [43,58,59,60,61], which are difficult to resolve with the observations. Three very different models of prolate cores have been proposed, which we discuss below.

Helical Fields and Prolate Cores. Fiege & Pudritz [22] (hereafter FP3) proposed a conceptually simple model of prolate cores, based on the idea that these objects relax from the condensations that form as helically magnetized filamentary clouds fragment (See Sect. 9) and relax to equilibrium. The basic idea is that both the poloidal and toroidal flux is inherited from the parent filament, and together shape the core. The toroidal field both squeezes the gas radially, and supports it against self-gravity along the symmetry axis of the filament, while the poloidal field offers only radial support. Cores that are toroidally dominated therefore tend to have a prolate shape.

All of the models derived from FP filaments were found to be at least slightly prolate, with intrinsic axis ratios ranging from $0.1 - 0.9$ and projected axis ratios ranging from $0.2 - 1$. Both field components were of similar strengths in the parent filament [4], but after relaxing to equilibrium, $M_{pol} > M_\phi$ for most models, where

$$M_{pol} \equiv \frac{1}{8\pi} \int (B_r^2 + B_z^2) dV$$
$$M_\phi \equiv \frac{1}{8\pi} \int B_\phi^2 dV, \tag{55}$$

[4] The maximum poloidal field is concentrated along the central spine of the filament and is generally stronger than the maximum toroidal field. The toroidal field generally dominates in the outer envelope, however.

and V is the volume. Nevertheless, the stresses exerted by the toroidal field are more effective at shaping the core since they are simultaneously compressive in the radial direction and expansive along the symmetry axis. Thus, prolate models can be produced with rather modest toroidal fields. Counter-intuitively, it is actually the models with the *lowest* values of M_ϕ/M_{pol} that are the most elongated. Generally, the less gravitationally dominated models in their parameter space exploration were found to be more elongated; projected axis ratios < 0.6 generally required $M/M_{B.E.} < 0.7$, where $M_{B.E.}$ is the Bonnor-Ebert critical mass defined in (1). FP3 also found that the critical mass is reduced by about 20% from the Bonnor-Ebert critical mass for most of their models.

The primary advantage of this model is that it is a natural consequence of the framework built by the FP filamentary cloud models. It can also be tested directly by polarimetry, since the magnetic structure of a helically magnetized core results in a very distinct polarization pattern. This pattern will be discussed in Fiege, Matthews, & Moriarty-Schieven [19].

Tidally Stretched Prolate Cores. Curry [14] published a hydrostatic model of prolate cores embedded in filamentary molecular clouds. In Curry's model, hydrostatic filaments fragment into periodic condensations, as discussed in Sect. 9. These fragments are tidally stretched by their neighbours as they evolve toward equilibrium, so that the final equilibrium is that of a periodic sequence of unmagnetized cores residing within a filament whose overall geometry is that of a hydrostatic Ostriker [49] filament.

One difficulty with this model, noted by the authors, is that the cores need to be spaced somewhat further in the model than they are in reality. Another difficulty is that the parent filament is an Ostriker filament with an r^{-4} density profile, which is steeper than observations allow (see Sect. 6.3).

One must also note that this hydrostatic model cannot be an accurate representation of filamentary clouds and their cores for the simple reason that these objects are known to be magnetized, and the magnetic fields are believed to be strong enough to be dynamically significant. Nevertheless, it would be interesting and worthwhile to try to extend this model to the slightly more complex case of tidally stretched magnetized cores embedded within a filament threaded by a poloidal field.

Note that the Curry [14] model relies on tidal stretching, which applies to multiple systems of embedded cores. This model does not explain isolated prolate cores, or filaments with a single embedded core. This is one important difference between the FP3 model and the Curry model, which could conceivably form the basis of an observational test.

Magnetic Confinement by Poloidal Field Tension. Curry & Stahler [15] recently published another class of prolate core models, in which the magnetic confinement is due the tension in a poloidal field that bends around the core, opposite to the conventional hourglass magnetic field. These models can result in prolate cores, but only if the magnetic field outside of the core is greater than

the mean field inside. Although these models are conceptually interesting, it is difficult to imagine how this field configuration could arise.

11 Sub-millimetre Polarimetry

11.1 What Can We Learn from Polarimetry?

Observations of polarized continuum emission at far infra-red and sub-millimetre wavelengths provide the best available method for studying the geometry of the magnetic field in molecular clouds. At these wavelengths, the polarized flux is due to the thermal emission of dust grains, which are predominantly non-spherical and partially aligned with their long axes orthogonal to the local field direction (see reviews in Hildebrand et al. [29] and Weintraub, Goodman, & Akeson [65]). Each grain along a given line of sight through the cloud contributes polarized emission parallel to the projection of its long axis in the plane of the sky, and therefore perpendicular to projection of the field $\mathbf{B}_\perp$ on the sky. Thus, polarization maps provide an indication of the mean field direction in the plane of the sky, which is appropriately weighted by the number density of grains, grain temperatures and cross sections, geometrical terms related to the field orientation, and efficiency factors that account for the degree of alignment and the randomization of the grain orientations by turbulence [62,64].

Polarimetry does not provide any information pertaining to the component of the magnetic field $B_\parallel$ along the line of sight. Thus, the full 3-dimensional field structure cannot be obtained directly from the data. Interpretations of the underlying 3-dimensional field from polarization maps suffer from a certain amount of degeneracy resulting from this incomplete information. Nevertheless, they can provide useful and challenging tests to theoretical models, since they may provide enough information to reject models that cannot be reconciled with the data. Zeeman measurements, combined with polarization maps, provide an opportunity to map the full 3-dimensional field structure, since Zeeman measurements obtain $B_\parallel$ directly. However, these measurements are very difficult and cannot yet be used to create well-sampled maps of the line of sight component of the magnetic field. Future technological advances might improve our ability to make Zeeman maps and help to resolve this degeneracy.

Another limitation of polarimetry is that it does not *directly* provide any information pertaining to the field strength. However, the field strength can be estimated using the classical technique of Chandrasekhar & Fermi [12], which is based on the dispersion of polarization directions about the mean. Recent corrections to this technique have been published by Padoan et al. [51] and Heitsch et al. [28], who have demonstrated its effectiveness, using MHD turbulence simulations, as a field strength estimator in regions dominated by turbulence.

At present, well-sampled polarization maps at sub-millimetre wavelenths have generally been limited to dense filaments and cores in molecular clouds. Existing detectors do not have the wide-field capabilities and high sensitivities needed to map entire molecular clouds or relatively low density cloud material on scales larger than cores. However, the maps that have been obtained of

dense, star-forming regions are fascinating. For examples, see Schleuning [55], Matthews & Wilson [37], Matthews, Wilson & Fiege [38], and Matthews, Fiege, & Moriarty-Schieven [36]. They generally show that sources are quite strongly polarized, with a maximum polarization percentage of typically $10-15\%$. Most maps that have been obtained to date suggest that dense regions of molecular clouds are threaded by an ordered magnetic field that is geometrically quite complex. There are no examples that are consistent with uniform magnetic fields.

It is not possible at present to model the polarization pattern expected for a given distribution of density and magnetic field from first principles, since many of the grain properties and the details of their alignment are not well constrained [34]. However, Fiege & Pudritz [23] (hereafter FP4) provide a simple heuristic technique to generate such models by combining all of the unknown grain properties into a single parameter, which can be constrained observationally. This technique was used by Matthews, Wilson, and Fiege [38] to model the polarization structure of the integral-shaped filament in Orion A, by Matthews, Fiege, and Moriarty-Schieven [36] to model the polarization of the NGC 2024 system of cores in Orion B, and by Padoan et al. [51] to synthesize polarization maps from self-gravitating cores formed by their simulations of MHD turbulence in molecular clouds. A related technique has been used by Heitsch et al. [28] to model the polarization patterns expected from their turbulence simulations.

11.2 Polarization Maps of the Fiege & Pudritz Model

Carlqvist & Kristen [10] first considered the polarization characteristics of helically magnetized filaments. Their calculation was specifically performed for absorption polarimetry at optical wavelengths, using a substantially different helical field model, but some of their geometrical arguments apply to our case as well. They demonstrated that the polarization vectors are always aligned either parallel or perpendicular to the filament, as long as the optical depth is low. The reason for this behaviour is that any line of sight always passes through any given radius, or equivalently any given flux surface, twice because of the symmetry of the filament. By symmetry, the magnetic field vectors, projected on the plane of the sky, are mirror images at these intersection points, regardless of the inclination of the filament. Therefore, grains located at these points appear tilted at equal but opposite angles with respect to the projected axis of the filament. At low optical depths, the contributions from each of these grains to the Stokes vectors partially cancel so that only the dominant component, either parallel to the filament or perpendicular, survives. Note that this becomes exact in the limit of zero optical depth. The same argument applies at submillimetre wavelengths, where the polarization is caused by emission rather than absorption, and absorption is negligible.

Figure 5 a) shows the Matthews, Wilson, & Fiege [38] map of the integral-shaped filament in the Orion Molecular Cloud, which was re-reduced using improved software from the Matthews & Wilson [37] data set. The polarization vectors are very nearly parallel to the filament at the northern end, which is

consistent with a FP model with a toroidally dominated magnetic field. However, that the polarization vectors bend substantially in the southern part of the filament. This is not consistent with the basic FP model, because of the symmetry argument given above. However, a slight extension to the model can account for this discrepancy.

Figure 5 b) compares the map to a predicted polarization map generated from a FP filament model, whose parameters are given in the caption. Note that the model has been slightly modified by gradually bending the southern end of the filament in a circular arc. This has the effect of breaking the symmetry discussed above, which allows the polarization vectors to become misaligned with respect to the filament. The density and magnetic field are transformed self-consistently by making use of the Lagrangian forms of the mass conservation equation and the induction equation of MHD:

$$\frac{\mathbf{B}}{\rho} = \mathbf{J}\frac{\mathbf{B_0}}{\rho_0}$$
$$\rho = \frac{\rho_0}{|\mathbf{J}|}, \tag{56}$$

where subscript "0" refers to the original unbent FP filament, and $\mathbf{J}$ is the Jacobian of the coordinate transformation responsible for the bend. We observe that this model, when viewed from an appropriate range of angles, reproduces most of the qualitative features of the polarization map.

An interesting and important feature of the map in Fig. 5 a) is that it appears to be depolarized toward the regions of brightest emission. This polarization hole effect is usually seen in polarization maps at sub-millimetre wavelengths, and has frequently been explained by postulating that a grain population exists at high optical depths, whose emission is only weakly polarized because the grains are either nearly spherical or poorly aligned with respect to the magnetic field. Alternatively, geometric factors caused by helical fields can cause the depolarization. This occurs because the greatest cancellation in the Stokes vectors, discussed above, usually occurs along the projected axis of toroidally dominated models, where the poloidal field is the strongest. Examples of strong geometric depolarization are shown in FP4. It is possible that geometric depolarization and poorly polarizing grains might work in concert to explain the polarization hole effect. We note that the model shown in Fig. 5 b) is not sufficiently depolarized in the central regions to account for the data, and might be improved by including poorly polarizing grains in the central regions.

12 Discussion

The primary aim of this paper was to review the FP helical field model of filamentary molecular clouds. In the preceding sections, we have discussed their virial properties (Sect. 5), radial structure (Sect. 6), dynamics(Sect. 7), stability (Sect. 7.2), fragmentation (Sect. 9), cores(Sect. 10), and polarimetry observations in support of the model (Sect. 11). The FP model agrees with the virial

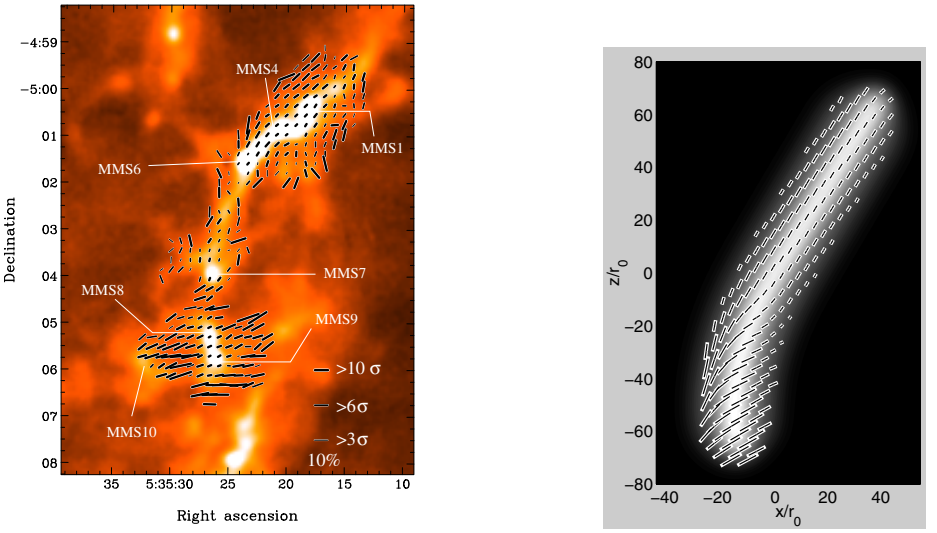


Fig. 5. (From [38]) a) (left) 850 μm polarization pattern across OMC-3 superimposed on a portion of the 850 μm intensity map of Johnstone & Bally [30]. The thinnest vectors have a signal-to-noise in polarization percentage, $\sigma_p > 3$, while the medium thickness vectors have $\sigma_p > 6$. Most vectors are bold and have $\sigma_p > 10$. b) (right) A qualitative model of the polarization pattern produced by a FP model with parameters given by $C = 1.2$, $\tilde{\Gamma}_z = 13$, and $\tilde{\Gamma}_\phi = 18$. The southern half of the filament has been bent into an arc with a radius of curvature of 6.4 times the filament's diameter. The whole filament is inclined to the plane of the sky at an angle of 20° and then rotated in the plane of the sky by 225°. The bending of the filament breaks the symmetry of the models presented in FP4 causing the polarization pattern from the inner region of the bend to dominate, as discussed in the text.

data for filaments, the radial density structure, the fragmentation properties of filaments, the predominance of prolate cores in filamentary clouds, and existing polarization maps. We have also demonstrated that pressure-truncated, self-gravitating, helically magnetized filaments are extremely stable against radial perturbations. Filaments can oscillate radially about their equilibrium state, but they cannot collapse radially nor expand into the surrounding gas. If processes occurring within the ISM or molecular clouds manage to generate such a filament, it will probably survive until gravitational fragmentation breaks it into a roughly periodically spaced sequence of cores. This is especially likely in relatively quiescent environments, where filaments are less prone to being sheared apart by the surrounding turbulence.

Future tests of the FP model will rely heavily on polarimetry mapping. FP4 have provided a technique to model polarization patterns, which can then be compared directly to the observations. The main difficulty with polarimetry is that it only provides information related to the component of the field that is parallel to the plane of the sky. The line of sight component cannot be obtained, which makes the interpretation of the field structure non-unique. There are two

possible solutions to this problem. At some point in the future, it may be possible to make well-sampled Zeeman maps, which could be combined with the polarization data to provide a very detailed 3-dimensional picture of the field. A more immediate solution is to focus on multi-wavelength polarization studies. In particular, if polarization maps can be made at a sequence of wavelengths, including both long wavelengths where the gas is optically thin, and shorter wavelengths where the optical depth rises to order unity, we would be able to probe the orientation of the field at various optical depths. The modeling technique presented in FP4 could be readily extended to take this into account, and a detailed comparison with the observations would provide a rigorous observational test.

Acknowledgements

This work was supported by the National Research Council of Canada and fellowships from CITA and the Natural Sciences and Engineering Research Council of Canada. The author wishes to thank Edith Falgarone and Jacques Vallée for useful comments, and an anonymous referee for a very thorough review of a preliminary version of this paper.

References

1. Alves J., Lada C.J., Lada E.A., Kenyon S.J., Phelps R., 1998, Ap.J., 506, 292
2. Arons J., Max C.E., 1975, Ap.J., 196, L77
3. Ballesteros-Paredes J., Vazquez-Semadeni E., Scalo J., 1999, Ap.J., 515, 286
4. Bally J., 1989, in Proceedings of the ESO Workshop on Low Mass Star Formation and Pre-main Sequence Objects, ed. Bo Reipurth; Publisher, European Southern Observatory, Garching bei Munchen
5. Bally J., Stark A.A., Wilson, R.W., Langer, W.D., 1987, Ap.J., 312L, 45
6. Bonnor W.B., 1956, MNRAS, 116, 351
7. Boulares A., Cox D.P., 1990, Ap.J., 365, 544
8. Carlqvist P., 1998, Ap.&S.S., 144, 73
9. Carlqvist P., Gahm G., 1992, IEEE Trans. on Plasma Science, vol. 20, no. 6, 867
10. Carlqvist P., Kristen H., 1997, A&A, 324, 1115
11. Caselli P., Myers P.C., 1995, Ap.J., 446, 665
12. Chandrasekhar S., Fermi E., 1953, Ap.J. 118, 116
13. Chromey F.R., Elmegreen B.G., Elmegreen D.M., 1989, Ap.J., 98, 2203
14. Curry, Charles L., 2000, Ap.J., 541, 831
15. Curry C.L, Stahler S.W., 2001, Ap.J., 555, 160
16. Dutrey A., Langer W.D., Bally J., Duvert G., Castets A., Wilson R.W., 1991, A&A, 247, L9
17. Ebert R., 1955, Z.Astrophys., 37, 217
18. Falgarone E., Pety J., Phillips T.G., 2001, Ap.J., 555, 178
19. Fiege J.D., Matthews B.C., Moriarty-Schieven G., 2002, in preparation
20. Fiege, J.D.; Pudritz, R.E., 2000, MNRAS, 311, 85 (FP1)
21. Fiege, J.D.; Pudritz, R.E., 2000, MNRAS, 311, 105 (FP2)
22. Fiege, J.D.; Pudritz, R.E., 2000, Ap.J., 534, 291 (FP3)
23. Fiege, J.D.; Pudritz, R.E., 2000, Ap.J., 544, 830 (FP4)

24. Fuller G.A., Myers P.C., 1992, Ap.J., 384, 523
25. Fuller G.A., Myers P.C., 1993, Ap.J., 418, 273
26. Gehman C.S., Adams F.C., Watkins R., 1996, Ap.J., 472, 673
27. Habe A., Uchida Y., Ikeuchi S., Pudritz R.E., 1991, PASJ, 43, 703
28. Heitsch F., Zweibel E.G., Mac Low M., Li P., Norman M.L., 2001, Ap.J., 561, 800
29. Hildebrand, Roger H., 1988, QJRAS, 29, 327,
30. Johnstone D., Bally J., 1999, Ap.J., 510, 49L
31. Johnstone, Doug; Fich, Mike; Mitchell, George F.; Moriarty-Schieven, G., 2001, Ap.J., 559, 307
32. Jones, C. E, Basu, S., Dubinski, J., 2001, Ap.J., 551, 387
33. Lada C.J., Alves J., Lada E.A., 1999, Ap.J. 512, 250
34. Lazarian, A.; Goodman, Alyssa A.; Myers, Philip C., 1997, Ap.J., 490, 273
35. Maddalena R.J., Morris M., Moscowitz J., Thaddeus P., 1986, Ap.J., 303, 375
36. Matthews B.C., Fiege J.D., Moriarty-Schieven G., 2002, Ap.J., 569 304
37. Matthews B.C., Wilson C.D., 2000, Ap.J., 531, 868
38. Matthews B.C., Wilson C.D., Fiege J.D., 2001, Ap.J., 562, 400
39. McCrea W.H., 1957, MNRAS, 117, 562
40. McKee C.F., Zweibel E., 1995, Ap.J., 440, 686
41. McLaughlin D.E., Pudritz R.E., 1996, Ap.J., 496, 194
42. Mitchell, George F.; Johnstone, Doug; Moriarty-Schieven, Gerald; Fich, Michel; Tothill, N. F. H., 2001, Ap.J., 556, 215
43. Mouschovias T., 1976, Ap.J., 207, 141
44. Myers P.C., Fuller G.A.,Goodman A.A., Benson P.J., 1991, Ap.J., 376, 561
45. Myers P.C., Goodman A.A., 1988, Ap.J., 326, L27
46. Myers P.C., Goodman A.A., 1988, Ap.J., 329, 392
47. Nagasawa M., 1987, Prog. Theor. Phys., 77, 635
48. Nakamura F., Hanawa T., Nakano T., 1993, PASJ, 45, 551
49. Ostriker J., 1964, Ap.J., 140, 1056
50. Ouyed R., Pudritz R. E., 1997, ApJ, 482, 712
51. Padoan P, Goodman A., Draine B.T, Juvela M., Nordlund A, Rognvaldsson O.E., 2001, Ap.J., 556, 1005
52. Ryden, B.S., 1996, Ap.J., 471, 822
53. Sanders, D. B.; Scoville, N. Z.; Solomon, P. M., 1985, Ap.J., 289, 373
54. Schneider S., Elmegreen B.G., 1979, Ap.J., 41, 87
55. Schleuning D.A., 1998, Ap.J., 493, 811
56. Stodólkiewicz J.S., 1963, Acta Astron., 13, 30
57. Spitzer L. Jr., 1978, *Physical Processes in the Interstellar Medium* (John Wiley & Sons: New York)
58. Tomisaka K., Ikeuchi S., Nakamura T., 1988a, Ap.J., 326, 208
59. Tomisaka K., Ikeuchi S., Nakamura T., 1988b, Ap.J., 335, 239
60. Tomisaka K., Ikeuchi S., Nakamura T., 1989, Ap.J., 341, 220
61. Tomisaka K., Ikeuchi S., Nakamura T., 1990, Ap.J., 362, 202
62. Lee H.M., Draine B.T., Ap.J., 1985, 290, 211
63. Vazquez-Semadeni E., Ostriker C., Passot T., Gammie C.F., Stone J.M., 2000, in *Protostars and Planets IV*, ed. Mannings V., Boss A.P., Russell, S.S. (Tucson: University of Arizona Press), p.3
64. Wardle M., Königl A., 1990, Ap.J., 362, 120
65. Weintraub D.A., Goodman A.A., Akeson R.L., 2000, in *Protostars and Planets IV*, ed. Mannings V., Boss A.P., Russell, S.S. (Tucson: University of Arizona Press), in press

Numerical Simulations of MHD Turbulence in Accretion Disks

Steven A. Balbus and John F. Hawley

Dept. of Astronomy, University of Virginia, Charlottesville, VA 22901

Abstract. We review numerical simulations of MHD turbulence. The last decade has witnessed fundamental advances both in the technical capabilities of direct numerical simulation, and in our understanding of key physical processes. Magnetic fields tap directly into the free energy sources in a sufficiently ionized gas. The result is that adverse angular velocity and adverse temperature gradients, not the classical angular momentum and entropy gradients, destabilize laminar and stratified flow. This has profound consequences for astrophysical accretion flows, and has opened the door to a new era of numerical simulation experiments.

1 Introduction

Magnetized, differentially rotating plasmas are subject to a powerful linear instability, whose maximum growth rate is a factor of $\sim 10^2$ per orbit (Balbus & Hawley [8]). This is the magnetorotational instability, or MRI. Despite the fact that planar Couette flow is exquisitely sensitive to nonlinear disturbances and flow lamina quickly breakdown into turbulence, decades of investigation have failed to find any comparable mechanism in a non-magnetized Keplerian disk. The problem is that local Coriolis forces are larger than the effective disruptive inertial force caused by the presence of shear. The result is that incompressible disturbances respond in a wavelike manner in a disk, whereas no such response is possible in non-rotating shear flow. Without rotation, displacements of perturbed fluid elements are (linearly) unbounded, eventually becoming ensnared in the shear flow, feeding a breakdown to turbulence.

The astrophysical significance of this is associated with accretion sources. By dint of angular momentum conservation, accretion onto compact objects invariably involves differentially rotating flow with a centrifugal barrier enshrouding the central mass. The presence of a magnetic field leads to the development of the MRI, which causes fluid elements to lose their specific angular momentum, and leads to the accretion process itself. Happily, for the purposes of this conference, many of the details of this process may be studied via large-scale numerical simulations.

In this paper we will briefly review the history and contributions of numerical MHD simulations of accretion disks. Numerical investigations of this problem began in earnest only a decade ago, and their impact has been profound. They have taught us not only *what* happens in a highly complex accretion flow; by varying flow parameters we have often learned why it happens as well. As a recent example of this, we shall discuss in some detail the structure of nonradiating flows

E. Falgarone and T. Passot (Eds.): LNP 614, pp. 329–348, 2003.
© Springer-Verlag Berlin Heidelberg 2003

(associated with black hole accretion) that numerical simulations have revealed in the last year or so. The topic is important, fascinating, and not entirely free of controversy.

The review proceeds along the following outline. In Sect. 2, we review the physics of the magnetorotational instability in its most general form. Section 3 discusses simulations that focus on a local patch of an unstable Keplerian disk. By limiting dynamical range, these calculations may include more sophisticated physics. In Sect. 4, global MHD simulations are summarized. Given their much larger dynamical range requirements, it is not yet possible to treat the more complex fluids amenable to a local analysis, but even these simple fluids may be of astrophysical relevance. There is ample evidence now for a non-radiative flow at the Galactic Center, and there is reason to believe this is not a special case. We conclude with a summary in Sect. 5.

2 The Magnetorotational Instability

We begin with a review of the MRI (Balbus & Hawley [10]). The dynamics of this instability is very simple, involving only the notion of magnetic tension in the presence of rotational forces.

2.1 Formal Calculation

The simplest case involves an axisymmetric disk threaded by a weak, vertical magnetic field, $\boldsymbol{B} = B\boldsymbol{e_Z}$, where $\boldsymbol{e_Z}$ is a vector in the Z direction. (We are using a standard (R, ϕ, Z) cylindrical coordinate system.) The undisturbed flow consists of fluid elements on circular orbits. We make a displacement $\boldsymbol{\xi} = (\xi_R, \xi_\phi, 0)$ in the disk plane. The time and space dependence of the displacement is a simple plane wave,

$$\boldsymbol{\xi} = \boldsymbol{\xi'} \exp(ikZ - i\omega t), \tag{1}$$

which defines the wavenumber k and angular frequency ω. The wave propagates along the magnetic field line at the Alfvén speed

$$v_A^2 = \frac{B^2}{4\pi\rho},$$

and is a consequence of the restoring field line tension $-(kv_A)^2\boldsymbol{\xi}$.

We go into a frame rotating at the angular velocity Ω of a fiducial orbit. In this frame, in addition to the magnetic tension force, we must add a Coriolis force $-2\boldsymbol{\Omega}\times d\boldsymbol{\xi}/dt$, and a centrifugal force $R\Omega^2\boldsymbol{e_R}$. The latter is exactly balanced by the inward gravitational force just at the location of the fiducial orbit, and the residual tidal force amounts to $-\boldsymbol{\xi}d\Omega^2/d\ln R$. The equations of motion for a fluid displacement are thus

$$\frac{d^2\xi_R}{dt} - 2\Omega\frac{d\xi_\phi}{dt} = -\left[(kv_A)^2 + \frac{d\Omega^2}{d\ln R}\right]\xi_R, \tag{2}$$

$$\frac{d^2\xi_\phi}{dt} + 2\Omega\frac{d\xi_R}{dt} = -(kv_A)^2\xi_\phi. \tag{3}$$

The coefficients Ω and $d\Omega^2/d\ln R$ are now taken to be constant, since our calculation is local. Note as well that pressure forces are unimportant since $\boldsymbol{k}\cdot\boldsymbol{\xi} = 0$.

Plane wave solutions of these equations satisfy the dispersion relation

$$\omega^4 - \omega^2[\kappa^2 + 2(kv_A)^2] + (kv_A)^2\left[(kv_A)^2 + \frac{d\Omega^2}{d\ln R}\right] = 0, \tag{4}$$

where κ is known as the epicyclic frequency

$$\kappa^2 = 4\Omega^2 + \frac{d\Omega^2}{d\ln R} = \frac{1}{R^3}\frac{d(R^4\Omega^2)}{dR}, \tag{5}$$

i.e., κ^2 is just proportional to the angular momentum gradient. In the absence of a magnetic field, fluid displacements would oscillate about their unperturbed circular orbit radius at a frequency κ. The displacements can easily be calculated, and they appear in the rotating frame as elliptical paths, with the elements moving in a retrograde sense relative to the unperturbed circular orbits. It is these "epicycles" that give κ its name.

The relation (4) is a quadratic equation in ω^2, and it is straightforward matter to show that if $d\Omega^2/dR < 0$, there are $\omega^2 < 0$ unstable modes. By way of contrast, it is *angular momentum*, not the angular velocity, that must decrease outwards for instability in an unmagnetized disk. This has major astrophysical consequences, because disks in nature almost all have a specific angular momentum profile that rises with increasing radius, but an angular velocity profile that decreases in the same direction. Ignoring even a highly subthermal field in a stability analysis is a grave error, leading to qualitatively incorrect conclusions. Magnetized and unmagnetized fluids behave very, very differently.

The maximum growth rate of the instability is a quantity of interest,

$$|\omega_{max}| = \frac{1}{2}\left|\frac{d\Omega^2}{d\ln R}\right|, \tag{6}$$

which students of galactic structure will recognize as the Oort A value. In a Keplerian disk it amounts to 0.75Ω, an enormous growth rate. It has been speculated that it is impossible for any instability feeding off differential rotation to grow more rapidly (Balbus & Hawley [9]).

2.2 Qualitative Description

The linearized equations of motion have precisely the same form as those that emerge in a simple mechanical problem. Imagine two masses connected by a spring, in orbit about a central mass. If one replaces $(kv_A)^2$ by a spring constant, say K, our systems are identical. This gives an easy way to envision the behavior of the instability, with magnetic tension acting like a spring obeying Hooke's law.

Denote the first mass as M_1, and assume it is on an orbit slightly closer to the center than mass M_2, which orbits farther out. M_1 travels slightly faster than M_2 if the angular velocity increases inwards, and the spring tension pulls back on it. On the other hand, M_2 is pulled forward in its orbital trek. This means that there is a positive torque pulling forward on M_2, and a retarding torque pulling back on M_1. Thus M_1 loses angular momentum, M_2 gains angular momentum. With additional angular momentum, M_2 (the outer mass) moves to a more distant orbit, while M_1 (the inner mass), moves to an orbit farther in. This stretches the spring yet further, the tension rises, and the process runs away. This is the underlying cause of the magnetorotational instability. Magnetized angular momentum transport is an intrinsically unstable process, because the more fluid elements separate, the greater the rate of angular momentum exchange that is driving the separation in the first place.

2.3 General Stability Criteria

It is possible to study the behavior of axisymmetric modes in great generality. By way of comparison, we first give the results for an unmagnetized gas. The local stability of adiabatic perturbations is then governed by the Høiland criteria (e.g. Tassoul [54]):

$$-\frac{3}{5\rho}(\boldsymbol{\nabla} P)\cdot\boldsymbol{\nabla}\ln P\rho^{-5/3}+\frac{1}{R^3}\frac{\mathrm{d}R^4\Omega^2}{\mathrm{d}R}\geq 0, \tag{7}$$

$$\left(-\frac{\mathrm{d}P}{\mathrm{d}Z}\right)\left(\frac{\mathrm{d}R^4\Omega^2}{\mathrm{d}R}\frac{\mathrm{d}\ln P\rho^{-5/3}}{\mathrm{d}Z}-\frac{\mathrm{d}R^4\Omega^2}{\mathrm{d}Z}\frac{\mathrm{d}\ln P\rho^{-5/3}}{\mathrm{d}R}\right)\geq 0, \tag{8}$$

which allow for the presence of both vertical and radial gradients in the host medium. (Here, P is pressure and ρ is mass density.)

In the presence of a magnetic field, all angular momentum gradients are replaced by angular velocity gradients in the stability criteria (Balbus [6]):

$$-\frac{3}{5\rho}(\boldsymbol{\nabla} P)\cdot\boldsymbol{\nabla}\ln P\rho^{-5/3}+\frac{\mathrm{d}\Omega^2}{\mathrm{d}\ln R}\geq 0, \tag{9}$$

$$\left(-\frac{\mathrm{d}P}{\mathrm{d}Z}\right)\left(\frac{\mathrm{d}\Omega^2}{\mathrm{d}R}\frac{\mathrm{d}\ln P\rho^{-5/3}}{\mathrm{d}Z}-\frac{\mathrm{d}\Omega^2}{\mathrm{d}Z}\frac{\mathrm{d}\ln P\rho^{-5/3}}{\mathrm{d}R}\right)\geq 0. \tag{10}$$

This generalizes the result found in our simple example from the previous section, that the angular velocity, not the angular momentum, must increase outward when $\Omega=\Omega(R)$. We find here that angular velocity gradients are quite generally the proper stability discriminants in a magnetized gas.

Finally, we include the effect of a Coulomb conductivity. In the presence of a magnetic field, dilute astrophysical plasmas conduct heat only along magnetic field lines, and then of course only if there is a gradient along the field line. This is a regime of some relevance to black hole accretion sources, in which the flow is characterized by very high temperatures and low densities. In this case, we find



Fig. 1. Development of thermoclinic instability in a Schwarzschild-stable layer. Magnetic lines of force are shown after one Alfvén crossing time, initial seeding with rms 1% random initial vertical velocity perturbations. Initial thermal energy density is 1600 times magnetic; initial field lines are isothermal and horizontal; vertical axis is stratified (Z) direction, horizontal is orthogonal (X) direction. Vertical grid runs from $Z = 1$ to 2, initial temperature profile is $1/Z$, gravitational field is $1/Z^2$; grid is 128×64. (From Balbus [11], simulation performed by J. M. Stone.)

that the stability criteria may be obtained from the previous simply by changing the entropy gradients to temperature gradients (Balbus [7]):

$$-\frac{1}{\rho}(\boldsymbol{\nabla} P) \cdot \boldsymbol{\nabla} \ln T + \frac{\mathrm{d}\Omega^2}{\mathrm{d} \ln R} \geq 0, \tag{11}$$

$$\left(-\frac{\mathrm{d}P}{\mathrm{d}Z}\right) \left(\frac{\mathrm{d}\Omega^2}{\mathrm{d}R}\frac{\mathrm{d} \ln T}{\mathrm{d}Z} - \frac{\mathrm{d}\Omega^2}{\mathrm{d}Z}\frac{\mathrm{d} \ln T}{\mathrm{d}R}\right) \geq 0. \tag{12}$$

Note that these requirements are independent of the thermal conductivity coefficient. This is a most surprising result: a nonrotating, adiabatic stratification of a temperature plus a small sprinkling of magnetic field is highly unstable! Figure 1 shows the results of a full nonlinear simulation of this instability, as far as it could be followed. The restriction that the heat flows only along field lines is critical here, as a scalar thermal conductivity (e.g. radiation) would serve only to stabilize perturbations by dissipation. Further discussion of this delicate point may be found in Balbus [7].

The three forms of the stability criteria show an "evolution" of replacing gradients of the extensive variables entropy S and angular momentum L with gradients of the intensive quantities T and Ω. Changes in the former are energy sources,

$$dE = T\, dS + \Omega\, dL + ..., \tag{13}$$

while changes in the latter are *free* energy sources,

$$d(E - TS - \Omega L) = -S\, dT - L\, d\Omega + ..., \tag{14}$$

Gradients in the free energy generally mean that lower energy or higher entropy equilibrium states are nearby. When a dynamical path becomes available to these states, instabilities are triggered. Transitions are swift: in all cases, the characteristic growth times are dynamical, either a sound crossing or a shearing time.

3 Local Nonlinear Simulations

With the general stability conditions in hand, it is natural to ask what are the nonlinear consequences of their violation? In general the answer is turbulent flow, and further progress depends upon numerical simulation.

3.1 Governing Equations

Start with the fundamental R and ϕ equations of motion:

$$\rho\frac{\mathrm{d}v_R}{\mathrm{d}t}+\rho(\boldsymbol{v}\cdot\boldsymbol{\nabla})v_R-\rho\frac{v_\phi^2}{R}=-\rho\frac{\mathrm{d}\Phi}{\mathrm{d}R}-\frac{\mathrm{d}P_{tot}}{\mathrm{d}R}+(\boldsymbol{B}\cdot\boldsymbol{\nabla})\frac{B_R}{4\pi}-\frac{B_\phi^2}{4\pi R}, \tag{15}$$

$$\rho\frac{\mathrm{d}v_\phi}{\mathrm{d}t}+\rho(\boldsymbol{v}\cdot\boldsymbol{\nabla})v_\phi+\rho\frac{v_\phi v_R}{R}=-\frac{1}{R}\frac{\mathrm{d}P_{tot}}{\mathrm{d}\phi}+(\boldsymbol{B}\cdot\boldsymbol{\nabla})\frac{B_\phi}{4\pi}+\frac{B_\phi B_R}{4\pi R}, \tag{16}$$

where

$$P_{tot}\equiv P+\frac{B^2}{8\pi}$$

is the total gas plus magnetic pressure. To leading order, the disk is a simple Keplerian system, with angular velocity

$$R\Omega^2(R)=\frac{GM}{R^2}\equiv\frac{\mathrm{d}\Phi}{\mathrm{d}R} \tag{17}$$

where M is the central mass. Thermal and magnetic forces are small compared with the central gravitational force, though of course this "smallness scale" is precisely the one we are concerned with! We may express this quantitatively as

$$|(v_\phi-R\Omega)|\sim c_S\ \text{(sound speed)}\sim v_A\ll R\Omega. \tag{18}$$

Note that the azimuthally-averaged ϕ equation may be written

$$\rho\frac{\mathrm{d}(R\rho v_\phi)}{\mathrm{d}t}+\boldsymbol{\nabla}\cdot\left[\rho R\left(v_\phi\boldsymbol{v}-v_{A\phi}\boldsymbol{v_{Ap}}\right)\right]=0 \tag{19}$$

where $\boldsymbol{v_{Ap}}$ is the poloidal Alfvén velocity, and all the angular momentum density and flux are understood to be averaged quantities. The ϕ-averaged radial angular momentum flux may be read off directly from (19):

$$\mathcal{F}_{\mathcal{J}\,R}=R\langle\rho(v_\phi v_R-v_{A\phi}v_{AR})\rangle, \tag{20}$$

which is an important quantity.

In the *local approximation*, the idea is to focus on a small patch of the disk, fixing the new origin to corotate with disk fluid orbiting at a particular angular velocity, Ω_0. We measure all velocities relative to $R\Omega_0$. We ignore curvature effects in the local computational patch of interest. Formally, we work in the limit,

$$R \to \infty, \quad v_\phi \to \infty, \quad \Omega \to \text{finite}. \tag{21}$$

We define the velocity $\boldsymbol{w}$:

$$\boldsymbol{w} = \boldsymbol{v} - R\Omega_0\, \boldsymbol{e_\phi}. \tag{22}$$

The value of R at which $\Omega = \Omega_0$ will be denoted R_0. In general we shall consider only small radial excursions from R_0,

$$R = R_0 + X, \quad X \ll R_0. \tag{23}$$

Thus,

$$R\Omega_0^2 - \frac{\mathrm{d}\Phi}{\mathrm{d}R} = R(\Omega_0^2 - \Omega^2(R)) \simeq -\, X \frac{d\Omega^2}{d\ln R}\bigg|_{R=R_0} \tag{24}$$

to leading order. Local Cartesian coordinates can be defined by aligning the X and Y axes along R and ϕ. In a frame rotating at Ω (dropping the subscript), the local equations of motion for w_X and w_Y are

$$\frac{\mathrm{d}w_X}{\mathrm{d}t} + \boldsymbol{w}\cdot\boldsymbol{\nabla} w_X - 2\Omega w_Y + X\frac{d\Omega^2}{d\ln R} = -\frac{1}{\rho}\frac{\mathrm{d}P_{tot}}{\mathrm{d}X} + \boldsymbol{B}\cdot\boldsymbol{\nabla} B_X \tag{25}$$

$$\frac{\mathrm{d}w_Y}{\mathrm{d}t} + \boldsymbol{w}\cdot\boldsymbol{\nabla} w_Y + 2\Omega w_X = -\frac{1}{\rho}\frac{\mathrm{d}P_{tot}}{\mathrm{d}Y} + \boldsymbol{B}\cdot\boldsymbol{\nabla} B_Y \tag{26}$$

The forms of the remaining dynamical equations remain unaffected by the change to rotating coordinates. They are the Z equation of motion

$$\frac{\mathrm{d}w_Z}{\mathrm{d}t} + \boldsymbol{w}\cdot\boldsymbol{\nabla} w_Z = -\frac{1}{\rho}\frac{\mathrm{d}P_{tot}}{\mathrm{d}Z} + \boldsymbol{B}\cdot\boldsymbol{\nabla} B_Z, \tag{27}$$

the equation of mass conservation

$$\frac{\mathrm{d}\rho}{\mathrm{d}t} + \boldsymbol{\nabla}\cdot(\rho\boldsymbol{w}) = 0, \tag{28}$$

the internal energy equation,

$$(\gamma - 1)\rho\left(\frac{\mathrm{d}}{\mathrm{d}t} + \boldsymbol{w}\cdot\boldsymbol{\nabla}\right)\frac{P}{\rho} = -P\boldsymbol{\nabla}\cdot\boldsymbol{w} \tag{29}$$

The induction equation for the magnetic field is simply

$$\frac{\mathrm{d}\boldsymbol{B}}{\mathrm{d}t} = \boldsymbol{\nabla}\times(\boldsymbol{w}\times\boldsymbol{B}). \tag{30}$$

The set of equations (25)–(30) completely describes the local behavior of a magnetized accretion disk. We must first, however, specify the boundary conditions (BC) at the edges of the computational domain. The simplest approach is to use hard walls at each of the radial, azimuthal, and vertical boundaries. The first nonlinear numerical investigations of local accretion disk behavior were those of Hawley & Balbus [22], who used hard wall BC. These first simulations were highly restrictive, but sufficed to demonstrate the existence of the weak field MRI, and confirm its analytic growth rates and its independence of an azimuthal field. These simulations used a vertical field and more complex loop geometries as starting configurations. Because no periodic BC were needed, these early simulations actually retained the curvature terms in the equations, so that non WKB terms were present. The detailed agreement between analytic calculation and numerical results left no doubt whatsoever of what was then a completely unexpected result: the combination of Keplerian rotation and a weak magnetic field is extremely unstable.

Extended simulations require what are known as shearing-box BC. In this approach, the azimuthal boundary conditions are always periodic. Vertical BC may be taken either as periodic or pure outflowing, depending upon the problem on interest. Because of the presence of large scale shear, the radial BC are more complex (Hawley, Gammie, & Balbus [26]). They may be described as "quasi-periodic:" the computational domain is thought of as one brick in a wall extending to infinity, each brick with the same fluid configuration as the next. The velocity shear is continuous across the entire brick wall, so a layer of bricks must slide with respect to the layer above and below! As one layer slides over another, a fluid element leaving the computational domain is replaced by its image on the opposite wall, but not at the azimuth it has just vacated. Instead it reappears at the azimuth from the sliding brick in contact with the computational domain. The mathematical formulation of these BC may be found in Balbus & Hawley [10].

3.2 Local Axisymmetric Flow

To run simulations over many local shear times, the shearing box formalism must be implemented. This was first done by Hawley & Balbus [23] for two-dimensional axisymmetric flow. A surprise emerged. Though the evolution of an initial uniform vertical magnetic field was in complete accord with analytic theory in the linear stages of its development, the nonlinear stages hardly appeared turbulent at all. In fact, the linear stage seemed to continue indefinitely, with exponentially growing streaming motions persisting. This contrasted sharply with the nonlinear behavior of a shearing box starting with a vertical field whose mean value was zero—half upwards and half downwards say, or sinusoidally varying. In that case, turbulence quickly developed after a few linear growth times, and then gradually decayed, leaving no field at all at the end of the simulation!

This behavior can be understood with the help of Cowling's anti-dynamo theorem (Moffatt [36]). The theorem states that dynamo amplification is impossible in an isolated, dissipative axisymmetric system. To understand what is

meant by "isolated," we work with the azimuthal component of the vector potential, denoted A. In axisymmetry, this component alone suffices to determine the poloidal magnetic field. The mathematical heart of the theorem is that in the absence of resistance, the integrated form of Faraday's induction equation may be manipulated into the form

$$\frac{\mathrm{d}}{\mathrm{d}t}\int A^2\,dV = -\int A^2\,\boldsymbol{v}\cdot\boldsymbol{dS}, \tag{31}$$

where the left integral is over a volume containing the fluid (perhaps infinite), and the right integral is over a bounding surface (perhaps at infinity). If periodic boundary conditions are used, or if A^2 falls off sufficiently rapidly, then the surface integral vanishes. Hence, the average of A^2 remains constant for the fluid. The presence of any dissipation then causes an inevitable decline; there is nothing to offset it.

When the mean vertical field vanishes, the vector potential satisfies smooth periodic boundary conditions, and the anti-dynamo theorem applies directly. The magnetic field depends upon derivatives of A, and its growth therefore requires a sort of continuous kneading of the fluid, bringing different values of A (which, absent dissipation, is a fluid element label) ever closer together. This continues on smaller and smaller scales, but eventually the grid scale is hit, and growth ceases at that point, with reconnection ensuing.

The presence of a nonvanishing mean magnetic field implies that A is nonperiodic: it must have a component proportional to X. Since periodic boundary conditions no longer hold, the surface integral in (31) no longer vanishes, and poloidal field components may grow at the expense of the free energy of differential rotation. There is nothing unphysical about this set-up; real disks certainly can be threaded by a magnetic field. The question is do such disks really exhibit the streaming behavior discussed above?

This was examined authoritatively by Goodman & Xu [19]. These authors pointed out a remarkable fact: starting with a vertical field, linear plane wave eigensolutions are in fact exact *nonlinear* solutions to the equations of motion. In the local approximation, the gas really does appear to act like two orbiting masses connected by a spring. The existence and persistence in two dimensions of coherent disk streams provided an explanation for a recurring puzzling behavior seen in a number of earlier axisymmetric *global* MHD disk simulations. Uchida & Shibata [56] and Shibata & Uchida [47] were interested in the creation of MHD jets and investigated this problem by threading a disk of gas with a vertical magnetic field. By imparting less than the Keplerian value of the angular momentum to the orbiting fluid elements, they hoped to duplicate the effects of slow radial accretion—but without turbulence. The infall produced radial fields that became wrapped up by differential rotation into strong toroidal fields, whose gradients in turn drove dynamical outflows along the vertical field lines. Some of these simulations, however, began with a Keplerian disk embedded in a vertical magnetic field. Such disks also collapsed on a dynamic time scale. At the time, the reason for this was not at all clear. This may now be understood as a global manifestation of the streaming solutions studied by Goodman and Xu [19].

The question was whether in three dimensions the fluid behavior will prove to be qualitatively different from the two dimensional case. Do the streams remain stable in three dimensions? Goodman & Xu noted that this ostensibly nonlinear question reduces to a *linear* stability problem, but a linear stability problem perturbed about a most unusual equilibrium solution. The presence of vertically periodic velocity streams renders the problem amenable to Floquet analysis (Bender & Orszag [12]), for which powerful mathematical techniques are available. The conclusion was that the new equilibrium of streaming motions should be unstable. The most important instability is a magnetized Kelvin-Helmholtz mode, which appears for radial wavelengths in excess of the streaming equilibrium flow's vertical wavelength.

3.3 Local Three-Dimensional Simulations

To determine the ultimate fate of the streams of the Goodman-Xu solution or to study dynamo amplification requires implementation of three-dimensional MHD codes. By now, many shearing box studies have been carried out, and a wide variety of models explored. The simplest consists of a homogeneous box, in which only the radial component of the large-scale gravitational field is retained, and the magnetic field is initially uniform (Hawley et al. [26]; Matsumoto & Tajima [33]). The initial field geometry in these studies had some combination of vertical and toroidal components. A more complicated initial field configuration, important for understanding dynamo activity, is to let the initial field have a random character with vanishing mean (Hawley, Gammie, & Balbus [27]). The presence of the vertical component of the gravitational field produces a density stratification (Brandenburg et al. [16]; Stone et al. [51]; Matsuzaki et al. [34]), which introduces the possibility of magnetic buoyancy, a new effect. Furthermore, by bringing the pseudoscalar quantity $\Omega \cdot \boldsymbol{\nabla}\rho$ into the problem, stratification breaks chiral symmetry, i.e., the flow acquires a "handedness." This result is potentially important for the development of local mean helicity in the turbulence, a feature upon which much of classical kinematic dynamo theory is based (Moffatt [36]).

Stone et al. [51] carried out a series of such simulations spanning two vertical scale heights in an initially isothermal disk. With the onset of the MRI, magnetic field rises out of the disc to establish a highly magnetized corona. The amplitude of the turbulence near the midplane is determined more by local dissipation than by buoyant losses, and the resulting stress levels are not greatly modified from the nonstratified simulations. The presence of a corona, however, could have important observational consequences if a significant amount of dissipational heating occurs there. Concerns about the effects of the close-in vertical boundary conditions led Miller & Stone [37] to carry out simulations with a larger computational domain covering 5 scale heights above and below the equator. They demonstrated that indeed a strongly magnetized corona ($P_{\rm gas} \ll P_{\rm mag}$) forms naturally from little more than a disk and a weak seed field.

Three-dimensional studies also show the breakdown of the two-dimensional streaming solutions (Hawley et al. [26]). If the computational box is large enough

to allow an unstable radial wavelength, streaming is disrupted within a few orbits, as the Kelvin-Helmholtz instability noted by Goodman & Xu [19] leads to fluid turbulence, which is the basis of significant outward angular momentum transport. This is because the turbulence is inherently anisotropic; perturbations in the R and ϕ components of the magnetic field and the velocity are highly correlated. Velocity power spectra of the simulations are weighted toward the lowest wavenumbers, i.e., there are significant fluctuations on scales comparable to the computational domain size. These large-scale fluctuations contain most of the magnetic energy and contribute the lion's share of the stress.

Angular momentum is transported outward by the the MRI. To understand what is meant by that let

$$\boldsymbol{v} = R\Omega(R)\boldsymbol{e_\phi} + \boldsymbol{u} \tag{32}$$

i.e., $\boldsymbol{u}$ is the velocity in excess of the local Keplerian rotation. (Note that $\boldsymbol{w}$ is by contrast the velocity in excess of the local solid body rotation $R\Omega_0\boldsymbol{e_\phi}$.) Then the radial angular momentum flux from (20) is

$$\mathcal{F}_{\mathcal{J}\,R} = R^2\Omega\langle\rho u_R\rangle + R\langle\rho(u_R u_\phi - v_{AR}v_{A\phi})\rangle. \tag{33}$$

The first term is simply the Keplerian angular momentum carried directly by the mass accretion. The second term is transport that is present whether or not there is any accretion present. This need not imply that an individual fluid element is losing angular momentum (though it generally is). In its linear stage, an axisymmetric *unmagnetized* convective instability transports angular momentum by having low and high angular momentum fluid elements interpenetrate, with no net mass flux. This means an *inward* angular momentum flux in a Keplerian disk, a result which is preserved in nonlinear three-dimensional simulations (Stone & Balbus [50]).

3.4 Radiative Effects

In general, thermal diffusion effects are not included in simulations. If the object of study is a classical optically thick Keplerian disk, this is a sensible approach to the dynamics. Radiative diffusion regulates the vertical temperature profile of the disk, but it does not greatly influence dynamical stability.

The linear stability of a magnetized, stratified, radiative gas was recently addressed by Blaes & Socrates [15]. Despite the complexity of the full problem, the MRI emerges at the end of the day unscathed, its classical stability criterion $d\Omega^2/dR > 0$ remaining intact. The maximum growth rate can be lowered however, particularly when the azimuthal field approaches or exceeds thermal strength. (The same is true for the ordinary MRI, as shown by Blaes & Balbus [14]).

The nonlinear problem has been carried through by Turner, Stone, & Sano [55], who set up a local, radiative, axisymmetric, shearing box flow. The linear calculations of Blaes & Socrates [15] were confirmed in detail, and the nonlinear flow was analyzed. As in standard MRI simulations, the stress is dominated by the Maxwell component, which is a factor of a few larger than the Reynolds

terms. Photon diffusion plays a dual role in keeping the matter nearly isothermal, and in creating over-dense clumps of thermally-supported gas. The clumping occurs when radiation pressure support is lost via diffusion on a dynamical time scale at large wave numbers. In this regime, radiative disks may be highly inhomogeneous. At larger optical depths and longer diffusion times, the disks look similar to their nonradiative counterparts.

3.5 Low Ionization Disks

Protostellar disks, and possibly cataclysmic variable disks, contain regions of low ionization fraction—so low that the assumptions of ideal MHD break down. Ohmic dissipation becomes important, together with the Hall inductive terms. (The latter arises because the distinction between the mean electron and mean fluid velocities becomes important.) The effect of Hall electromotive forces on the MRI have been studied by Wardle [57] and by Balbus & Terquem [11]. Nonlinear numerical simulations including ohmic resistivity have been done by Fleming, Stone, & Hawley [17]; Sano & Stone [45] have done simulations including both the ohmic and Hall processes. It can be shown that under rather general conditions, if Ohmic dissipation is important, Hall electromotive forces are also important (Balbus & Terquem [11]; Sano & Stone [45]).

The most interesting feature introduced by Hall electromotive forces is helicity: the relative orientation of the angular velocity $\boldsymbol{\Omega}$ and magnetic field $\boldsymbol{B}$ vectors matters (Wardle [57]; Balbus & Terquem [11]). The key point is that $\boldsymbol{\Omega} \cdot \boldsymbol{B} > 0$ configurations raise the maximum growth rate in the presence of ohmic losses, and this aligned configuration results in more vigorous transport in the local axisymmetric simulations of Sano & Stone [45]. In configurations where $\boldsymbol{\Omega} \cdot \boldsymbol{B}$ vanishes on average, the level of turbulence is much more sensitive to the size of the ohmic dissipation term.

In the simulations of Fleming et al. [17], a critical magnetic Reynolds number Re_M emerged below which turbulence is suppressed (Re_M is defined here as the ratio of the product of the scale height and sound speed to the resistivity). When the mean field vanishes, Re_M was found to be surprisingly high, $\sim 10^4$. In the presence of a mean field, the critical $Re_M \sim 10^2$. The interesting and important question is whether the inclusion of Hall electromotive forces changes these numbers by lowering them, i.e., making it easier to support turbulence. The Sano & Stone [45] axisymmetric simulations did not reveal a large change, but questions on the maintenance of turbulence more properly await a three-dimensional treatment.

4 Global Disk Simulations

Full global simulations are a demanding computational problem, requiring extended evolutions at high resolution. The fundamental difficulties with global simulations are worth reiterating. First one has a severe problem with length scales. The goal is to evolve accretion disks from first principles, which means

computing the angular momentum transport self-consistently. Angular momentum transport is due to MHD turbulence, and the most-unstable MRI modes will typically be much smaller than the disk pressure scale height H. One would also like to resolve the turbulent cascade through its inertial range, if possible. The disk itself extends from the black hole horizon, r_S, out to thousands of horizon radii. Since all time scales are more or less $\propto \Omega$, Kepler's law imposes a difficulty as well.

One must be practical and work within well-chosen limitations. It is often helpful to forgo density stratification, for example. One may also restrict the dimensionality of the problem, by adopting axisymmetry. Finally, and most critically, one must make a judicious choice of problem before embarking on a full global simulation.

4.1 Two-Dimensional Simulations

Axisymmetric MHD disk simulations have been used extensively in the past, particularly to study jet formation processes. Such simulations date back at least to Uchida & Shibata [56]. It is only more recently that two-dimensional simulations have directed primarily at the internal dynamics of the disk itself and the resulting accretion flow, rather than to the launching and collimation of jets.

From the astrophysical side, there has been considerable recent interest in under-luminous accreting compact objects, inspired in no small part by X-ray observations of the Galactic center (Melia & Falcke [35]). Despite the inferred presence of a $2\times10^6\,M_\odot$ mass black hole, little of the expected radiation has been detected by Chandra. If the gas is able to radiate more rapidly than it accretes, there seems little doubt that it forms a thin Keplerian disk along the lines of Shakura & Sunyaev [46]. If the flow is either too optically thick for the radiation to diffuse outward over an accretion time, or too optically thin to radiate, then the fate of the gas is much less clear. Several ideas have circulated in the literature, dating from the earliest days of accretion theory. A small sample: Pringle & Rees [43], Ichimaru [30], Begelman & Meier [13], Rees et al. [44], Abramowicz et al. [1], Narayan & Yi [39, 40], Narayan, Igumenschev, & Abramowicz [38]; Abramowicz et al. [2].

From the numericist's perspective, it is a most welcome development that there is evidence for, and intense interest in, nonradiating flows. For this is precisely the type of flow that is well-suited to numerical simulation. The first two-dimensional MHD simulations of nonradiating accretion flows were carried out by Stone & Pringle [52]. Here, accretion begins from an initial equilibrium torus, and is driven by MHD stresses arising from the onset of the MRI. The initial infall phase is relatively smooth and dominated by the vertical field "channel flow" of the MRI (cf. Sect. 3). Subsequent evolution is decidedly turbulent. The resulting flow consists of an approximately barotropic disk near the midplane, with constant Ω contours parallel to cylindrical radii. There is no strong correlation between the specific angular momentum and entropy, as was found in hydrodynamical simulations (Stone, Pringle, & Begelman [53]). A substantial

magnetized coronal outflow enveloped the disk. Finally, there was very little difference in the character of the solution if dissipative losses were retained as heat or simply ignored.

While most of these features were destined to survive in three-dimensional runs, two dimensions is generally a significant limitation for the study of MHD turbulence (Hawley [20]). Poloidal magnetic field cannot be indefinitely maintained in axisymmetry, by the anti-dynamo theorem (e.g., Moffat [36]). Indeed, toward the end of the Stone & Pringle [52] simulations the turbulence clearly begins to die away. Axisymmetric simulations also over-emphasize streaming modes, and have a tendancy to produce coherent radial magnetized flows rather than the more generic MHD turbulence. Finally, toroidal field instabilities cannot be simulated in axisymmetry. Nothwithstanding these limitations, by allowing a rapid investigative turnover of plausible accretion scenarios, two-dimensional simulations have proven to be a valuable tool, and a remarkably reliable one as well.

4.2 "Cylindrical Disks"

The "cylindrical disk" is a global three-dimensional system allowing for full radial and azimuthal dynamics, but ignoring the vertical component of the central gravitational field. Hence, there is no vertical stratification. Turbulence and magnetic field can be easily sustained however, and far fewer vertical grid zones are required compared with true three-dimensional simulations.

Armitage [4] carried out the first such calculation using the standard ZEUS code and a grid covering the full 2π in azimuth, running from $R = 1$ to 4 with a reflecting inner boundary and an outflow outer boundary, and covering a length of 0.8 in Z. Vertical boundary conditions were periodic. The initial magnetic field was vertical and proportional to $\sin(kR)/R$, with the sine function argument linearly varying over 2π between $R = 1.5$ and 3.5. From this initial state, turbulence rapidly developed along with significant angular momentum transport.

A more extensive set of cylindrical simulations (Hawley [21]) reaffirmed many of the conclusions of the local box models for accretion disk turbulence driven by the magnetorotational instability, as well as testing several technical aspects of global simulations. Because the driving instability is local, a reduction in the azimuthal computational domain to some fraction of 2π does not create large qualitative differences. Similarly, the choice of either an isothermal or adiabatic equation of state has little impact on the initial development of the turbulence. Simulations that begin with vertical fields have greater field amplification and higher ratios of stress to magnetic pressure compared with those beginning with toroidal fields.

Recently, cylindrical disk calculations have been applied to the problem of the star-disk boundary layer (Armitage [5]; Steinacker & Papaloizou [49]). Such simulations represent an important step forward, as previous studies of boundary layers were modeled by a hydrodynamic viscosity. Significant dynamo activity was reported by both sets of investigators. In particular, Armitage [5] finds an

order of magnitude larger field energy density in the boundary layer compared with the average disk field. Dissipative heating in the boundary layer has yet to be simulated.

4.3 Three-Dimensional Simulations

Global three-dimensional MHD disk simulations with full vertical structure were presented by Hawley [20]. These models started with equilibrium tori containing either a weak poloidal or a weak toroidal magnetic field. A torus is a useful initial condition for global simulations because it can be well-resolved and wholly contained within the grid. The thickness of the torus depends upon the angular momentum distribution, with constant specific angular momentum l associated with the thickest tori. Such structures may actually be generic in active galactic nuclei, where they feed an inner disk (Krolik [31]). As the system evolution proceeds, the MRI develops rapidly. Stresses are dominated by the Maxwell component (a factor of several larger than the Reynolds stress), which immediately redistributes angular momentum from the initial non-Keplerian profile to a nearly Keplerian state. At later times, the disks show rapid time variability, tightly wrapped, low-m spiral structure, and significant internal stress.

Machida, Hayashi, & Matsumoto [32] studied the evolution of a constant angular momentum torus containing toroidal magnetic field. Turbulence soon develops and magnetically dominated ($\beta < 1$) filaments form. The buoyancy of such fields leads to the development of a strongly magnetized corona.

Steinacker & Henning [48] have recently revisited the question of the influence of a large-scale vertical field on an accretion disk, previously studied in axisymmetry. They found strong accretion in the disk (the authors described it as a collapse) driven by the MRI. Coupling to a magnetized corona drives outflows, with the degree of collimation depending on the strength of the field.

The nature of the magnetic stress near the marginally stable orbit of black hole ($6GM/c^2$ for a Schwarzschild hole) has received renewed attention. These disk hydrodynamic treatments using a turbulent viscosity lead to a vanishing stress in this region (Abramowicz & Kato 1989), but matters can be more complicated when magnetic stresses are involved (e.g. Novikov & Thorne [41]; Gammie [18]; Algol & Krolik [3]). The question of what happens to the stress is of interest because of the possibility of extracting energy from the hole's rotation and communicating it to the disk. It can be addressed via three-dimensional MHD simulations.

Hawley & Krolik [28, 29] investigated the evolution of a magnetized accretion torus lying near the marginally stable orbit in a pseudo-Newtonian model potential for the black hole (Paczyński & Wiita [42]). The work focused on the behavior of the stress as the gas flows through the "plunging region" inside of the marginally stable orbit. These simulations used the highest three-dimensional resolution to date, with up to $256 \times 64 \times 192$ zones in (R, ϕ, Z), and as many as 70 radial grid zones lying between the marginally stable orbit and the horizon. They find that in contrast to standard models, the disk does not sharply truncate at the marginally stable orbit, nor does the stress vanish. For both poloidal

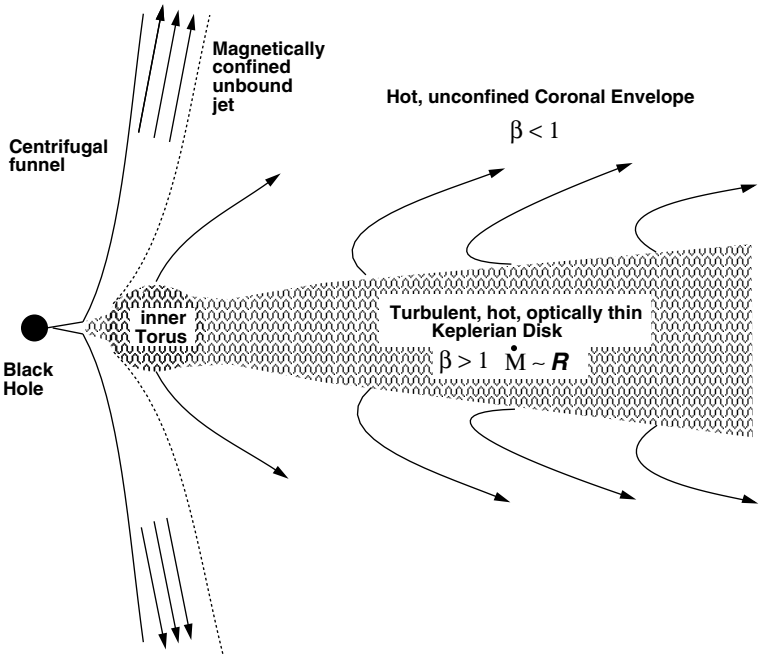


Fig. 2. A schematic diagram of a nonradiative accretion flow, highlighting its principal features. A turbulent, nearly Keplerian gas-dominated hot disk is surrounded by an active, diffuse, magnetic-dominated coronal envelope. Near the marginally stable orbit, the flow thickens into a small inner torus. A centrifugally-evacuated funnel lies along the axis, surrounded by a jet confined by magnetic pressure in the corona. From Hawley & Balbus [24].

and toroidal fields, the stress actually increases somewhat in the plunging region. The nature of the flow shifts from MHD turbulence to flux freezing, and in doing so drags out and shears the field lines, increasing the correlation between B_R and B_ϕ, and hence the Maxwell stress. Variability over large spatial scales is the rule here, over a broad range of time scales.

The general accretion problem occurs over much larger scales, at least hundreds of gravitational radii. This global nonradiative MHD accretion flow, first considered in two dimensions by Stone & Pringle [52], was extended to three dimensions in Hawley, Balbus & Stone [25] and Hawley & Balbus [24]. The accretion flow originates with a constant specific angular momentum torus initially centered at 100 gravitational radii. MHD turbulence ensues and the resulting flow seems to settle into three well-defined dynamical structures. The main accretion is through a hot, thick, rotationally-dominated Keplerian disk. The MRI is so efficient at transporting angular momentum, the transformation from constant angular momentum to Keplerian profile occurs within a few orbital times at the pressure maximum. Evidently, Keplerian profiles are characteristic of warm and cool disks alike. The total pressure scale height in this disk is comparable

to the vertical size of the initial torus. Surrounding this disk is a magnetized corona with vigorous circulation and possibly outflow. Gas pressure dominates only near the equator; magnetic pressure is more important in the surrounding corona. This coronal pressure confines a jet-like outflow, pinning it against the centrifugal funnel wall. (Though the Z axis has been removed from the computational grid, the jet seems to be a genuine dynamic feature, and is associated with an effective equipotential surface.)

Runs with and without Ohmic heating were performed; very few differences were found. The flow was somewhat hotter in the resistive runs, and the turbulence slightly subdued, but the dynamics remain dominated by MHD turbulence. There was no evidence of a convective envelope (Abramowicz et al. [2]).

Figure 2 shows schematically the appearance of the disk, corona, and jet as they appear in a typical run. The inner torus is pressure-thickened, but remains predominately supported by rotation. It is a highly transitory structure, forming and collapsing over the course of the simulation.

Contours of specific angular momentum are shown in Fig. 3. The disk emerges sharply in this diagram as the zone of cylindrical contours. This appears to be a consequence of the relatively small magnetic to thermal energy density ratio, and a barotropic equation of state.

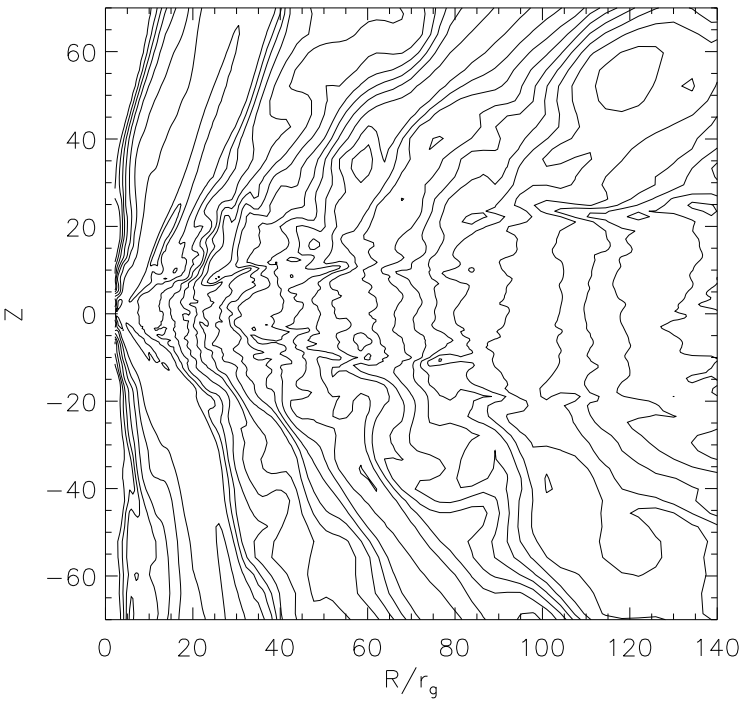


Fig. 3. Contours of azimuthally-averaged specific angular momentum. The disk is associated with the contours stratified nearly on cylinders. From Hawley & Balbus [24].

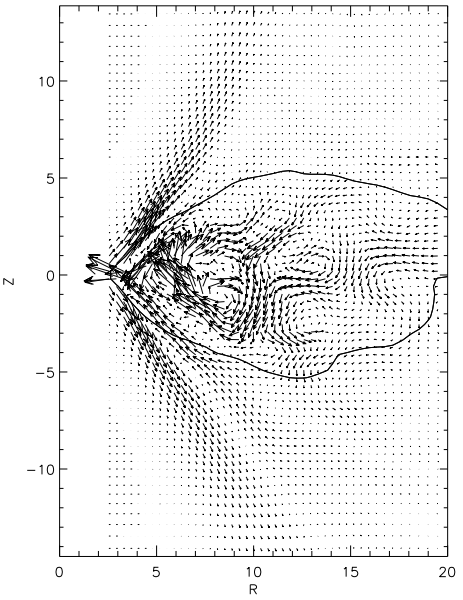


Fig. 4. Azimuthally-averaged momentum vectors in the within 20 gravitatinal radii of the black hole. The overlaid contour is density, showing the shape of the inner torus. The jet outflow along the centrifugal barrier is clearly evident. The magnetic pressure in the corona confines the jet externally. From Hawley & Balbus [24].

Figure 4 is a momentum plot of the inner 20 gravitational radii of the disk, showing the internal dynamical structure of the inner torus. The distinct jet structure stands out particularly clearly.

5 Summary

Our understanding of accretion phenomena has grown enormously in the past decade, and the physical process underlying the anomalous viscosity of accretion disks has been elucidated. Indeed, "anomalous viscosity" is a misnomer, since magnetic fields do so much more than fill the role occupied by a hydrodynamical Navier-Stokes viscosity, and we should move away from this mode of thought. Magnetized fluids are far too subtle for this approach to be successful.

The combination of a magnetic field and outwardly decreasing differential rotation is highly unstable, as is the combination of a magnetic field and an upwardly decreasing temperature profile. Local simulations of Keplerian disks show that this magnetorotational instability leads to a turbulence-enhanced stress tensor that transports energy and angular momentum outwards, allowing accretion to proceed. The typical dimensionless value of the stress (normalized to a fiducial pressure) ranges from 5×10^{-3} to 0.6 depending upon field geometry, and is highly variable in both space and time. Local simulations have become very sophisticated in the class of problems they are able to investigate. Work has

begun on the local disk dynamics of radiation-dominated and non-ideal MHD systems.

Fully global three-dimensional MHD simulations are now a reality. These runs show that initially non-Keplerian angular momentum distributions rapidly evolve to Keplerian. Efficient angular momentum transport builds up rotation in the outer regions in the early stages, and the gas expands. Any initial radial pressure gradient is "inflated" to zero, and a Keplerian distribution emerges.

Chandra observations have provided compelling evidence that very low luminosity accretion flows are present in Nature, and such flows are amenable to numerical investigation. The most detailed studied to date (Hawley & Balbus [24]) reveals a three component structure: a warm Keplerian disk, a highly magnetized corona, and an axial jet. None of these structures were present in the initial condition, which is a simple constant angular momentum torus located 100 gravitational radii from the hole. Significant fluctuations in all flow quantities are present, both in time and in space. There was no indication that thermal convection was dominating the dynamical flow structure (Abramowicz, et al. [2]).

The results of global three-dimensional MHD numerical simulations comprise a vast and extremely useful data base, a true numerical laboratory. At the time of this writing, relatively little has been done to translate the formal numerical flows into photons impinging upon instruments. The exploitation of this largely untapped resource should provide advances as singular as those of the last extraordinary decade.

Acknowldgements

Our research is supported by NASA grants NAG-10655, NAG5-9266, and NSF grant AST-0070979. The authors are most grateful to the editors E. Falgarone and T. Passot for their patience and forbearance.

References

1. Abramowicz, M. A., Czerny, B., Lasota, J.-P., & Szuszkiewicz, E. 1988, ApJ, 332, 646
2. Abramowicz, M. A., Igumenschev, I. V., Quataert, E., & Narayan, R. 2002, ApJ, 565, 1101
3. Algol, E., & Krolik, J. H. 2000, ApJ, 528, 161
4. Armitage, P. 1998, ApJ, 501, L189
5. Armitage, P. 2002, MNRAS, in press (astro-ph/0110670)
6. Balbus, S. A. 1995, ApJ, 453, 380
7. Balbus, S. A. 2001, ApJ, 562, 909
8. Balbus, S. A., & Hawley, J. F. 1991, ApJ, 376, 214
9. Balbus, S. A., & Hawley, J. F. 1992, ApJ, 392, 662
10. Balbus, S. A., & Hawley, J. F. 1998, RMP, 70, 1
11. Balbus, S. A., & Terquem, C. 2001, ApJ, 552, 235
12. Bender, C. M., & Orszag, S. A. 1978, Advanced Mathematical Methods for Scientists and Engineers (New York: McGraw–Hill), p. 560

13. Begelman, M. C., & Meier, D. L. 1982, ApJ, 253, 873
14. Blaes, O. M., & Balbus, S. A. 1994, ApJ, 421, 163
15. Blaes, O., & Socrates, A. 2001, ApJ, 553, 987
16. Brandenburg, A., Nordlund, Å, Stein, R. F. & Torkelsson, U. 1995, ApJ, 446, 741
17. Fleming, T. P., Stone, J. M., & Hawley, J. F. 2000, ApJ, 530, 464
18. Gammie, C. F. 1999, ApJ, 522, L57
19. Goodman, J., & Xu, G. 1994, ApJ, 432, 213
20. Hawley, J. F. 2000, ApJ, 528, 462
21. Hawley, J. F. 2001, ApJ, 554, 534
22. Hawley, J. F., & Balbus, S. A. 1991, ApJ, 376, 223
23. Hawley, J. F., & Balbus, S. A. 1992, ApJ, 400, 595
24. Hawley, J. F., & Balbus, S. A., 2002, ApJ, submitted
25. Hawley, J. F., Balbus, S. A., & Stone, J. M. 2001, ApJ, 554, L49
26. Hawley, J. F., Gammie, C. F., & Balbus, S. A. 1995, ApJ, 440, 742
27. Hawley, J. F., Gammie, C. F., & Balbus, S. A. 1996, ApJ, 464, 690
28. Hawley, J. F, & Krolik, J. H. 2001, ApJ, 548, 348
29. Hawley, J. F, & Krolik, J. H. 2002, ApJ, 566, 164
30. Ichimaru, S. 1977, ApJ, 214, 840
31. Krolik, J. H. 1999, Active Galactic Nuclei (Princeton; Princeton Univ. Press)
32. Machida, M., Hayashi, M. R., & Matsumoto, R. 2000, ApJ, 532, 67
33. Matsumoto, R., & Tajima, T. 1995, ApJ, 445, 767
34. Matsuzaki, T., Matsumoto, T., Tajima, T., & Shibata, K. 1997, in Accretion Phenomena and Related Outflows, eds. D. Wickramsinghe, L. Ferrario, and G. Bicknell (San Francisco: ASP), p. 766
35. Melia, F., & Falcke, H. 2001, ARA&A, 39, 309
36. Moffatt, K. 1978, Magnetic Field Generation in Electrically Conducting Fluids, (Cambridge: Cambridge Univ. Press), p. 113
37. Miller, K. & Stone, J. M. 2000, 534, 398
38. Narayan, R., Igumenschev, I. V., & Abramowicz, M. A. 2000, ApJ, 539, 798
39. Narayan, R., & Yi, I. 1994, ApJ, 428, L13
40. Narayan, R., & Yi, I. 1995, ApJ, 452, 710
41. Novikov, I. D., & Thorne, K. S. 1973, in Black Holes—Les Astres Occules, ed. C. DeWitt (New York: Gordon and Breach)
42. Paczyński, B., & Wiita, P. 1980, A&A, 88, 23
43. Pringle, J. E., & Rees, M. J. 1972, A&A, 21, 1
44. Rees, M. J., Phinney, E. S., Begelman, M. C., & Blandford, R. D. 1982, 1982, Nature, 295, 17
45. Sano, T., & Stone, J. M. 2002, ApJ, submitted (astro-ph/0201179)
46. Shakura, N. I., & Sunyaev, R. A. 1973, A&A, 24, 337
47. Shibata, K., & Uchida, Y. 1986, PASJ, 38, 631
48. Steinacker, A., & Henning, T. 2001, ApJ, 554, 514
49. Steinacker, A., & Papaloizou, J. C. B. 2002, ApJ, in press
50. Stone, J. M., & Balbus, S. A. 1996, ApJ, 464, 364
51. Stone, J. M., Hawley, J. F., Gammie, C. F. & Balbus, S. A. 1996, ApJ, 463, 656
52. Stone, J. M., & Pringle, J. E. 2001, MNRAS, 322, 461
53. Stone, J. M., Pringle, J. E., Begelman, M. C. 1999, MNRAS, 310, 1002
54. Tassoul, J.–L. 1978, *Theory of Rotating Stars*, (Princeton University Press: Princeton)
55. Turner, N. J., Stone, J. M., & Sano, T. 2002, ApJ, 566, 148
56. Uchida, Y., & Shibata K. 1985, PASJ, 37, 31
57. Wardle, M. 1999, MNRAS, 307, 849

Recent Developments in Collisionless Reconnection Theory: Applications to Laboratory and Astrophysical Plasmas

Amitava Bhattacharjee[1], Zhiwei W. Ma[1], and Xiaogang Wang[2]

[1] Center for Magnetic Reconnection Studies, Department of Physics and Astronomy, University of Iowa, Iowa City, IA 52242, USA
[2] Department of Physics, Dalian University of Technology, Dalian, China 116024

Abstract. Recent developments in the theory and simulation of nonlinear collisionless reconnection hold the promise for providing solutions to some outstanding problems in laboratory and astrophysical plasma physics. Examples of such problems are sawtooth oscillations in tokamaks, magnetotail substorms, and impulsive solar and stellar flares. In each of these problems, a key issue is the identification of fast reconnection rates that are insensitive to the mechanism that breaks field lines (resistivity and/or electron inertia). The classical models of Sweet-Parker and Petschek sought to resolve this issue in the realm of resistive magnetohydrodynamics (MHD). However, the plasmas mentioned above are weakly collisional, and hence obey a generalized Ohm's law in which the Hall current and electron pressure gradient terms play a crucial role. Recent theoretical models and simulations on impulsive (or triggered) as well as quasi-steady reconnection governed by a generalized Ohm's law are reviewed. In the impulsive reconnection problem, not only is the growth rate fast but the time-derivative of the growth rate changes rapidly. In the steady-state reconnection problem, explicit analytical expressions are obtained for the geometric characteristics (that is, length and width) of the reconnection layer and the reconnection rate. Analytical results are tested by Hall MHD simulations. While some of the geometric features of the reconnection layer and the weak dependence of the reconnection rate on resistivity are reminiscent of Petschek's classical model, the underlying wave and particle dynamics mediating the reconnection dynamics in the presence of the Hall current and electron pressure gradient are qualitatively different. Quantitative comparisons are made between theory and observations. Open and unresolved issues are identified.

1 Introduction

During the last four decades, the Sweet-Parker [1,2] and the Petschek model [3] of steady-state reconnection have been the focal points of research on nonlinear reconnection dynamics. Both models were motivated by observations of solar flares and based on resistive magnetohydrodynamics (MHD). To fix ideas, let us consider a sheared magnetic field,

$$\mathbf{B} = B_P \tanh(z/a)\hat{\mathbf{x}} + B_T\hat{\mathbf{y}}, \tag{1}$$

where B_P and B_T are positive constants. The poloidal component of the magnetic field changes sign across the so-called neutral line at $z = 0$. In the Sweet-Parker model, assuming that the plasma is incompressible, steady-state reconnection occurs in the vicinity of the neutral line on a characteristic time scale

E. Falgarone and T. Passot (Eds.): LNP 614, pp. 351–375, 2003.
© Springer-Verlag Berlin Heidelberg 2003

$\tau_{SP} \equiv (\tau_A \tau_R)^{1/2} = S^{1/2}\tau_A$ where $\tau_A \equiv a/v_A = a(4\pi\varrho)^{1/2}/B_P$ is the poloidal Alfvén time, $\tau_R \equiv 4\pi a^2/\eta c^2$ is the resistive diffusion time, and $S \equiv \tau_R/\tau_A$ is the Lundquist number. (Here ϱ is the mass density, η is the resistivity of the plasma, and c is the speed of light.) In the Sweet-Parker model, the reconnection layer has a Y-point geometry [4] and its length is of the order of the system size. For weakly collisional systems such as the solar corona, the Lundquist number S is typically very large ($\sim 10^{12} - 10^{14}$) and hence, the time scale τ_{SP} is of the order of hours. Since τ_{SP} is much too long to account for fast events such as solar flares, Petschek proposed another steady-state model which maintains an X-point geometry for all times. In contrast with the Sweet-Parker model, Petschek's model yields a reconnection time scale that has a weak logarithmic dependence on S. For the high-S solar corona, the Petschek time scale is of the order of minutes, much closer to the relevant time scale for flares.

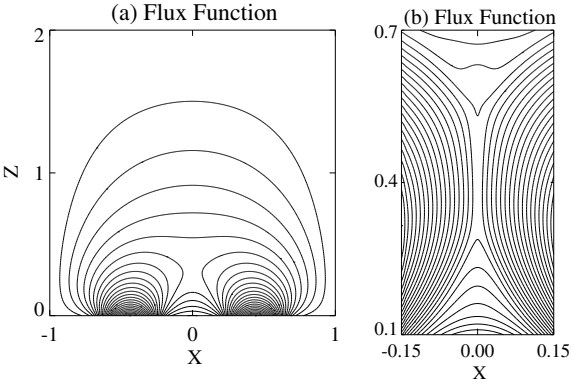


Fig. 1. (a) Flux function for a numerically generated vacuum solution of a two-dimensional coronal arcade containing an X-point in the initial state. (b) Flux function for a high-Lundquist-number resistive MHD simulation in the nonlinear regime, showing the spontaneous formation of Y points.

Since the mid-eighties, high-resolution computer simulations have produced some of the most illuminating tests of Petschek's model. A persistent feature seen in these simulations is that in the high-S regime, even if one begins with an equilibrium state containing an X-point that would appear to favor Petschek, one ends up obtaining an extended reconnection layer with Y-point structure typical of Sweet-Parker. Figure 1, taken from simulations by Ma *et al.*[5], shows this feature. The initial configuration, shown in Fig. 1a, represents a two-dimensional vacuum magnetic field in the xz-plane containing an X-point. Reconnection is induced in this configuration by the motion of magnetic field footpoints on the photosphere, idealized here as the $z = 0$ plane. Figure 1b shows a plot of the flux function for a high-S simulation in the nonlinear regime in the vicinity of the reconnection layer. We note that the initial X-point geometry has evolved to form Y-points, and as discussed by Ma *et al.*, reconnection occurs on the Sweet-

Parker time scale in the nonlinear regime. Thus, the classical models of Sweet-Parker and Petschek leave us with a quandary for high-S plasmas. Whereas the Sweet-Parker time scale is realizable dynamically in the high-S regime, it is too slow. On the other hand, the Petschek model, which yields a faster time scale, appears not to be realizable in the high-S regime.

In addition to the problem discussed above, there is a subtler issue that is impossible to resolve within the framework of all steady-state reconnection models like Sweet-Parker and Petschek. By definition, steady-state models can provide but one time scale—that of steady reconnection (proportional to $S^{1/2}$ for Sweet-Parker and $\ln S$ for Petschek). Steady reconnection is not a generic condition. It is a strong theoretical assumption and one that is frequently violated in many dynamical situations of great physical interest. In particular, there are phenomena involving magnetic reconnection in laboratory as well as space plasmas where the dynamics exhibits an impulsiveness, that is, a sudden increase in the time-derivative of the growth rate. This is often referred to as the "trigger problem"—-the magnetic field configuration evolves slowly for a long period of time, only to undergo a sudden dynamical change over a much shorter period of time. As the classical steady-state reconnection models of Sweet-Parker and Petschek do not include time-dependency, they cannot account for the time-evolution of the growth rate.

In this paper, we review recent developments in collisionless reconnection theory that hold the promise for providing solutions to some of the outstanding problems mentioned above. The plasmas in high-temperature tokamaks, the Earth's magnetotail or the solar corona are characterized by very different plasma parameters, geometry, and boundary conditions, but they have one feature in common: their classical Lundquist number is often very high ($S \geq 10^8$). We describe such plasmas as "collisionless." As discussed below, these plasmas tend to develop thin and intense current sheets in the reconnection layer. As these thin current sheets become localized and intense, and their width $\triangle_\eta$ falls in the collisionless range $d_e \equiv c/\omega_{pe} < \triangle_\eta \leq d_i \equiv c/\omega_{pi}$ (where ω_{pe} and ω_{pi} are the electron and ion plasma frequencies, respectively), it is not valid to neglect the collisionless terms in the generalized Ohm's law

$$\mathbf{E} + \frac{\mathbf{v} \times \mathbf{B}}{c} = \eta \mathbf{J} + \frac{4\pi}{{\omega_{pe}}^2} \frac{D\mathbf{J}}{Dt} - \frac{\nabla p}{ne} + \frac{\mathbf{J} \times \mathbf{B}}{nec}, \tag{2}$$

where $\mathbf{E}$ is the electric field, $\mathbf{B}$ is the magnetic field, $\mathbf{v}$ is the plasma flow velocity, $\mathbf{J}$ is the current density, p is the electron pressure (assumed to be a scalar), n is the electron density, e is the magnitude of the electron charge, and $D/Dt \equiv \partial/\partial t + \mathbf{v} \cdot \bigtriangledown$ is the total convective derivative. By collisionless terms, we mean the last three terms on the right-hand-side of (2). The second term is proportional to finite electron inertia, the third to the electron pressure gradient, and the fourth to the Hall current. To keep the terminology simple, we will refer to the last two terms on the right-hand-side of (2) collectively as the Hall MHD terms. In (2), we have retained the resistive diffusion as well as the electron inertia term both of which can break field lines. If the electron pressure were a tensor, it could also provide another mechanism for breaking field lines. A scalar

electron pressure (with simple adiabatic closure) cannot break field lines, but can, nonetheless, contribute significantly to the parallel electric field. We demonstrate by analysis and high-resolution numerical simulation that Hall MHD terms alter qualitatively the geometrical and temporal features of reconnection dynamics obtained from resistive MHD. In particular, we show that Hall MHD effects provide not only a fast reconnection time scale but also a plausible explanation of the trigger problem. It is thus possible to account for significant observational features of magnetic reconnection in different physical contexts from a common perspective.

In the spirit of a review, we provide some historical perspective. The influence of the Hall current and electron pressure gradient terms on nonlinear reconnection dynamics has been the focus of extensive research since the early nineties in fusion [6-12] as well as space physics [13-21]. The impetus for these studies have come primarily from two sources. The first of these was an attempt to understand the so-called "trigger problem" for sawtooth oscillations in tokamaks [22] which led to the discovery of the role of the electron pressure gradient as a trigger in a simulation [6] and its subsequent explanation by an analytical model [7]. The second has been a broader search for the underlying mechanism(s) of quasi-steady fast collisionless reconnection in space plasmas, such as in the magnetopause and the magnetotail. This search has recently culminated in the Global Environment Modeling (GEM) challenge in which a suite of particle, hybrid and Hall MHD codes were run with the same initial condition and shown to be in agreement in their predictions of the steady-state reconnection rate. (See [23] for an overview of the GEM challenge.) Whereas the work on the trigger problem in tokamaks, which have a strong toroidal (or guide) field, emphasized the role of the electron pressure gradient, the GEM challenge, which employed an initial condition without a guide field, emphasized the role of the Hall current.

In the present paper, we will review these recent developments on nonlinear collisionless reconnection, following a historical chronology. We begin in Sect. 2 with a discussion of the trigger problem for sawtooth oscillations in a tokamak. In Sect. 3, we discuss the problem of impulsive enhancement of the current density at near-earth distances in the magnetotail just before substorm onset, and explain how Hall MHD provides a satisfactory explanation of this enhancement. In Sect. 4, we review recent developments in models of quasi-steady collisionless reconnection. In Sect. 5, we discuss the problem of a fast trigger for flare dynamics. In each of the sections, we mention open questions that remain a challenge for theory. We conclude in Sect. 6 with a summary.

2 Trigger for Sawtooth Collapse in High-Temperature Tokamaks

Sawtooth crashes in tokamaks are characterized by a quasi-periodic collapse of the plasma temperature in the core. During a sawtooth cycle, the hot core plasma deforms under the influence of a non-ideal helical instability. The sawtooth crash occurs suddenly, producing a nearly flattened temperature profile. After the

crash, the core temperature recovers slowly. While the entire sawtooth cycle lasts about $100 - 200ms$, the crash occurs in about $120\mu s$.

These observations pose a substantial challenge for reconnection theory. In the seventies, essential features of the data available then were thought to be explained by the Kadomtsev model [24]. To fix ideas, we will assume that a toroidal tokamak of minor radius a and major radius R_0 can be modeled as a straight cylindrical plasma of radius a and periodicity length $2\pi R_0$ along which there is a strong axial magnetic field B_T. (In other words, the curved toroidal magnetic field in the ϕ-direction is now assumed to be the axial magnetic field in a cylinder which, to be consistent with (1), is assumed to point in the y-direction.) The q-profile is defined as $q(r) = rB_T/RB_\theta(r)$ where r is the radial coordinate in the straight cylinder. For equilibria that depend only on the radius r, normal modes can be written in the Fourier form $\exp(im\theta - iny/R)$, where θ is the poloidal angle. Near the rational surface for the $m = 1, n = 1$ kink-tearing mode (hereafter referred to simply as the $m = 1$ mode) where $q = 1$,we can rewrite the equilibrium magnetic field in a form essentially equivalent to (1), with z measuring the radial distance $r - r_s$ from the rational surface $r = r_s$.

Kadomtsev proposed that the nonlinear dynamics of the resistive $m = 1$ mode can account for salient features of sawtooth collapse in tokamaks. He showed that in the nonlinear regime of this mode, reconnection occurs at the separatrix on the characteristic Sweet-Parker time scale $\tau_K = (\tau_A \tau_R)^{1/2}$. For most tokamaks operating during the seventies, the time scales τ_A and τ_R were typically of the order of $10^{-7}s$ and $10^{-1}s$, respectively, which gave $\tau_K \sim 100\mu s$, in agreement with the characteristic time scale that was observed then. However, discrepancies with the predictions of the Kadomtsev model became evident in larger and hotter tokamaks such as JET and TFTR where the predicted time scale is one to two orders of magnitude larger than the observed collapse time. It is not uncommon in these devices to obtain collapse times in the range $20-100\mu s$ whereas $\tau_K \sim 2 - 10ms$.

It is possible to bring the theoretical prediction of the collapse time closer to observations by invoking a resistivity larger than classical in the reconnection layer. However, careful analysis [25], which is well-supported by numerical simulations [26,27] shows that the resistive $m = 1$ kink-tearing instability slows down nonlinearly, with the island width w increasing algebraically with time according to the relation $w \sim \eta t^2$. With such slow and gradual algebraic growth, it is not possible to account for the fast trigger of the sawtooth collapse.

Aydemir [6] made an important discovery that shows a possible way out of this theoretical impasse. He found by numerical integration of reduced two-fluid equations that parallel electron pressure gradients cause a rapid acceleration in the nonlinear growth of the $m = 1$ mode. Wang and Bhattacharjee [7] developed an analytical model, which reproduces the salient features of nonlinear $m = 1$ growth seen in Aydemir's simulation. In what follows, we review the main results obtained in [6] and [7]. (Similar results have also been obtained in [10].)

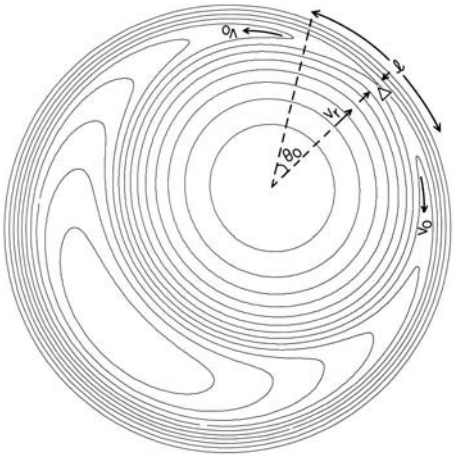


Fig. 2. Geometry of the nonlinear reconnection layer for the $m = 1$ nonideal kink mode (Courtesy A. Aydemir).

The geometry of the configuration is shown schematically in Fig. 2. Assuming that the plasma is incompressible, mass conservation implies that

$$\triangle v_0 \approx \ell v_r, \tag{3}$$

where ℓ and $\triangle$ are, respectively, the length and half-width of the reconnection layer, and v_r and v_0 are the (radial) inflow and (poloidal) outflow speeds. Note that in (1), $v_r = v_r(\theta = 0)$, and $v_0 = v_\theta(\theta_0)$, where θ_0 represents the angle subtended at the center of the plasma by the half-length of the reconnection layer. During the early nonlinear phase, the angle θ_0 decreases rapidly as the length of the thin current sheet decreases. From energy conservation arguments [7, 26, 27], it can be shown quite simply that v_0 is approximately equal to the helical Alfvén speed at the edge of the reconnection layer, defined by the relation, $v_{Ah} = (1-q)B_\theta/(4\pi\varrho)^{1/2} \approx r_s q'(r_s)B_\theta(r_s)/(4\pi\varrho)^{1/2}$. More precisely, it can be shown that $v_0 \approx (\xi/r_s)v_{Ah}$ where ξ is the radial displacement of the inner plasma column associated with the $m = 1$ mode. It can also be shown from geometric considerations [7, 29] that the width w of the $m = 1$ island is related to ξ by the relation $w = 2\xi$. Noting that $v_r = (1/2)d\xi/dt = (1/4)dw/dt$ and $v_0 = (\xi/r_s)v_{Ah}$, we obtain

$$\omega_A{}^{-1} dw/dt \approx (2\triangle/\ell)w \tag{4}$$

where $\omega_A = v_{Ah}/r_s$. As we explain below, in the presence of the electron pressure gradient term in the generalized Ohm's law, we obtain $\triangle \propto w$ which causes a near-explosive growth phase. In this phase the island width tends to blow up as inverse time, that is, $w \propto (t_c - t)^{-1}$ where t_c is a positive constant. (This should be contrasted with algebraic nonlinear growth $w \sim \eta t^2$ discussed above.)

To demonstrate that $\triangle \propto w$, we need to use an approximate integral form of the parallel component of the momentum equation [7]:

$$\int_0^{\theta_0} \frac{\partial}{\partial\theta}(p + \frac{1}{2}\varrho {v_\theta}^2) \approx 0.$$

To require that the integrand itself vanishes is a stronger requirement, used in our earlier studies [7], but not actually necessary. We now calculate the time-evolution of magnetic flux in the ideal region (where convection dominates) and the reconnection layer (where the electron pressure gradient dominates) and match the two results in the asymptotic overlap region to obtain

$$\frac{v_r(\theta)}{c} \approx -\frac{1}{B_T r_s n e}\frac{\partial p}{\partial\theta}. \tag{5}$$

Integrating (5) from 0 to θ_0 and using (3), we obtain

$$\triangle \approx v_0/(2\Omega_i)) \approx w/(4r_s v_{Ah}\Omega_i), \tag{6}$$

where $v_r = v_r(\theta = 0)$ and $\Omega_i = eB_T/(m_i c)$ is the ion Larmor frequency. We have thus shown that $\triangle \propto w$, as claimed earlier. Under these conditions, the island width shows a tendency to blow up in finite-time. This explosive growth accounts for the sawtooth trigger.

It should be noted that the tendency for finite-time blowup discussed above is merely a tendency and actually arrested at later times. This is because the derivation given above assumes that the $m = 1$ island is thin, that is, $X \equiv w/2r_s << 1$. This assumption breaks down when the island size becomes a significant fraction of the plasma radius. For large islands, Wang and Bhattacharjee [7] show that the island equation is given by

$$\frac{dX}{d\tau} \cong \frac{1}{2}\left(\frac{\omega_A}{\Omega_i}\right)^{1/2} X^2(1-X)^{3/2} + \frac{2d_e}{r_s}X\left\{\frac{\Omega_i}{\omega_A}(1-X)\right\}^{1/2}, \tag{7}$$

where $\tau \equiv \omega_A t$. The nonlinear growth rate $\gamma \equiv d\ln X/d\tau$ of the island is computed from (7). Figure 3 shows the time-history of the island growth rate, from its early exponential to near-explosive growth, followed by the decay phase. This picture shows a strong qualitative resemblance to the numerical results of Aydemir [6] based on reduced two-fluid equations.

There are three principal observational features of sawtooth oscillations that a viable theory must explain. The first is the rapidity of the sawtooth collapse, the second is the sudden onset of the collapse, and the third is the incomplete relaxation of the current profile during which the q-value at the center of the plasma column remains nearly fixed at a value lower than unity while the temperature profile relaxes completely [28-30]. Although collisionless reconnection mediated by the electron pressure gradient appears to provide a satisfactory account of the first two features, it fails to explain the third. It is possible that a pressure-driven instability such as the ballooning instability intervenes by affecting the electron pressure gradient that controls the reconnection process, but

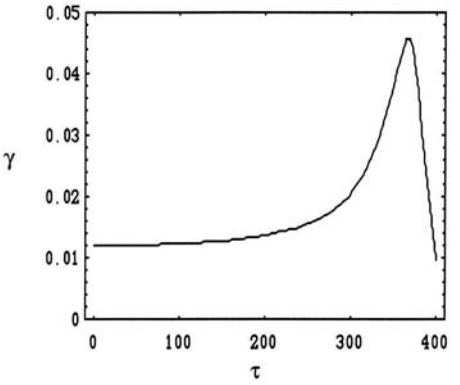


Fig. 3. The nonlinear growth rate for the $m = 1$ island width predicted by the analytical model of [7].

more theoretical work needs to be done to determine if this scenario provides an explanation of the third feature.

Having discussed an example of impulsive or triggered reconnection in a laboratory plasma, we now move on to consider a similar phenomenon in the Earth's magnetotail during substorms.

3 Sudden Enhancement of Thin Current Sheets in the Earth's Magnetotail Prior to Substorm Onset

The key to substorm onset lies in the growth phase when the magnetotail is prepared for the relaxation dynamics that follows. During the growth phase, it is now widely accepted that the cross-tail current intensifies, resulting in substantial stretching of the nightside magnetic field lines at near-Earth distances ($< 10R_E$ where R_E denotes the radius of the earth) [31]. Observations show that the thickness of the thin current sheet can be reduced to less than $1R_E$ in the near-Earth region before the onset of the expansion phase [32,33]. The onset of the expansion phase involves a sudden reduction (usually referred to as disruption) of the cross-tail current density in the near-Earth region ($7 - 11R_E$). The current sheet disruption region is localized ($< 1R_E$) and the disruption of the current is also partial, typically involving about 20% of the cross-tail current. At the onset of the expansion phase, the magnetic field configurations at low latitudes becomes dipolar, and the plasma sheet expands.

Multi-point observations of current disruption at near-Earth distances [34,35] provide significant constraints on theoretical models of substorms. From local magnetic field and particle measurements, it is inferred in these observations that, after a period of sluggish growth ($\sim 0.5 - 1.5hr$), the cross-tail current density exhibits rapid, impulsive growth during a short interval ($<$ 1min) just before the onset of the expansion phase (although the identification of the second phase is not always unambiguous). Following the impulsive growth phase, the

current disrupts on a very short time scale ($\sim 10s$). These results, are of great interest because they show the presence of two time scales in the growth phase, followed by a third time scale during which rapid disruption of the tail current occurs.

In this section, we present the results of a numerical simulation of 2D substorm dynamics in the Earth's magnetotail using the compressible Hall MHD equations beginning from a realistic magnetotail equilibrium [36]. We assume that y is an ignorable coordinate. In order to maintain adequate spatial resolution in the high-S regime, we consider a two-dimensional domain on the nightside, extending from $6R_E$ to distances just above $46R_E$ down the tail. All variables are cast in dimensionless form: in particular, distances are scaled by $1R_E$, and time is scaled by the characteristic Alfvén time $\tau_A = 1R_E/V_A(\sim 6s)$ where the Alfvén speed is computed at the corner $x = -6R_E, z = 12R_E$ of the computational domain. The magnetic field $\mathbf{B}$ is represented as $\mathbf{B} = \nabla\psi \times \hat{\mathbf{y}} + B_y\hat{\mathbf{y}}$. The initial equilibrium is obtained by solving the Grad-Shafranov equation

$$\nabla^2\psi = J_y(\psi) = -\frac{dp(\psi)}{d\psi}, \tag{8}$$

numerically with $B_y = 0$. (The inclusion of a B_y field in the initial equilibrium does not change our results qualitatively.) We start with a flux function ψ_0 that reproduces the magnetic field in the Earth-noon meridian plane of the 1989 Tsyganenko model [37], specify the cross-tail current $J_y(\psi)$ as a flux function,

$$J_y(\psi) = -J_0 - J_d{\psi_d}^3/(\psi^2 + {\psi_d}^2)^{3/2}, \tag{9}$$

where J_0, J_d and ψ_d are constants, and iterate numerically the equilibrium equation $\nabla^2\psi_{n+1} = J_y(\psi_n)$ beginning from the initial source term $J_y(\psi_0)$ until the convergence condition $\mid \nabla^2\psi_n - J_y(\psi_n) \mid < 10^{-9}$ is satisfied. Such an equilibrium is generally stable with respect to resistive tearing instabilities at near-Earth distances due to the presence of a significant B_z-component. Since free reconnection is thus ruled out, we are motivated to investigate the role of forced reconnection. An electric field, given by

$$E = E_0\{0.6 + 0.4\tanh[(3(x_1 + x) + 2x_2)/x_2]\}, \tag{10}$$

is imposed in the x-direction at the upper and lower boundaries at $z = \pm 12R_E$, where $E_0 = 0.06$, $x_1 = 6R_E$, and $x_2 = 40R_E$. The functional form (10) is a simple representation of an electric field that peaks at near-earth distances and decays monotonically with increasing $|x|$. The maximum equatorward inflow v_z due to the electric field (10) is of the order of 60km/s, which is less than 6% of the characteristic Alfvén speed ($\sim$ 1000km/s), roughly consistent with observations. Free boundary conditions are used at the earthward as well as tailward boundaries of the simulation box for all dependent variables except the flux function ψ which obeys $d\psi/dt = 0$.

In the simulations that follow we have omitted the electron inertia term in (2). This is because numerical constraints limit the maximum Lundquist number

in our simulations, including the effect of numerical diffusion, to 10^5. At such values of S, the effect of electron inertia on the simulation dynamics cannot be resolved numerically. It is possible that at much higher values of the Lundquist number, typical of the magnetotail but beyond the scope of the present study, electron inertia may have significant effects in the growth phase of the thin current sheet.

In Fig. 4, we show a plot of the level surfaces of the cross-tail current density. In early stages of the simulation, an X-type neutral line with an extremely small separatrix angle is formed in the region $x = -30R_E$ where the B_z-field is initially weak. Evidence for the formation of such an X-line during the growth phase is strong in recent GEOTAIL observations [38], and consistent with one of the predictions of the so-called near-Earth neutral line model of substorms [39]. After a period of slow algebraic growth in time, the cross-tail current density at near-Earth distances exhibits a sudden impulsive enhancement at a sub-Alfvénic growth rate that is insensitive to the value of the Lundquist number. This is demonstrated in Fig. 5 where we plot the time evolution of the cross-tail current density J_y at a typical near-Earth distance $x = -12R_E$ near the center of the plasma sheet for $S = 10^5$ (solid line), 2×10^4 (dotted line), and 10^4 (dashed line). Note in all cases an impulsive enhancement of the amplitude of the current density at $t \sim 300\tau_A$, with the rate of enhancement weakly dependent on the magnitude of the Lundquist number. Due to the inclusion of the Hall terms in the generalized Ohm's law, the time of impulsive enhancement is of the order of 1 min., which is comparable with the time scale reported in [37]. This feature of impulsive enhancement is qualitatively similar to that found in our earlier Hall MHD simulation of a Harris sheet [16].

The cross-tail current density is enhanced by an order of magnitude compared with its value at the end of the slow growth phase and the thickness of the current sheet is reduced approximately to $0.2R_E$. The geometry of the thin current sheet and reconnection region undergoes rapid temporal changes, evolving from an extended Y-point structure to a structure shorter in length and thinner than that seen in a resistive MHD simulation run with the same initial condition [40].

In the present simulation, the impulsive growth phase ends with a partial current disruption, brought about by near-Earth reconnection. However, we cannot conclude that reconnection is a mechanism for substorm onset at near-Earth distances because two significant discrepancies with observations remain. First, the time scale of current disruption and dipolarization in our simulation is of the order of several minutes, whereas observations appear to suggest that the time scale is of the order of ten seconds. In other words, the time scale observed in our simulation at near-Earth distances is too slow by nearly an order of magnitude when compared with observations, and this discrepancy cannot be cured by simply appealing to a larger Lundquist number. Second, the magnetic field in the simulations does not truly dipolarize. Since our simulation is two-dimensional (that is, y is an ignorable coordinate), three-dimensional instabilities that break the y-symmetry are excluded, and such instabilities can alter qualitatively the dynamics seen in the late stages of our simulation. For example, we have demon-

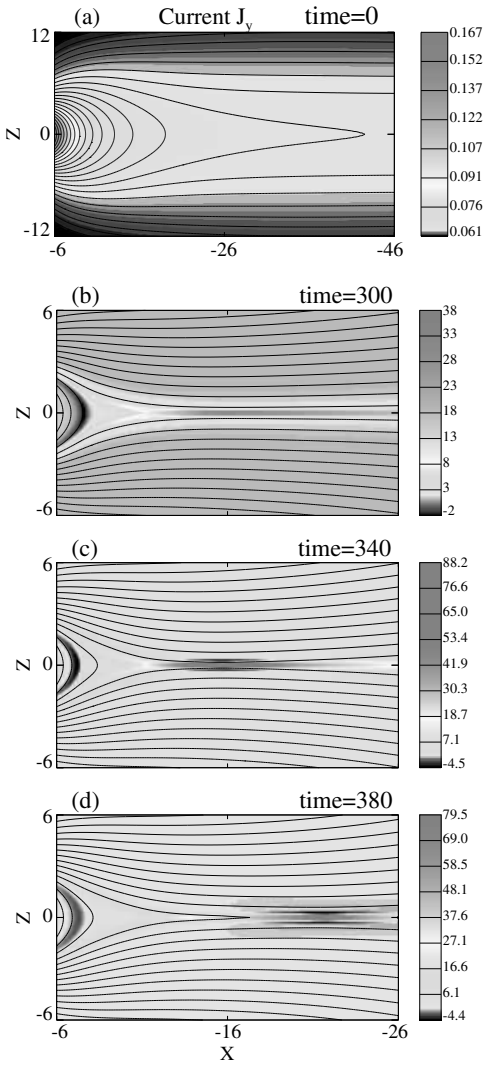


Fig. 4. Projection of level surfaces of the cross-tail current density J_y on the $x - z$ plane at different times during the time development of thin current sheets. The solid black lines represent the projection of flux surfaces on the $x - z$ plane.

strated elsewhere [41] that due to the build-up of strong pressure gradients in the impulsive growth phase at near-Earth distances, the thin current sheet is unstable to a rapidly growing linear ideal ballooning instability (with very large wavenumber along y). On the basis of an asymptotic reduction of the full MHD equations using the ballooning-mode representation, Hurricane *et al.* [42] suggest t hat such a linear instability will grow explosively in the nonlinear regime. While these results are promising, more analytical and numerical work needs to

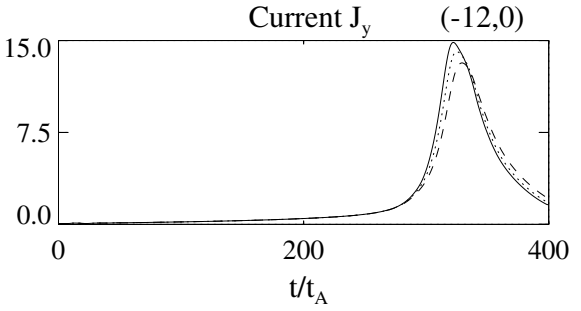


Fig. 5. Time history of J_y at a typical near-Earth distance $x = -12R_E, z = 0$ near the center of the plasma sheet.

be done with the *full* nonlinear MHD equations in realistic magnetotail geometry before one can claim that ballooning is indeed a mechanism for dipolarization and current diversion at substorm onset.

4 Quasi-steady Reconnection

In this Section, we discuss quasi-steady collisionless reconnection in which both the Hall current and the electron pressure gradient are treated on the same footing. This complements our earlier discussion on the trigger problem given in Sects. 2 and 3. In Sect. 4.1, we discuss the dispersion relations of electromagnetic waves that mediate collisionless reconnection for the initial equilibrium (1), including a non-zero guide field and the effects of plasma beta. When $B_T = 0$, the reconnection layer supports obliquely propagating Alfvén-whistler waves, and the Hall current dominates reconnection dynamics. When $B_T/B_P \geq 1$, the reconnection layer supports kinetic/inertial Alfvén waves, and reconnection dynamics is dominated by the electron pressure gradient. In Sect. 4.2, on the basis of the scaling properties of the generalized Ohm's law, we estimate the nonlinear reconnection rate which, at leading order, is controlled by ions and is independent of the mechanism that breaks field lines (resistivity or electron inertia). We show that this leading-order result depends on a combination of local parameters, such as the ion skin depth, and global parameters such as those determined by the boundary conditions driving reconnection. In Sect. 4.3, we illustrate and test the predictions of the analytical model with a Hall MHD code.

4.1 Dispersion Relations

The equations of Hall MHD, cast in dimensionless form, are the generalized Ohm's law

$$\mathbf{E} + \mathbf{v} \times \mathbf{B} = \frac{1}{S}\mathbf{J} + \frac{{d_e}^2}{n}\frac{d\mathbf{J}}{dt} + \frac{d_i}{n}\Big(\mathbf{J} \times \mathbf{B} - \beta_{ep}\nabla p_e\Big), \tag{11}$$

the equation of state,

$$\frac{dp_e}{dt} = \frac{d_i}{n}\mathbf{J}\cdot\nabla p_e - \gamma p_e \nabla\cdot\mathbf{v}_e, \tag{12}$$

and the momentum equation,

$$\frac{d\mathbf{v}}{dt} = \frac{1}{n}(\mathbf{J}\times\mathbf{B} - \beta_p\nabla p), \tag{13}$$

where $d/dt = \partial/\partial t + \mathbf{v}\cdot\nabla$. To (4)-(6), we add Faraday's and Ampere's laws

$$\frac{\partial\mathbf{B}}{\partial t} = -\nabla\times\mathbf{E}, \tag{14}$$

$$\mathbf{J} = \nabla\times\mathbf{B}, \tag{15}$$

respectively. In (4)-(8), we have redefined the following variables to make them dimensionless:

$$\begin{aligned}
&c\mathbf{E}/(B_P v_A)\to\mathbf{E}, \mathbf{B}/B_P\to\mathbf{B}, \mathbf{v}/v_A\to\mathbf{v},\\
&a\nabla\to\nabla, 4\pi a\mathbf{J}/(cB_P)\to\mathbf{J},\\
&p_e/(n_0T_e)\to p_e, n/n_0\to n, p/[n_0(T_e+T_i)]\to p,\\
&d_i/a\to d_i, d_e\to d_e/a,\\
&\beta_{ep}\equiv 4\pi n_0 T_e/{B_P}^2, \beta_p\equiv 4\pi n_0(T_e+T_i)/{B_P}^2.
\end{aligned}$$

In the definitions above, n_0 is the average ion and electron density (in a hydrogen plasma) and the ion temperature T_i is assumed to be constant. From the ion flow velocity $\mathbf{v}_i\approx\mathbf{v}$, we can compute the electron flow velocity by the relation

$$\mathbf{v}_e \approx \mathbf{v} - (d_i/n)\mathbf{J}. \tag{16}$$

For the initial equilibrium (1), y is an ignorable coordinate. We assume, furthermore, that $\partial/\partial y = 0$ for all times. Under these conditions, we can represent the magnetic field as

$$\mathbf{B} = \nabla\psi\times\hat{\mathbf{y}} + B_y\hat{\mathbf{y}}, B_y\equiv B_T + b(x,z,t), \tag{17}$$

where $b(x,z,t)$ is the perturbation to the B_y-field. From (17), it follows that

$$\mathbf{J} = \nabla b\times\hat{\mathbf{y}} - \nabla^2\psi\hat{\mathbf{y}}. \tag{18}$$

Note that the gradient operator in (17) and (18) is two-dimensional, that is,

$$\nabla = \hat{\mathbf{x}}\frac{\partial}{\partial x} + \hat{\mathbf{z}}\frac{\partial}{\partial z}. \tag{19}$$

In [43], we have given a derivation of the linear electromagnetic waves mediating collisionless reconnection by systematic reduction of the finite-beta Hall MHD

equations. It is helpful to determine the linear waves for two important reasons. First, the nature of the waves in the reconnection layer determine the collective motions of the electrons and ions in the reconnection layer, and the signature of the electric and magnetic field perturbations associated with these waves can possibly be tested by *in situ* observations. Second, the properties of the waves impose constraints on the accuracy and stability of the simulation codes used to study collisionless reconnection.

We summarize below our main results on linear waves in the reconnection layer. In the exterior or ideal region,$d_i \ll 1/k < 1$, the dominant waves are essentially shear-Alfvén, with the dispersion relation, $\omega \approx k_\| v_A$, where ω is the frequency. In the interior region or the reconnection layer, described by the inequality $d_e < 1/k < d_i$, when the equilibrium toroidal or guide field is zero (that is, $B_T = 0$), the dispersion relation is given by

$$\omega \approx k_\| k {d_i}^2 v_A = k_\| k {d_i}^2 \Omega_{ci} = k_\| k {d_e}^2 \Omega_{ce}. \tag{20}$$

The dispersion relation (20) is manifestly independent of the particle mass m_i or m_e. We suggest that it is appropriate to characterize the dispersion relation (20) as that of an Alfvén-whistler wave for the following reasons. An Alfvén wave lies in the low-frequency range $\omega \ll \Omega_{ci} \ll \Omega_{ce}$, and is non-dispersive because its frequency is linearly proportional to the wave number. In contrast, a whistler wave lies in the frequency range $\Omega_{ci} \ll \omega \ll \Omega_{ce}$, and is dispersive because its frequency has a quadratic dependence on the wave number. Since $kd_i \sim k_\perp d_i \sim 1$ in the reconnection layer, the magnitude of the wave frequency predicted by the dispersion relation (20) is essentially in the Alfvén range, but the wave is dispersive like the whistler.

In the interior region or the reconnection layer, when $\beta_p \leq 1$ and $B_T/B_P \geq 1$, the dispersion relation is given by,

$$\omega \approx {\beta_{ep}}^{1/2} k_\| k d_i v_A = k_\| k \varrho_s v_A = {\beta_{ep}}^{1/2} k_\| k {d_i}^2 \Omega_{ci} = {\beta_{ep}}^{1/2} k_\| k {d_e}^2 \Omega_{ce}, \tag{21}$$

where $\varrho_s \equiv {\beta_{ep}}^{1/2} d_i$. Equation (21) looks deceptively like the Alfvén-whistler wave dispersion relation (20) except for one crucial difference—the multiplicative factor ${\beta_{ep}}^{1/2}$ that is a clear signature of the electron pressure term in the generalized Ohm's law (2). This dispersion relation actually is an asymptotic form of the kinetic Alfvén wave dispersion relation in the limit $k^2 {\varrho_s}^2 >> 1 >> k^2 {d_e}^2$ [44].

In the regime $m_e/m_i < \beta_p \ll 1$, which is also the regime of very strong guide field $B_T >> 1$,we obtain the well-known dispersion relation for kinetic/inertial Alfvén waves/44/:

$$\left(\frac{\omega}{k_\|}\right)^2 \approx \frac{{B_{t0}}^2 + k^2 {\varrho_s}^2}{1 + k^2 {d_e}^2}. \tag{22}$$

This limit is relevant to fusion devices, and in particular, to the problem of sawtooth collapse in low-beta, weakly collisional tokamaks, discussed in Sect. 2. We caution the reader that the fluid equations in this paper are, strictly speaking, not valid in the regime $k^2 {\varrho_i}^2 >> 1$ when kinetic equations should

be used. However, for cold ions, one can have $k^2{\varrho_s}^2 >> 1 >> k^2{\varrho_i}^2$, since ${\varrho_s}^2/{\varrho_i}^2 = T_e/T_i >> 1$. In this regime, (22) continues to be a good approximation for the kinetic Alfvén wave dispersion relation.

4.2 Nonlinear Estimates of the Reconnection Rate

When the equilibrium guide-field is zero, the Hall current term controls the reconnection rate. Then the (dimensionless) generalized Ohm's law yields

$$\frac{\partial \psi}{\partial t} \approx d_i(\mathbf{J} \times \mathbf{B})_y \sim \frac{d_i}{\ell} b_s B_R, \tag{23}$$

where b_s is the nonlinearly saturated b-field and B_R is the magnitude of the poloidal field at the outer limit of the reconnection layer, assumed to be of half-length ℓ in the x-direction and half-width $\triangle$ in the z-direction. Reconnection is induced in this configuration by imposing inward flows at the boundaries on two sides of the neutral line at $z = 0$ [43]. Under steady-state conditions, we can estimate b_s from the relation

$$\frac{db}{dt} = \mathbf{v} \cdot \nabla b \approx -\frac{d_i}{n}\Big[\psi, \nabla^2\psi\Big] = \frac{d_i}{n}\Big[\psi, J_y\Big]. \tag{24}$$

Representing the inflow into the reconnection layer by V_{in} and the outflow by V_{out}, (24) yields the estimate

$$\left(V_{in}\frac{b_s}{\triangle} + V_{out}\frac{b_s}{\ell}\right) \sim d_i B_x \frac{\partial J_y}{\partial x} \sim d_i B_R \frac{J_y}{\ell} \sim d_i \frac{{B_R}^2}{\triangle \ell}. \tag{25}$$

Making use of the continuity relation,

$$V_{in}\ell \approx V_{out}\triangle \approx B_R\triangle, \tag{26}$$

in (25), we obtain

$$b_s \sim \frac{d_i B_R}{2\triangle}. \tag{27}$$

From the inner limit of the exterior or ideal region, we obtain

$$\frac{\partial \psi}{\partial t} \approx V_{in} B_R. \tag{28}$$

Matching (23) and (28), we obtain

$$V_{in} \sim \frac{d_i}{\ell} b_s. \tag{29}$$

From (26), (27), and (29), we obtain the reconnection layer width

$$\triangle \sim d_i/\sqrt{2}, \tag{30}$$

and the reconnection rate

$$V_{in} \sim \frac{d_i B_R}{\sqrt{2\ell}}. \tag{31}$$

We point out that the parameters B_R as well as ℓ depend on the boundary conditions of the system. The length ℓ of the reconnection layer can be estimated as follows. If reconnection is forced by a sinusoidal perturbation of wavenumber k, the magnetic flux function can be written $\psi(z,x) = \psi_0(z) + \tilde{\psi}(z)\cos kx$, where $\psi_0(z) = z^2/2$ is the equilibrium flux near the neutral line. In the region close to the X-point ($x, z << 1$), we can write

$$\psi(z,x) \approx \frac{1}{2}z^2 + \tilde{\psi}(0)\left(1 - \frac{k^2x^2}{2}\right) = \frac{1}{2}z^2 + \frac{{w_s}^2}{2}\left(1 - \frac{k^2x^2}{2}\right), \tag{32}$$

where w_s is the island half-width and related to $\tilde{\psi}$ by the relation $w_s \approx (2\tilde{\psi}(0))^{1/2}$. From (28), the reconnection electric field at the inner limit of the outer (or ideal region) is given by

$$\frac{\partial \psi}{\partial t} \approx v_{in} B_R \approx \frac{\triangle}{\ell} v_{out} B_R \approx \frac{\triangle}{\ell} {B_R}^2, \tag{33}$$

where B_R is the magnitude of the poloidal field at the inner (outer) limit of the ideal (reconnection) region and $v_{out} \approx v_A = B_R$ (in dimensionless variables). Downstream of the reconnection layer (along x), we have

$$\frac{\partial \psi}{\partial t} \approx -v_x \frac{\partial \psi}{\partial x}\bigg|_{x \approx \ell} \sim \frac{1}{4}k^2 {w_s}^2 \ell B_R. \tag{34}$$

Matching (33) and (34), we obtain

$$\ell \approx \frac{2\sqrt{B_R \triangle}}{k w_s}. \tag{35}$$

As shown by (30), $\triangle \propto d_i$ which implies that $\ell \propto {d_i}^{1/2}$. The quantities B_R and w_s can be calculated once the boundary conditions are specified.

In the presence of a significant guide field ($B_T \geq 1$), the ion dynamics in the inner or reconnection region is dominated by the electron pressure gradient term in the parallel component of (the dimensionless) Ohm's law. It can be shown that [43]

$$V_{in} \sim \frac{d_i {B_R}^2}{2\ell B_T}. \tag{36}$$

Both expressions (35) and (36) are closely related to analogous relations derived for the sawtooth problem, discussed in Sect. 2. For instance, as shown in [7], the length of the thin current sheet due to the nonlinear collisionless $m = 1$ instability is given by $\ell \approx r_s(\omega_A/\Omega_{ci})^{1/2}$ which is proportional to ${d_i}^{1/2}$, and so is the reconnection rate.

4.3 Comparison with Numerical Simulations

We now illustrate some of the results obtained in Sect. 4.2 with the Hall MHD code developed at the University of Iowa [16]. The equilibrium (1) is driven by inward flows of the form

$$\begin{aligned}\mathbf{v}(x, z = \pm a, t) &= \mp\hat{\mathbf{z}} V(t)(1 + \cos kx) \\ &= \mp\hat{\mathbf{z}} V_0 \tanh\Big(\frac{t}{\tau}\Big)\left(1 - \tanh\Big(\frac{t - t_0}{\tau}\Big)\right)(1 + \cos kx). \end{aligned} \quad (37)$$

The imposition of inward flows for a finite time (of the order t_0) and their subsequent switch-off ensures the realization of a quasi-saturated state that eventually decays in time. Periodic boundary conditions are imposed along x. It is easy to show that [45]

$$B_R \approx 2\left(1 + \frac{k}{\sinh k}\right) \int_0^\infty V(t)dt. \quad (38)$$

Equation (38) makes it clear that B_R, which is the magnitude of the poloidal field at the outer limit of the reconnection layer, is determined by the form of the boundary conditions. In what follows, we choose the dimensionless values $V_0 = 0.005$,$\tau = 2$, $k = \pi/2$, and $t_0 = 8$. For these parameters, we obtain $B_R \approx 0.24$.

We run the Hall MHD simulations with the parameters $d_i = 0.1$ and $S \approx 7 \times 10^3$. The system size along x and z is given by $L_x = 40d_i$ and $L_z = 20d_i$. In the figures that follow, unless stated otherwise, we plot only a quarter of the

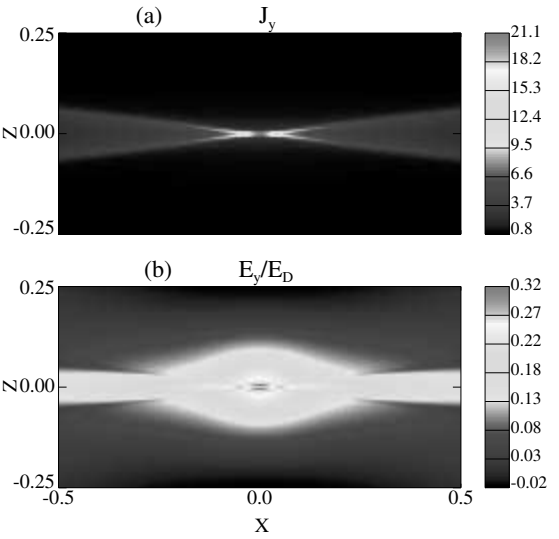


Fig. 6. Spatial scale separation between J_y and E_y when $B_T = 0$ (a) Thin current sheet with an X-point structure. (b) E_y, scaled by the Dreicer field E_D.

system size along each direction. Figure 6 shows the geometry of the thin current sheet in the quasi-saturated regime when $B_T = 0$. Although the geometry of the thin current sheet resembles an X-point structure of the type envisioned by Petschek, there are significant qualitative differences in the underlying physics. We note that there is a clear separation of spatial scales between E_y and J_y (For reference, E_y is scaled to the (dimensionless) Dreicer field $E_D = u_{Te}/(Sd_i)$ with $u_{Te} = 10V_A$.) The spatial scale of E_y is the ion skin depth, but the spatial scale of J_y is narrower and controlled by the Lundquist number.

The separation of scales between E_y and J_y is made possible by the decoupling of ion and electron flows in Hall MHD that is beyond the scope of the resistive MHD model used by Petschek. The ion flow can be calculated by setting $\mathbf{v}_i \approx \mathbf{v}$. In the ideal (or outer) region, $\mid z \mid \gg d_i$, the electrons and ions move together and $\mathbf{v}_i \approx \mathbf{v}_e$. The non-ideal (or inner region) is composed of two asymptotic regions: a region of characteristic width $\triangle_\eta$ where the thin current sheet forms, and a broader region of characteristic width $\triangle$. In the region $\triangle_\eta \ll \mid z \mid < \triangle$, the electrons and ions are decoupled and $\mid \mathbf{v}_i \mid \ll \mid \mathbf{v}_e \mid$. In the innermost asymptotic region $\mid z \mid \leq \triangle_\eta$, around the X-point, $\mathbf{v}_i \approx 0$ and the current is carried almost entirely by electrons.

There is an important point of distinction between collisionless reconnection phenomena with and without a guide field. When the guide field is zero in the initial state, a finite guide (B_y) field with quadrupolar symmetry [46, 47] is generated spontaneously by the Hall current. In the presence of a significant guide field ($B_T = 1$), the quadrupolar symmetry is destroyed. For a more detailed discussion, the reader is referred to [43].

Figure 7 shows a test of the dependence of the reconnection rate (31) on the ion skin depth d_i, all other parameters remaining fixed for the case $B_T = 0$. The plot demonstrates that the maximum reconnection electric field defined by $E_{max} \equiv V_{in} B_R$ is proportional to $\sqrt{d_i}$. For a more detailed discussion of the structure and properties of the reconnection layer, the reader is referred to [48].

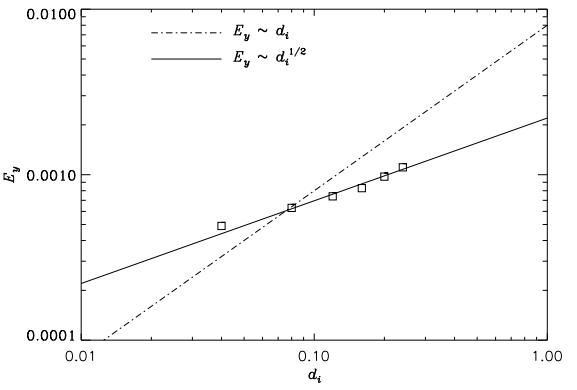


Fig. 7. Plot of the maximum reconnection electric field as a function of the ion skin depth d_i.

5 Fast Trigger for Flare Dynamics in the Solar Corona

A frequently observed feature of a solar flare is its sudden growth from a relatively quiescent background. (See, for instance, the monograph by Tandberg-Hanssen and Emslie [49], and other references therein.) The impulsive phase is characterized by a sudden release of the magnetic energy stored during the pre-onset phase. Recent time-resolved measurements with devices such as the Yohkoh soft X-ray telescope have allowed unprecedented opportunities for continuous observations of the pre-onset and onset phases of a flare as well as the recovery phase after a flare. Based on these observations, Tsuneta and co-workers [50,51] have presented several pieces of strongly suggestive evidence that magnetic reconnection is responsible for the energy release.

Numerical simulations of flares are generally based on resistive MHD equations, employing either constant or anomalously enhanced resistivity [52, 53, 54]. While these simulations attempt to incorporate many of the physical effects relevant to a flare, they do not appear to address the issue of the trigger which is the main concern of the present study. We repeat for emphasis that observations of the impulsive phase impose a significant constraint on magnetic reconnection models; not only should the time scale be fast, but the time - derivative of the growth rate curve must increase suddenly. The traditional steady-state reconnection models of Sweet-Parker and Petschek can, at best, provide a fast time scale, but since they do not include time-dependency, they cannot account for the time-evolution of the growth rate.

Magnetic reconnection is involved in the temporal development of near-singular current density in spatially localized structures regions in a coronal plasma. Parker [55] proposed that such intense concentrations of current density ("current sheets"), which can grow from a smooth initial state (such as a vacuum magnetic field) due to shearing motions of the photospheric footpoints, can be an effective mechanism for coronal heating. Small events that produce current sheets and release energy ($\sim 10^{17} J$) are classified as nanoflares. Current sheets are not only a defining feature of nanoflares, but are also expected to occur in the high-Lundquist-number corona during the development of microflares and flares if magnetic reconnection is the underlying mechanism liberating large amounts of energy ($\sim 10^{20} J$ for microflares and $\sim 10^{26} J$ for flares). Hence, the time-development of current sheets and the associated dc electric field at the separatrix must be characterized by rapid and impulsive growth if magnetic reconnection is the physical mechanism underlying a flare.

The classical Lundquist number of the solar corona is very high. If we consider a pre-flare coronal loop with length $L \sim 10^9 cm$, magnetic field strength $B \sim 300$ G, density $n \sim 10^{10} cm^{-3}$, and temperature $T \sim 2 \times 10^6$ K [49], the classical Lundquist number is given by $S = \tau_R/\tau_A = (4\pi L^2/\eta c^2)/(L/V_A) \sim 10^{14}$. Since the Ohmic energy dissipation at such high values of the Lundquist number is too low, Spicer [56] has considered a number of ways in which anomalous resistivity due to microinstabilities of the thin current sheets at multiple sites of reconnection can provide a mechanism for enhancing the dissipation rate. In such a scenario, it is argued plausibly that the microinstabilities are triggered when

the current sheets become thin enough, and the anomalous resistivity thus produced then broadens the thin current sheet to evolve it into a state of marginal stability. In the process, energy dissipation due to anomalous resistivity can be enhanced one to two orders of magnitude over the classical value. If the anomalous resistivity is turned on suddenly at the onset of the relevant microinstability and turns off when the instability becomes marginally stable, it can provide an explanation of the trigger if the resistivity is large enough to match the required reconnection rate. We do not question the plausibility of such a scenario, but explore the possibility that collisionless effects outside the scope of the simple Ohm's law $\mathbf{E} + \mathbf{v} \times \mathbf{B}/\mathbf{c} = \eta \mathbf{J}$ can play an important role and provide a viable mechanism for the trigger without the need to invoke microinstabilities.

To fix ideas, we consider a two-dimensional coronal arcade with an X-type neutral line where current sheets can be realized by imposing footpoint displacements [57-64]. Since it is clear where current sheets (and reconnection) occur in this geometry, an important question is how rapidly in time such sheets develop from a given smooth initial state. This question has been investigated in some detail by Wang and Bhattacharjee [64] and Ma et al. [5]. The initial state used by Ma et al.[5] is similar to that shown in Fig. 1: it describes a coronal configuration in the $x - z$ plane, with the footpoints of the magnetic field intersecting the photosphere which is taken to be a plane perpendicular to the plane of the paper. The initial force-free equilibrium containing an X-type neutral line is obtained by solving Laplace's equation for a potential field in the whole physical domain as a boundary-value problem. We assume the resistivity to be constant, corresponding to a Lundquist number of approximately 10^5. At the bottom boundary which represents the photosphere, we impose shearing motion of the footpoints according to the relation,

$$v_y(x, y = 0, t) = v_0 \tanh(\frac{t}{t_0}) \exp\Big[- (\frac{x - x_0}{\triangle x})^2 \Big], \tag{39}$$

where t_0 is a rapid "turn-on" time, $x = x_0$ is the point at which the outer separatrix intersects the x axis in the simulation box (which is half of the physical domain), and $\triangle x$ characterizes the spatial extent of the sheared profile.

In Figs. 8a and 8b, the dotted line indicates the time-evolution of the maximum amplitude of J_y at the separatrix and the maximum parallel electric field in the resistive simulation with a localized footpoint shear profile ($\triangle x = 0.16a$). We note that the growth of current sheets as well as the parallel electric field in this case has no impulsive phase.

With the same initial and boundary conditions, we now present simulation results from the compressible Hall MHD equations (Figs. 8a and 8b, solid lines). As mentioned above, the Hall terms come into play when the thickness of the current sheet becomes smaller than the ion skin depth. Prior to this time (which includes the linear phase), magnetic reconnection is controlled by the resistivity. Figures 8a and 8b show, respectively, the time-dependence of the maximum current sheet amplitude as well as the parallel electric field for the Hall MHD run with $d_i = 0.1$. While the peak amplitude of the Hall MHD current sheet is somewhat smaller than the resistive MHD current sheet (Fig. 8a), the former shows

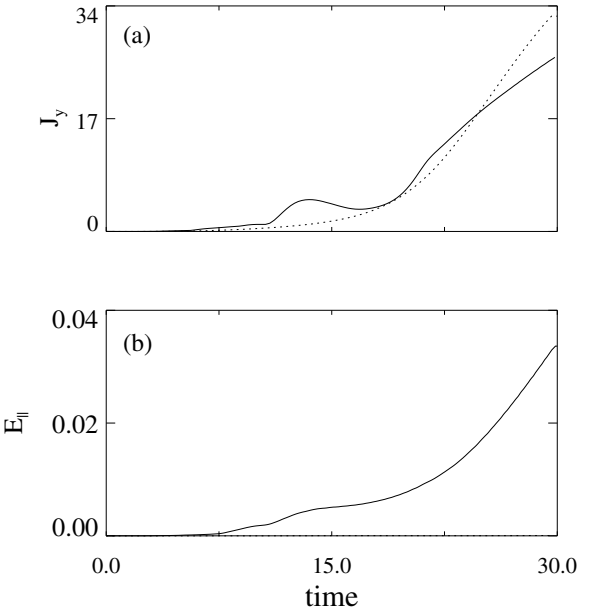


Fig. 8. (a). Maximum amplitudes J_y and (b) $E_\parallel$ for resistive MHD (dotted line) and Hall MHD (solid line) simulations as a function of time.

a much greater degree of impulsiveness. We remark that the non- monotonic time-evolution seen in the Hall MHD run can be explained as the outcome of two competing tendencies. On the one hand, the photospheric footpoint motion tends to build up the flux ψ near the separatrix which reconnects impulsively under the influence of Hall MHD effects. On the other hand, due to the build-up of the flux ψ the loop tends to expand outward, and this outward motion tends to reduce the flux and the reconnection rate near the separatrix.

Figure 9 shows image plots of J_y, E_y, and $E_\parallel$ for the resistive and Hall MHD runs at $t = 30$. Two features stand out. First, the Hall MHD simulation shows a significant change in the geometry of the reconnection region by reducing the length of the Y-type current sheet seen in the resistive MHD simulation. This qualitative change in the current sheet structure increases $E_\parallel$ and the magnetic reconnection rate which is strongly influenced by the length-to-width ratio of the current sheet. Second, there is a separation of spatial scales between E_y and J_y in the Hall MHD simulation, absent in the resistive MHD simulation. Whereas the scale of E_y is of the order of d_i and independent of the resistivity, the scale of J_y is much narrower and controlled by the resistivity. This decoupling of spatial scales in Hall MHD (which also occurs between $E_\parallel$ and $J_\parallel$) is due to the decoupling between ion and electron motions. As a consequence, we obtain impulsive as well as fast (near-Alfveénic) reconnection rates that are controlled by ions and are weakly dependent on the magnitude of the Lundquist number.

The initial geometry used in this simulation is simple, and we do not claim that we have simulated an actual flare. Any model of magnetic reconnection that

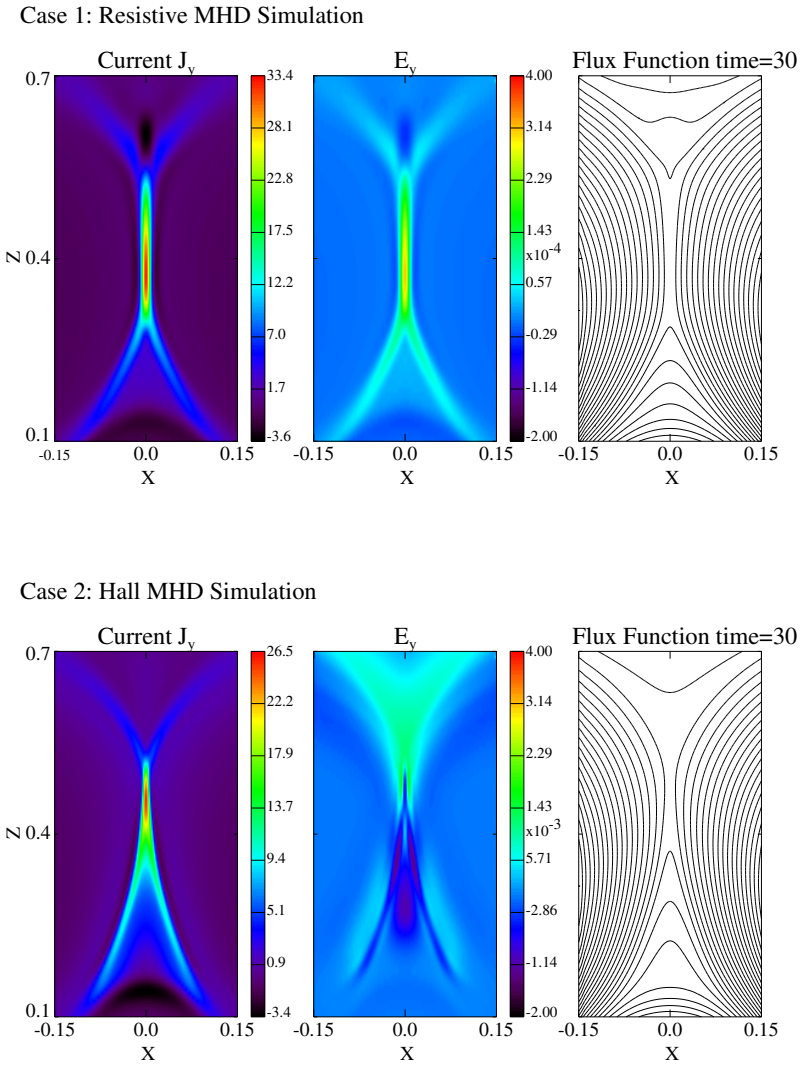


Fig. 9. Image plots of J_y, E_y, and $E_{\|}$ for resistive and Hall MHD simulations at $t = 30$ with the same initial conditions.

proposes to account for flare dynamics must show a convincing way to account for the energy liberated as well as the large number of energetic electrons produced, and this remains to be done. We have focused instead on the pivotal issues of fast time scale and trigger, and demonstrated that reconnection models based on Hall MHD can potentially account for these issues. A more complete model of impulsive flares that relies on magnetic reconnection as an underlying mechanism is left to future work, where the physical considerations discussed in this paper will be essential ingredients.

6 Summary

In this paper, we have reviewed recent developments in nonlinear collisionless reconnection theory involving impulsive (or triggered) as well as quasi-steady dynamics. For a discussion of impulsive reconnection, we have drawn upon three examples: sawtooth collapse in tokamaks, substorms in the Earth's magnetotail, and impulsive flares in the solar corona. Both types of plasmas have very high Lundquist numbers ($S \geq 10^8$), and hence are classified as collisionless. We have suggested that at such high values of the Lundquist number, the standard Ohm's law of resistive MHD needs to be replaced by the generalized Ohm's law (2). We have shown by means of analysis and simulation that the incorporation of electron pressure gradient and Hall current effects brings theory closer to observations in the sense that it provides a fast reconnection time scale (near-Alfvénic) as well as a trigger. In those special cases, when the dynamics of a system happens to evolve into a quasi-steady state, we have obtained analytical expressions for the geometric characteristics (that is, length and width) of the reconnection layer and the reconnection rate. These analytical results have been tested by Hall MHD simulations.

The reconnection dynamics in the three examples considered in this paper have some common features. If we contrast reconnection dynamics in Hall MHD with that in resistive MHD, three distinguishing features emerge. First, for impulsive as well as quasi-steady reconnection dynamics, we obtain sub-Alfvénic reconnection rates that are insensitive to the value of resistivity. Second, the geometry of the reconnection region changes dramatically as the Y-points of the resistive MHD current sheet shrink rapidly to form a structure that approaches an X-point. Consequently, there is a rapid enhancement in the reconnection rate. The underlying physical mechanism that makes the first two qualitative changes possible is the decoupling of the spatial scale of the parallel electric field from the spatial scale of the parallel current density in a way that is forbidden in resistive MHD. Whereas the reconnection electric field has a spatial scale controlled by the ions, the current sheet has a narrower spatial scale, determined by the high value of S.

The results mentioned above prompt comparisons with the Petschek model, though we repeat for emphasis that the nonlinear dynamics in cases involving impulsive reconnection are non-steady and are essentially dominated by collisionless effects, and that both of these features are qualitatively beyond the scope of Petschek's original model. Whereas Petschek obtained reconnection rates that depend logarithmically on the resistivity, the present model obtains reconnection rates that are sub-Alfvénic and independent of the resistivity at leading order. Whereas Petschek obtained such high reconnection rates by assuming that the reconnection dynamics sustains an X-point geometry for all times, the present model evolves in the high-S regime from an a Y-point geometry in the early nonlinear stage, characteristic of the models of Sweet-Parker, to a geometry approaching an X-point.

While the Hall MHD model is a point of departure for studies of collisionless reconnection dynamics, it omits significant kinetic effects such as those associ-

ated with tensor pressure and microinstabilities of current sheets. These continue to be topics of extensive ongoing research.

Acknowledgements

This research is supported by NSF Grant No. ATM 00-0203, and the DOE Cooperative Agreement No. DE-FC02-01ER59651 under the auspices of the program on Scientific Discovery through Advanced Computing.

References

1. P. A. Sweet, Nuovo Cimento Suppl.: Ser. X. **8**, 188 (1958).
2. E. N. Parker, Astrophys. J. Suppl. Ser. **8**, 177 (1963).
3. H. E. Petschek, in *AAS/NASA Symposium on the Physics of Solar Flares*. ed. by W. N. Hess (NASA, Washington, DC, 1964), p.425.
4. S. I. Syrovatsky, Sov. Phys. JETP **33**, 933 (1971).
5. Z.W. Ma, C. S. Ng, X. Wang, and A. Bhattacharjee, Phys. Plasmas **2**, 3184 (1995).
6. A. Y. Aydemir, Phys. Fluids B **4**, 3469 (1992).
7. X. Wang and A. Bhattacharjee, Phys. Rev. Lett. **70**, 1627 (1993); Phys. Plasmas **2**, 171 (1995); Phys. Plasmas **4**, 1173 (1997).
8. R. G. Kleva, J. F. Drake, and F. L. Waelbroeck, Phys. Plasmas **2**, 23 (1995).
9. D. Biskamp, E. Schwarz, and J. F. Drake, Phys. Rev. Lett. **75**, 3850 (1995).
10. B. Rogers and L. Zakharov, Phys. Plasmas **3**, 2411 (1996).
11. R. Horiuchi and T. Sato, Phys. Plasmas **4**, 277 (1997).
12. D. Grasso, F. Pegoraro, F. Porcelli, and F. Califano, Plasma Phys. Control. Fusion **41**, 1497 (1999).
13. H. J. Cai, D. Q. Ding, and L. C. Lee, J. Geophys. Res. **99**, 35 (1994).
14. M. E. Mandt, R. E. Denton, and J. F. Drake, Geophys. Res. Lett. **21**, 73 (1994).
15. D. Krauss-Varban and N. Omidi, Geophys. Res. Lett. **22**, 3271 (1995).
16. Z. W. Ma and A. Bhattacharjee, Geophys. Res. Lett. **23**, 1673 (1996).
17. Y. Lin and D. W. Swift, J. Geophys. Res. **101**, 19859 (1996).
18. J. Büchner and J.-P. Kuska, Adv. Space Res. **19**, 1817 (1997).
19. M. A. Shay and J. F. Drake, Geophys. Res. Lett. **25**, 3759 (1998).
20. R.-F. Lottermoser, M. Scholer, and A. P. Mathews, J. Geophys. Res. **103**, 4547 (1998).
21. M. Hesse, J. Birn and M. Kuznetsova, Phys. Plasmas **6**, 1781 (1999).
22. J. A. Wesson, Nucl. Fusion **30**, 2545 (1990).
23. J. Birn, J. F. Drake, M. A. Shay, B. N. Rogers, R. E. Denton, M. Hesse, M. Kuznetsova, Z. W. Ma, A. Bhattacharjee, A. Otto, and P. L. Prtichett, J. Geophys. Res. **106**, 3715 (2000).
24. B. Kadomtsev, Sov. J. Plasma Phys. **1**, 389 (1975).
25. F. L. Waelbroeck, Phys. Plasmas **1**, 2372 (1989).
26. W. Park, D. A. Monticello, and R. B. White, Phys. Fluids **27**, 137 (1984).
27. D. Biskamp, Phys. Fluids B **3**, 3353 (1991).
28. T. H. Osborne, R. N. Dexter, and S. C. Prager, Phys. Rev. Lett. **49**, 734 (1982).
29. H. Soltwisch, H., Rev. Sci. Instrum. **59**, 1599 (1988).
30. H. G. Weisen, H., G. Borg, B. Joye, A. J. Knight, and J. B. Lister, Phys. Rev. Lett. **62**, 434 (1989).

31. R. L. Kaufmann, J. Geophys. Res. **92**, 7471 (1987).
32. V. A. Sergeev, V. A., P. Tanskenen, K. Mursula, A. North, and R. C. Elphic, J. Geophys. Res. **95**, 3819 (1990).
33. A. T. Y. Lui, A. T. Y., R. E. Lopez, B. J. Anderson, K. Takahashi, L. J. Zanetti, R. W. McEntire, T. A. Potemra, D. M. Klumpar, E. M. Greene, and R. Strangeway, J. Geophys. Res. 97, 1461 (1992).
34. S. K. Ohtani, S. Kokobun, and C. T. Russell, J. Geophys. Res. **97**, 3129 (1992).
35. S. K. Ohtani, S., K. Takahashi, L. J. Zanetti, T. A. Potemra, R. W. McEntire, and T. Iijima, J. Geophys. Res. **97**, 19,311 (1992).
36. Z. W. Ma and A. Bhattacharjee, Geophys. Res. Lett. **25**, 3277 (1998).
37. N. A. Tsyganenko, Planet. Space Sci. **37**, 5 (1989).
38. T. Nagai, M. Fujimoto, Y. Saito, S. Machida, T. Terasawa, R. Nakamura, T. Yamamoto, T. Mukai, A. Nishida, and S. Kokubun, J. Geophys. Res. **103**, 4419 (1998).
39. E. W. Hones, Jr., Space Sci. Rev. **23**, 393 (1979).
40. Z. W. Ma, X. Wang and A. Bhattacharjee, Geophys. Res. Lett. **22**, 2985 (1995).
41. A. Bhattacharjee, Z. W. Ma and X. Wang, Phys. Plasmas **5**, 2001 (1998).
42. O. Hurricane, B. H. Fong, and S. C. Cowley, Phys. Plasmas **4**, 3565 (1997).
43. X. Wang, A. Bhattacharjee and Z. W. Ma, J. Geophys. Res. **105**, 27,633 (2000).
44. A. Hasegawa and C. Uberoi, The Alfvén Wave (U. S. Department of Energy, Washington, DC, 1982), p. 21.
45. X. Wang, Z. W. Ma, and A. Bhattacharjee, Phys. Plasmas **3**, 2129 (1996).
46. B. U. O. Sonnerup In: *Solar System Plasma Physics*. eds. L. J. Lanzerotti, C. F. Kennel, and E. N. Parker(North Holland Publishers, Amsterdam, 1979) p. 3.
47. T. Terasawa, Geophys. Res. Lett. **10**, 475 (1983).
48. X. Wang, A. Bhattacharjee, and Z.W. Ma, Phys. Rev. Lett. **87**, 265003 (2001).
49. E. Tandberg-Hanssen and A.G. Emslie: *The Physics of Solar Flares* (Cambridge University Press, New York 1988)
50. S. Tsuneta, Astrophys. J. **456**, 840 (1996).
51. S. Tsuneta, Astrophys. J. **478**, 787 (1997).
52. T.G Forbes and J.M. Malherbe, Sol. Phys. **135**, 361 (1991).
53. D. Biskamp, *Nonlinear Magnetohydrodynamics* (Cambridge University Press, New York 1993).
54. T. Yokoyama and K. Shibata, Astrophys. J. **474**, L61 (1997).
55. E. N. Parker, Astrophys. J. **174**, 499 (1972).
56. D.S. Spicer, Space Sci. Rev. **31**, 351 (1982).
57. B.C. Low and R. Wolfson, Astrophys. J. **324**, 574 (1988).
58. Z. Mikic, D.C. Barnes, and D.D. Schnack, Astrophys. J. **328**, 830 (1988).
59. T.H. Jensen, Astrophys. J. **343**, 507 (1989).
60. D. Biskamp and H. Welter, Sol. Phys. **120**, 49 (1989).
61. J.T. Karpen, S.K. Antiochos, and C.R. Devore, Astrophys. J. **356**, L67 (1990).
62. J.M. Finn and Y.T. Lau, Phys. Fluids B **3**, 2675 (1991).
63. G. Vekstein, E. Priest, and T. Amari, Astron. Astrophys. **243**, 492 (1991).
64. X. Wang and A. Bhattacharjee, Astrophys. J. **420**, 415 (1994).

Problems and Progress in Astrophysical Dynamos

Ethan T. Vishniac[1], Alex Lazarian[2], and Jungyeon Cho[2]

[1] Johns Hopkins Univ., Baltimore MD21218, USA
[2] Univ. of Wisconsin, Madison WI53706, USA

Abstract. Astrophysical objects with negligible resistivity are often threaded by large scale magnetic fields. The generation of these fields is somewhat mysterious, since a magnetic field in a perfectly conducting fluid cannot change the flux threading a fluid element, or the field topology. Classical dynamo theory evades this limit by assuming that magnetic reconnection is fast, even for vanishing resistivity, and that the large-scale field can be generated by the action of kinetic helicity. Both these claims have been severely criticized, and the latter appears to conflict with strong theoretical arguments based on magnetic helicity conservation and a series of numerical simulations. Here we discuss recent efforts to explain fast magnetic reconnection through the topological effects of a weak stochastic magnetic field component. We also show how mean-field dynamo theory can be recast in a form which respects magnetic helicity conservation, and how this changes our understanding of astrophysical dynamos. Finally, we comment briefly on why an asymmetry between small-scale magnetic and velocity fields is necessary for dynamo action, and how it can arise naturally.

1 Introduction

Magnetic fields have played a curious role in astrophysics, being both commonplace and poorly understood. They are ubiquitous in ionized systems, from the interiors of stars to the hot interstellar medium. The magnetic energy density is typically roughly comparable to the turbulent kinetic energy density. In stellar interiors, this means that magnetic fields tend to play a small role. In the interstellar medium, and in stellar coronae, their role is large, and consequently a matter of intense debate. In accretion disks the typical magnetic field energy density is probably an order of magnitude below the ambient gas pressure (e.g. [39,40,73,14,15]) but they play a critical role in the outward transfer of angular momentum and the dissipation of orbital energy. Moreover, in optically thin environments the presence of a strong magnetic field can have a dramatic effect on the luminosity and spectrum of an object. A clear understanding of the generation and dynamics of magnetic fields is important to astrophysics in many ways. Unfortunately, their dynamics has not been well understood, at least judging by the diversity of opinions found in the literature [20,66,50,80,16]. Consequently, arguments which cite magnetic fields as a dynamically important element in any particular object have tended to rely on phenomenology, rather than any sort of fundamental explanation.

Fortunately, over the last ten years, and especially quite recently, there has been significant progress in this area. First, although direct observations of high

E. Falgarone and T. Passot (Eds.): LNP 614, pp. 376–401, 2003.
© Springer-Verlag Berlin Heidelberg 2003

conductivity magnetic field dynamics are still restricted to the solar wind and the Sun, improvements in resolution have made it possible to watch magnetic fields evolve in real time [43], and to measure the power spectrum of magnetohydrodynamic (MHD) turbulence in the solar wind directly [57]. Second, numerical simulations have reached the point where it is possible to simulate simple MHD systems with $\sim 10^8$ cells over many dynamical times. Third, a better understanding has been reached in terms of MHD turbulence theory (for a review see the chapter by Cho, Lazarian & Vishniac in this volume).

These results encourage us to believe that the many remaining problems are ripe for further progress. These problems range from the nature of dynamo processes in stars, accretion disks, and galaxies, to the question of how magnetic fields reconnect on dynamical time scales, with apparent disregard for the constraint due to flux-freezing. To be more precise, in the limit of negligible resistivity, the magnetic field in a fluid medium is frozen, the sense that neither the magnetic flux threading a fluid element, nor the field topology, can change. Magnetic reconnection, the exchange of partners between adjacent field lines violates the second condition, while the generation of a large-scale field through dynamo action apparently violates the first.

Conventional *mean field dynamo* theory (see [62,65,49] for reviews) allows a large-scale magnetic field to grow exponentially from a seed field (see [69,51]) at the expense of small-scale turbulent energy through a process of spiral twisting and reconnection, illustrated in Fig. 1. This process starts with a set of large-scale parallel field lines pointing in some arbitrary direction. If the underlying turbulence has a tendency to twist the field lines into spirals with a preferred handedness (i.e. the velocity field has some net helicity), then reconnection on two dimensional surfaces between adjacent spirals will produce a new field, at right angles to the old one, provided that there is a systematic gradient in the strength of the spirals. The new field component is at right angles to both the gradient and the old field component. In a differentially rotating system, we can get a dynamo if the original field direction is in the $\hat{\phi}$ direction, and the dynamo process produces a radial field component. Differential shearing of B_r will then drive the azimuthal field component, closing the cycle. This is the '$\alpha - \Omega$ dynamo'. In the absence of global shear, we need a second round of dynamo action, which gives an α^2 dynamo. This process can be given a systematic mathematical treatment by a suitable choice of averaging procedures.

There are two ways in which this picture ignores, rather than solves, the difficulties imposed by flux-freezing. The more obvious point is that adjacent spiral field lines are assumed to reconnect quickly. Without this assumption the field will accumulate small-scale tangled knots which will quickly suppress dynamo action, and the large-scale magnetic field will saturate far below equipartition with the surrounding turbulence. Unfortunately, if reconnection happens at the rate allowed by the generally accepted Sweet-Parker model [64,78], it is far too slow. However, there are observations which suggest that this represents more of a challenge for theorists than a real constraint on the evolution of magnetic fields. If reconnection is slow, turbulence would cause many magnetic reversals

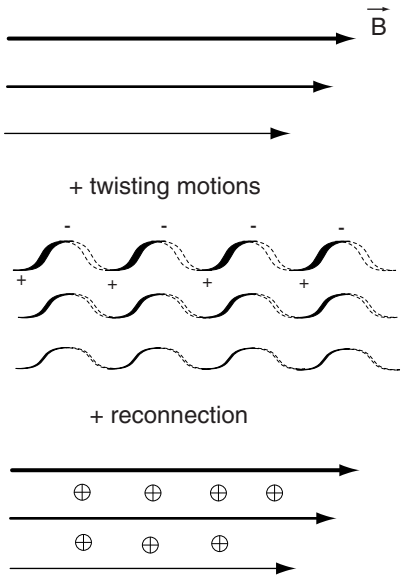


Fig. 1. The mean-field dynamo in action. Anisotropic turbulence twists the field lines into spirals. Reconnection restores the original field lines, but a vertical gradient in the strength of the spirals generates a net flux out of the page.

per parsec within the interstellar medium. Observations, on the contrary, show that magnetic field is coherent over the scales of hundreds of parsecs. This fact, as well as direct observations of large and small-scale Solar flares [27], suggest that the rate of reconnection is many orders of magnitude more rapid than allowed by the Sweet-Parker model. As this example shows, the importance of reconnection in astrophysics is not limited to understanding the dynamo process. The process of reconnection is an integral part of the transfer of magnetic energy to fluid and particle motion in stellar coronae and in the interstellar medium. More generally, it is impossible to claim that we understand MHD unless we can predict whether crossing magnetic flux tubes will reconnect or bounce from one another.

A more subtle difficulty arises from the process by which straight field lines are twisted into spirals. This is intuitively appealing if we consider field lines as isolated strings of infinitesimal radius. More realistically, the field occupies a non-zero volume. Twisting a tube into a spiral shape requires that we either allow the ends to slip, or allow parts of the tube to twist in the opposite sense. There is a geometrical constraint which is ignored in the standard picture. This objection can be given a rigorous mathematical form, the conservation of magnetic helicity, which we will describe in Sect. 3.

How do numerical simulations of dynamo activity compare to mean-field dynamo theory? Computer simulations of dynamos can be divided into two

classes. There are simulations in which some local instability (convection, the Balbus-Hawley instability etc.) is allowed to operate, and there are simulations in which the turbulence is driven externally, usually in such a way as to guarantee the presence of a net fluid helicity. The former simulations are often successful at generating large-scale magnetic fields whose energy density is at least as great as the turbulent energy density (e.g.[14,37,32]). The latter are less successful, in the sense that the energy density of the large-scale magnetic field is often quite modest (e.g. [60,4]). In particular there are simulations ([19,13]) which produce dynamos in a computational box, with forced helical turbulence. These dynamos show a steep inverse correlation between the dynamo growth rate and the conductivity. Naively extrapolating to astrophysical regimes suggests that magnetic dynamos driven by fluid helicity would take enormous amounts of time to grow. This conclusion is sharply at odds with evidence for rapid and efficient stellar dynamos.

Here we discuss recent work on the problems of fast reconnection and magnetic helicity conservation in astrophysical dynamos. For reconnection we concentrate on a generic reconnection scheme that appeals to magnetic field stochasticity as the critical property that accelerates reconnection ([53], see [54] for a review). Collisionless plasma effects which may also accelerate magnetic reconnection are addressed in the chapter by Bhattacharjee in this volume. We will typically assume that the evolution of the magnetic field is described by the simplest form of the induction equation

$$\partial_t \mathbf{B} = \nabla \times (\mathbf{v} \times \mathbf{B}) + \eta \nabla^2 \mathbf{B}, \tag{1}$$

although we will make reference to work which includes more realistic treatments of collisionless plasma effects.

2 Rates of Magnetic Reconnection

A simple dimensionless measure of the importance of resistivity, η, in a conducting fluid is the Lundquist number $\equiv V_A L/\eta$, where $V_A \equiv B/(4\pi\rho)^{1/2}$ is the Alfvén velocity and L is a typical scale of the system. When fluid velocities are of order the Alfvén speed, as is usual in astrophysics, this is a crude estimate of the ratio of the first and second terms in (1). Typically this number is very large under most astrophysical circumstances, and flux freezing should be a good approximation. More precisely, the coefficient of magnetic field diffusivity in a fully ionized plasma is $\eta = c^2/(4\pi\sigma) = 10^{13} T^{-3/2}$ cm^2 s^{-1}, where $\sigma = 10^7 T^{3/2}$ s^{-1} is the plasma conductivity and T is electron temperature. The characteristic time for field diffusion through a plasma slab of size L is L^2/η, which is large for any "astrophysical" L.

What happens when magnetic field lines intersect? Do they deform each other and bounce back or do they change their topology? This is the central question of the theory of magnetic reconnection. In fact, the whole dynamics of magnetized fluids and the back-reaction of the magnetic field depends on the answer.

2.1 The Sweet-Parker Scheme and Its Modifications

The literature on magnetic reconnection is rich and vast (see, for example, [68] and references therein). We start by discussing a robust scheme proposed by Sweet and Parker [64,78]. In this scheme oppositely directed magnetic fields are brought into contact over a region of length L_x (see Fig. 2). In general there will be a shared component, of the same order as the reversed component. However, this has only a minor effect on our discussion. The gradient in the magnetic field is confined to the current sheet, a region of vertical size Δ, within which the magnetic field evolves resistively. The velocity of reconnection, V_r, is the speed with which magnetic field lines enter the current sheet, and is roughly $\eta \approx V_r \Delta$. Arbitrarily high values of V_r can be achieved (transiently) by decreasing Δ. However, for sustained reconnection there is an additional constraint imposed by mass conservation. The plasma initially entrained on the magnetic field lines must escape from the reconnection zone. In the Sweet-Parker scheme this means a bulk outflow, parallel to the field lines, within the current sheet. Since the mass enters along a zone of width L_x, and is ejected within a zone of width Δ, this implies

$$\rho V_{rec} L_x = \rho' V_A \Delta \quad , \tag{2}$$

where we have assumed that the outflow occurs at the Alfvén velocity. This is actually an upper limit set by energy conservation. If we ignore the effects of compressibility, $\rho = \rho'$ and the resulting reconnection velocity allowed by Ohmic diffusivity and the mass constraint is

$$V_{rec,sweet-parker} \approx V_A \mathcal{R}_L^{-1/2}, \tag{3}$$

where $\mathcal{R}_L$ is the Lundquist number using the current sheet *length*. Depending on the specific astrophysical context, this gives a reconnection speed which lies somewhere between 10^{-3} (stars) and 10^{-10} (the galaxy) times V_A.

It is well known that using the Sweet-Parker reconnection rate it is impossible to explain solar flares. For the reasons given in the introduction, it is also well known that it is impossible to reconcile dynamo theory with observations without some substantially faster reconnection scheme. Consequently, for forty years discussions of reconnection speeds have tended to focus on mechanisms that might give reconnection speeds close to V_A, i.e. 'fast' reconnection. In general, we can divide schemes for fast reconnection into those which alter the microscopic resistivity, broadening the current sheet, and those which change the global geometry, thereby reducing L_x. Ultimately, a successful scheme should satisfy basic physical constraints without requiring contrived geometries or boundary conditions. In the near term, we can gain some insight into the likely nature of the solution by considering that reconnection is not always fast. Magnetic field lines in the solar corona and chromosphere which could reach a lower energy configuration through reconnection do not always immediately do so. Furthermore, a solution which relies entirely on collisionless effects, for example, would imply that field lines do not reconnect in dense environments, which would leave a major problem in understanding the nature of stellar dynamos.

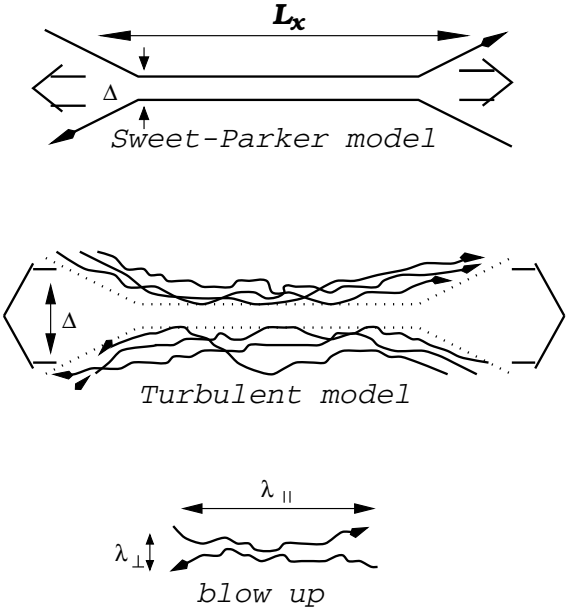


Fig. 2. Upper plot: Sweet-Parker scheme of reconnection. Middle plot: illustration of stochastic reconnection that accounts for field line noise. Lower plot: a close-up of the contact region. Thick arrows depict outflows of plasma. From [54].

Attempts to accelerate Sweet-Parker reconnection are numerous. We start by considering schemes to broaden the current sheet. Anomalous resistivity is known to broaden current sheets in laboratory plasmas. It is present in the reconnection layer when the field gradient is so sharp that the electron drift velocity is of the order of the thermal velocity of ions $u = (kT/m)^{1/2}$[65]. In other words, when $j > j_{cr} = Neu$. If the current sheet has a width δ with a change in the magnetic field ΔB then $4\pi j = c\Delta B/\delta$. The effective resistivity increases nonlinearly as j becomes greater than j_{cr}, thereby broadening the current sheet. We can find an upper limit to this effect by assuming that j never gets very much larger than j_{cr}, that is $\delta \approx \frac{c\Delta B}{4\pi Neu}$. Expressing δ in terms of the ion cyclotron radius $r_c = (muc)/(eB_{tot})$, where B_{tot} is the total magnetic field (including any shared component) we find

$$\delta \approx r_c \left(\frac{V_A}{u}\right)^2 \frac{\Delta B}{B_{tot}} \quad , \tag{4}$$

which agrees with [65] up to the factor $\Delta B/B_{tot}$, which equals 1 in that treatment. Combining (2) and (4) one gets [53]

$$V_{rec,anomalous} \approx V_A \frac{r_c}{L_x} \left(\frac{V_A}{u}\right)^2 \frac{\Delta B}{B_{tot}} \quad . \tag{5}$$

Equation (5) shows that the enhanced reconnection velocity is still much less than the Alfvén velocity if L_x is much greater than the ion Larmor (cyclotron)

radius. In general, "anomalous reconnection" is important when the thickness of the reconnection layer in the Sweet-Parker reconnection scheme is less than δ. However, for typical interstellar magnetic fields the Larmor radius r_c is $\sim 10^7$ cm and anomalous effects are negligible.

Tearing modes are a robust instability connected to the appearance of narrow current sheets [31]. The resulting turbulence will broaden the reconnection layer and enhance the reconnection speed. Here we give an estimate of this effect and show that while it represents a significant enhancement of Sweet-Parker reconnection of laminar fields, it leaves reconnection slow. One difficulty with many earlier studies of reconnection in the presence the tearing modes stemmed from the idealized two dimensional geometry assumed for reconnection. In two dimensions tearing modes evolve via a stagnating non-linear stage related to the formation of magnetic islands. This leads to a turbulent reconnection zone [59], but the current sheet remains narrow and its effects on the overall reconnection speed are unclear. This nonlinear stagnation stage does not emerge when realistic three dimensional configurations are considered [53]. In any realistic circumstances field lines are not exactly antiparallel. Consequently, we expect that instead of islands one finds nonlinear Alfvén waves in three dimensional reconnection layers. The tearing instability proceeds with growth rates determined by the linear growth phase while the resulting magnetic structures propagate out of the reconnection region at the Alfvén speed.

The dominant mode will be the longest wavelength mode, whose growth rate will be

$$\gamma \approx \frac{\eta}{\Delta^2}\left(\frac{V_A\lambda_\parallel}{\eta}\right)^{2/5}. \tag{6}$$

The transverse spreading of the plasma in the reconnection layer will start to stabilize this mode when its growth rate is comparable to the transverse shear $V_A/\lambda_\parallel$ [18]. At this point we have $V_{rec,local} \approx \gamma\Delta$ and [53]

$$V_{rec,tearing} = V_A\left(\frac{\eta}{V_A L_x}\right)^{3/10}, \tag{7}$$

which is substantially faster than the Sweet-Parker rate, but still very slow in any astrophysical context. Note that unlike anomalous effects, tearing modes do not require any special conditions and therefore should constitute a generic scheme of reconnection.

Finally, we note that there is a longstanding, but controversial suggestion, that ions tend to scatter about once per cyclotron period, 'Bohm diffusion' [12]. Even if this is correct, the effective diffusivity of magnetic field lines would still be only $\eta_{\rm Bohm} \sim V_A r_c$. While this would be a large increase over Ohmic resistivity, it produces fast reconnection, of order V_A, only if $r_c \sim L_x$. It therefore fails as an explanation for fast reconnection for the same reason that anomalous resistivity does.

2.2 X-point Reconnection

The failure to find fast reconnection speeds through current sheet broadening has stimulated interest in fast reconnection through radically different global geometries. Petschek [67] conjectured that reconnecting magnetic fields would tend to form structures whose typical size in all directions is determined by the resistivity ('X-point' reconnection). This results in a reconnection speed of order $V_A/\ln \mathcal{R}_L$. However, attempts to produce such structures in numerical simulations of reconnection have been disappointing. Typically the X-point region collapses towards the Sweet-Parker geometry as the Lundquist number becomes large [7,8,9,85,58].[1] One way to understand this collapse is to consider perturbations of the original X-point geometry. In order to maintain this geometry shocks are required in the original (Petschek) version of this model. These shocks are, in turn, supported by the flows driven by fast reconnection, and fade if L_x increases. Naturally, the dynamical range for which the existence of such shocks is possible depends on the Lundquist number and shrinks when fluid conductivity increases. The apparent conclusion is that, at least in the collisional regime, reconnection occurs through narrow current sheets.

One may invoke collisionless plasma effects to stabilize the X-point reconnection (for collisionless plasma). For instance, a number of authors [71,70,72] have reported that in a two fluid treatment of magnetic reconnection, a standing whistler mode can stabilize an X-point with a scale comparable to the ion plasma skin depth, $c/\omega_{pi} \sim (V_A/c_s)r_L$. The resulting reconnection speed is a large fraction of V_A, and apparently independent of L_x, which would suggest that something like Petschek reconnection emerges in the collisionless regime. This possibility is discussed at length in the chapter by Bhattacharjee (this volume). However, these studies have not yet demonstrated the possibility of fast reconnection for generic field geometries, since they assume that there are no bulk forces acting to produce a large-scale current sheet. Similarly, those studies do not account for fluid turbulence. Magnetic fields embedded in a turbulent fluid will give fluctuating boundary conditions for the current sheets. On the other hand, boundary conditions need to be fine tuned for a Petschek reconnection scheme [68].

Finally, we note that a number of researchers have claimed that turbulence may accelerate reconnection (for example, [75], where tearing modes are used as the source of the turbulence). The general idea is that turbulent motions can provide an effective transport coefficient $\sim \langle v^2 \rangle \tau$ [65]. However, a closer examination of this process has convincingly demonstrated that an unrealistic amount of energy is required to mix field lines unless they are almost exactly anti-parallel [66]. In the next section we will discuss a mechanism that, when it works, should produce reconnection under a broad range of field geometries, without regard to the particle collision rate.

[1] Recent plasma reconnection experiments [86] do not support Petschek scheme either.

2.3 Stochastic Reconnection

Two idealizations were used in the preceding discussion. First, we considered reconnection in only two dimensions. Second, we assumed that the magnetized plasma has laminar field lines. The Sweet-Parker scheme can easily be extended into three dimensions, in the sense that one can take a cross-section of the reconnection region such that the shared component of the two magnetic fields is perpendicular to the cross-section. In terms of the mathematics nothing changes, but the outflow velocity becomes a fraction of the total V_A and the shared component of the magnetic field will have to be ejected together with the plasma. This result has motivated researchers to do most of their calculations in 2D, which has obvious advantages for both analytical and numerical investigations.

However, physics in two and three dimensions are very different. This is true, for example, in hydrodynamic turbulence, partly because lines of vorticity have different dynamics when they are free to move around one another. Similarly, the ability of magnetic field lines to move past one another in three dimensions dramatically alters the topological constraints on their dynamics. In [53] we considered three dimensional reconnection in a turbulent magnetized fluid and showed that reconnection is fast. This result cannot be obtained by considering two dimensional turbulent reconnection (cf. [59]). This point has been the source of significant confusion. Turbulent reconnection has usually been used to refer to reconnection driven by the turbulent transport of magnetic flux, as discussed in the previous subsection. In other words, one looks for a net flux transport term, operating on microscales, that is proportional to magnetic field gradients and has a coefficient which is independent of the resistivity. This process was recently examined, and severely criticized, in [45], under the mistaken impression that it is the critical physical process in stochastic reconnection. Instead, stochastic reconnection is a geometric effect arising from the appearance of stochastic field line wandering in three dimensions, which gives rise to a broad outflow from the current sheet, but has little effect on the current sheet structure. Below we briefly discuss the idea of stochastic reconnection, while the full treatment of the problem is given in [53].

MHD turbulence guarantees the presence of a stochastic field component, although its amplitude and structure clearly depends on the amplitude and the turbulence driving mechanism. Our *model* of the field line stochasticity also depends on our ability to model generic MHD turbulence. We consider the case in which there exists a large-scale, well-ordered magnetic field, of the kind that is normally used as a starting point for discussions of reconnection. This field may, or may not, be ordered on the largest conceivable scales. However, we will consider scales smaller than the typical radius of curvature of the magnetic field lines, or alternatively, scales below the peak in the power spectrum of the magnetic field, so that the direction of the unperturbed magnetic field is a reasonably well defined concept. In addition, we expect that the field has some small-scale 'wandering' of the field lines. On any given scale the typical angle by which field lines differ from their neighbors is $\phi \ll 1$, and this angle persists for a distance

along the field lines $\lambda_\parallel$ with a correlation distance $\lambda_\perp$ across field lines (see Fig. 2).

The modification of the mass conservation constraint in the presence of a stochastic magnetic field component is self-evident. Instead of being squeezed from a layer whose width is determined by Ohmic diffusion, the plasma may diffuse through a much broader layer, $L_y \sim \langle y^2 \rangle^{1/2}$ determined by the diffusion of magnetic field lines. (Here 'y' is the axis perpendicular to the mean field direction. See Fig. 2.) This suggests an upper limit on the reconnection speed of $\sim V_A(\langle y^2 \rangle^{1/2}/L_x)$. This will be the actual speed of reconnection if the progress of reconnection in the current sheet itself does not impose a smaller limit. The value of $\langle y^2 \rangle^{1/2}$ can be determined once a particular model of turbulence is adopted, but it is obvious from the very beginning that this value is determined by field wandering rather than Ohmic diffusion, as in the Sweet-Parker model.

What about limits on the speed of reconnection that arise from considering the structure of the current sheet? In the presence of a stochastic field component, magnetic reconnection dissipates field lines not over their entire length $\sim L_x$ but only over a scale $\lambda_\parallel \ll L_x$ (see Fig. 2), which is the scale over which magnetic field line deviates from its original direction by the thickness of the Ohmic diffusion layer $\lambda_\perp^{-1} \approx \eta/V_{rec,local}$. If the angle ϕ of field deviation did not depend on the scale, the local reconnection velocity would be $\sim V_A\phi$, independent of resistivity. However, for any realistic model of MHD turbulence, ϕ $(= \lambda_\perp/\lambda_\parallel)$, does depend on scale. Consequently, the *local* reconnection speed $V_{rec,local}$ is given by the usual Sweet-Parker formula but with $\lambda_\parallel$ instead of L_x, i.e. $V_{rec,local} \approx V_A(V_A\lambda_\parallel/\eta)^{-1/2}$. Also, it is apparent from Fig. 2 that $\sim L_x/\lambda_\parallel$ magnetic field lines will undergo reconnection simultaneously (compared to a one by one line reconnection process for the Sweet-Parker scheme). Therefore the overall reconnection rate may be as large as $V_{rec,global} \approx V_A(L_x/\lambda_\parallel)(V_A\lambda_\parallel/\eta)^{-1/2}$. Whether or not this limit is important depends on the value of $\lambda_\parallel$.

The relevant values of $\lambda_\parallel$ and $\langle y^2 \rangle^{1/2}$ depend on the magnetic field statistics. This calculation was performed in [53] using the Goldreich-Sridhar model [33] of MHD turbulence, the Kraichnan model ([41,48]) and for MHD turbulence with an arbitrary spectrum (limited only by some basic physical constraints and which is in rough agreement with observations [1,52,74]). In all the cases the upper limit on $V_{rec,global}$ was greater than V_A, so that the diffusive wandering of field lines imposed the relevant limit on reconnection speeds. Among these, the Goldreich-Sridhar model provides the best fit to observations (e.g. [1,74]) and simulations [22,23]. In this case the reconnection speed was

$$V_{rec,up} = V_A \min\left[\left(\frac{L_x}{l}\right)^{\frac{1}{2}}, \left(\frac{l}{L_x}\right)^{\frac{1}{2}}\right]\left(\frac{v_l}{V_A}\right)^2, \tag{8}$$

where l and v_l are the energy injection scale and turbulent velocity at this scale respectively. We stress that the use of MHD turbulence models here is solely for the purpose of providing a well-defined model of field line stochasticity. The dynamics of the turbulent cascade are largely irrelevant and any process which

leads to small-scale field line stochasticity (e.g. footpoint motions for solar field lines) is a possible cause of fast reconnection.

In [53] we also considered other processes that can impede reconnection and find that they are less restrictive. For instance, the tangle of reconnection field lines crossing the current sheet will need to reconnect repeatedly before individual flux elements can leave the current sheet behind. The rate at which this occurs can be estimated by assuming that it constitutes the real bottleneck in reconnection events, and then analyzing each flux element reconnection as part of a self-similar system of such events. This turns out to limit reconnection to speeds less than V_A, which is obviously true regardless. As the result, (8) is not only an upper limit on the reconnection speed, but is the best estimate of its value.

Naturally, when turbulence is negligible, i.e. $v_l \to 0$, the field line wandering is limited to the Sweet-Parker current sheet and the Sweet-Parker reconnection scheme takes over. However, in practice this requires an artificially low level of turbulence that should not be expected in realistic astrophysical environments. Moreover, the release of energy due to reconnection, at any speed, will contribute to the turbulent cascade of energy and help drive the reconnection speed upward. This may be relevant to the slow onset, and rapid acceleration, of the reconnection process in solar flares.

We stress that the enhanced reconnection efficiency in turbulent fluids is only present if 3D reconnection is considered. In this case ohmic diffusivity fails to constrain the reconnection process as many field lines simultaneously enter the reconnection region. The number of lines that can do this increases with the decrease of resistivity and this increase overcomes the slow rates of reconnection of individual field lines. It is impossible to achieve a similar enhancement in 2D (see [87]) since field lines can not cross each other.

There is a limited analogy one can draw between the enhancement of reconnection speeds in X-point models and increased rate of reconnection due to field line stochasticity. In both cases one gets a boost from a reduced parallel length scale. In the case of X-point models this effect is, usually by design, enormous since $L_x \to \Delta$. Stochastic reconnection depends on a relatively modest enhancement, since $L_x \to \lambda_\parallel(\Delta) \gg \Delta$. The bulk of the effect comes from the simultaneous reconnection of many independent flux elements, and the steady diffusion of the ejected plasma away from the current sheet. The main problem with X-point reconnection models, their tendency to collapse to narrow current sheets, is absent in stochastic reconnection, since in the latter case the current sheets stay narrow, and the diverging field lines are separated by other field lines, rather than by unmagnetized plasma.

A more subtle difficulty arises from our prescription for the structure of the stochastic field near the current sheet. We have assumed that we can apply the statistically homogeneous prescription for field line perturbations in a turbulent medium near planes where there is a dramatic change in the structure of the large-scale magnetic field. This is not obvious. It may be that the presence of a strong shear in the field acts as a kind of internal surface, producing an altered,

and perhaps greatly reduced, level of stochasticity. This kind of internal 'shadowing' does not appear in current simulations, but there has been little attempt to look for it, and the issue can only be resolved when detailed numerical simulations of stochastic reconnection are performed. Similarly, one may wonder if the systematic ejection of plasma along the field lines might modify their topological connections. In this case it seems more plausible to suppose that this would lead to an increase in the diffusion rate, rather than a decrease, but again no simulations of this process are available.

2.4 Reconnection in Partially Ionized Gas

A substantial fraction of the ISM in our galaxy is partially ionized, as well as photospheres of most stars. This motivates studies of the effect of neutrals on reconnection and MHD turbulence. The role of ion-neutral collisions is not trivial. On one hand, neutral particles tend to have a substantially longer mean free path, so that drag between the neutrals and ions may truncate the turbulent cascade at a relatively large scale. On the other hand, the ability of neutrals to diffuse perpendicular to magnetic field lines enhances reconnection rates, at least in the Sweet-Parker model.

Reconnection in partially ionized gases has been studied by various authors ([63,88,84]) in the context of the Sweet-Parker reconnection model. Our comments here are based on [84] where we studied the diffusion of neutrals away from the reconnection zone. In general, in a partially ionized gas the reconnection zone consists of two distinct regions. A broad region, whose width is determined by the ambipolar diffusivity, $\eta_{ambi} \approx V_A^2/t_{ni}$ where t_{ni} is the neutral-ion collision rate, and a narrow region whose width is determined by the Ohmic diffusivity. Magnetic reconnection takes place in the narrow region, while the broader region allows a more efficient ejection of matter.

If the recombination time is short, then ions and neutrals are largely interchangeable and the reconnection speed is [84]

$$V_{rec} \approx V_A \left(\frac{V_A t_{in}}{L_x} \right)^{1/2} . \tag{9}$$

This is faster than the Sweet-Parker rate, but not fast in the sense of allowing reconnection speeds close to V_A. In practice, even this rate is typically unachievable. Under typical interstellar conditions the reconnection speed is limited by the recombination rate. That is, the rate at which ions recombine and leave the resistive region determines the speed of the whole process. Consequently, the ambipolar reconnection rates obtained in [84] are insufficient either for fast dynamo models or for the ejection of magnetic flux prior to star formation. In fact, the increase in the reconnection speed stems entirely from the compression of ions in the current sheet, with the consequent enhancement of both recombination[2] and

[2] In the model [84] it is assumed that the ionization is due to cosmic rays. In the case of photoionization of the heavy species, e.g. carbon, the recombination and therefore the reconnection rates are lower.

ohmic dissipation. This effect is small unless the reconnecting magnetic field lines are almost exactly anti-parallel. As above, we expect that including the effects of anomalous resistivity and tearing modes may enhance reconnection speeds appreciably, but not to the extent of producing fast reconnection.

None of this work included the effects of field line stochasticity, which is critical for producing fast reconnection in ionized plasmas. We expect that in this case also the presence of turbulence will lead to substantially higher reconnection speeds. However, whether or not this produces fast reconnection must depend on the nature of the turbulent cascade in a partially ionized gas. Recent work, which is discussed in detail in the chapter by Cho, Lazarian & Vishniac in this volume, show that the magnetic field in a partially ionized gas has a much more complex structure than it is usually assumed. In fact, in [25] we reported a new regime of MHD turbulence which is characterized by the existence of intermittent magnetic structures below the viscous cutoff scale. The root mean square perturbed magnetic field strength in these structures does not drop at smaller scales. However, the *curvature* scale (and therefore the divergence rate) for these structures does not decrease significantly as their perpendicular scale decreases. At sufficiently small scales the ions and neutrals will decouple, and a turbulent cascade, extending down close to resistive scales but involving only ions will appear.

The existence of strong magnetic field structures on small scales, and the reappearance of a strong turbulent cascade at very small scales, should lead to fast reconnection speeds through stochastic reconnection. However, it remains to be seen whether or not the intermediate scales, characterized by weak divergence of field lines, will impose a significant bottleneck on the reconnection plasma outflow. If it does, then the implication is that interstellar clouds with small ionized fractions may not allow fast reconnection. This conclusion would not pose any problems with galactic dynamo, but may be extremely important for other essential processes, e.g. star formation. This issue is examined further in [56].

3 The Dynamo Process

3.1 Conventional Theory and Its Problems

We start this section by briefly reviewing the standard approach to dynamo theory, and discussing various objections to it. Some of these objections center around the speed of reconnection, and can be safely ignored if reconnection is fast in a turbulent environment. In fact, since stochastic reconnection depends on small-scale structure in the magnetic field, the claim that small-scale structure tends to accumulate energy faster than the large-scale field [50] can be seen as self-limiting. A disproportionate growth in power on small scales will only continue until the reconnection speed is boosted to large fraction of V_A. However, as we have already mentioned, some objections to dynamo theory are more subtle and require substantial modification to mean-field dynamo theory.

The usual approach to the dynamo problem is to take (1), set $\eta = 0$, and divide the velocity field into small-scale turbulence and some large-scale rotational motion. In order to follow the evolution of the large-scale magnetic field, we write

$$\mathbf{B} \equiv \langle \mathbf{B} \rangle + \mathbf{b}. \tag{10}$$

The brackets here denote averaging over scales somewhat larger than the turbulent eddy size. In other words, they indicate a smoothing process which averages out all small-scale features. The field $\langle \mathbf{B} \rangle$ is the 'mean field'. The dynamo process can be written in mathematical terms by approximating the evolution of the small-scale field component, $\mathbf{b}$, as

$$\partial_t \mathbf{b} \approx \nabla \times (\mathbf{v} \times \langle \mathbf{B} \rangle), \tag{11}$$

and substituting the result into the evolution equation for the large scale field,

$$\partial_t \langle \mathbf{B} \rangle = \nabla \times \langle \mathbf{v} \times \mathbf{b} \rangle. \tag{12}$$

In a turbulent, incompressible and homogeneous plasma this implies

$$\partial_t \langle \mathbf{B} \rangle = \nabla \times (\alpha \cdot \mathbf{B}) + \nabla \cdot (\mathbf{D_T} \cdot \nabla) \langle \mathbf{B} \rangle. \tag{13}$$

Here α, the kinetic helicity, and $\mathbf{D_T}$, the turbulent diffusion tensor, are dyads given by

$$\alpha_{il} \equiv \epsilon_{ijk} \langle v_j \partial_l v_k \rangle \tau_c, \tag{14}$$

and

$$D_{T,ij} \equiv \langle v_i v_j \rangle \tau_c, \tag{15}$$

where τ_c is the eddy correlation time. The component of the electromotive force along the large scale field direction, $\langle \hat{\mathbf{B}} \rangle \cdot \langle \mathbf{v} \times \mathbf{b} \rangle$, is the piece that can drive an increase in the large scale magnetic field. (The component perpendicular to $\langle \mathbf{B} \rangle$ gives an effective large scale field velocity, that is, it affects the transport of the field rather than its generation.) The trace of α divided by τ_c is what is usually referred to as the kinetic helicity, and it is often assumed for convenience that α is a scalar times the identity matrix. In symmetric turbulence α vanishes, but $\mathbf{D_T}$ does not. In fact, since a successful dynamo requires non-vanishing diagonal components for α, we can see from this expression that a successful dynamo should require symmetry breaking along all three principal axes.

The appearance of D_T in (13) would seem to vindicate the use of turbulent diffusion in astrophysical MHD. There are two reasons why this is not quite right. First, fast reconnection is implicit in this kind of averaging argument. Rather than appealing to turbulent diffusion as an explanation for fast reconnection, we are actually using our understanding of fast reconnection to explain diffusion. The second point is less formal and more important. Equation (13) is not a realistic description of the evolution of $\langle \mathbf{B} \rangle$. As noted in Sect. 1, twisting magnetic field lines into spirals is not easily accomplished, and numerical simulations do not support the use of (13).

3.2 Magnetic Helicity Conservation Constraint

The fundamental problem is that there is an important mathematical constraint that follows from (1), which is not respected by (13). The magnetic helicity, defined as $H \equiv \mathbf{A} \cdot \mathbf{B}$ evolves according to

$$\partial_t H = -\nabla \cdot [\mathbf{A} \times (\mathbf{v} \times \mathbf{B} + \nabla \mathbf{\Phi})] - \eta \mathbf{B} \cdot \nabla \times \mathbf{B}, \tag{16}$$

where Φ is an arbitrary function of space and time. For the Coulomb gauge, which turns out to be a convenient choice, we require

$$\nabla^2 \Phi = \nabla \cdot (\mathbf{v} \times \mathbf{B}). \tag{17}$$

In the limit of vanishing resistivity, this not only implies that the volume integrated magnetic helicity vanishes, it also implies that the magnetic helicity of any individual flux tube is separately conserved [79].

For a non-zero, but very small, η, we can transfer magnetic helicity from one flux tube to another. However, since H is of order LB^2, where L is a characteristic scale of the field, it takes less energy to hold magnetic helicity on large scales than on eddy scales, and a divergent amount on infinitesimal scales. Consequently, in the limit of vanishing resistivity the resistive term in (16) does not affect the global conservation of helicity, even in the presence of fast reconnection, as long as reconnection only occurs in an infinitesimal fraction of the plasma volume. On the other hand, the conservation of magnetic helicity for individual flux tubes is completely lost. The implication is that global magnetic helicity conservation is a good approximation for laboratory plasmas, a point that was originally stressed by Taylor [79], and an even better one for astrophysical systems.

How does this affect dynamo theory? The large scale distribution of magnetic helicity can be divided into a piece carried by large scale magnetic structures and a piece carried by small-scale structures, or

$$\langle H \rangle = \langle \mathbf{A} \rangle \cdot \langle \mathbf{B} \rangle + \langle \mathbf{a} \cdot \mathbf{b} \rangle. \tag{18}$$

Henceforth we will use $h \equiv \langle \mathbf{a} \cdot \mathbf{b} \rangle$. The evolution of the first piece, in a perfectly conducting fluid, is

$$\partial_t (\langle \mathbf{A} \rangle \cdot \langle \mathbf{B} \rangle) = 2 \langle \mathbf{B} \rangle \cdot \langle \mathbf{v} \times \mathbf{b} \rangle - \nabla \cdot [\langle \mathbf{A} \rangle \times (\langle \mathbf{v} \times \mathbf{b} \rangle + \nabla \langle \Phi \rangle)] \,. \tag{19}$$

The second term on the right hand side is the magnetic helicity transport driven by mean-field terms. The first represents the exchange of magnetic helicity between large and small scales. This term is proportional to the component of the electromotive force which drives the dynamo process. In other words, the generation of a large scale magnetic field is a direct consequence of the transfer of magnetic helicity between large and small scales.

The point that MHD turbulence transfers magnetic helicity to the largest available scales, even if that scale is much larger than any eddy scale, is well known [30,76,77]. We can estimate the rate at which h is transferred to large

scale magnetic field structures by considering its role in biasing the value of the electromotive force [83]. The inverse cascade rate is

$$\tau_{cascade}^{-1} \sim \frac{V_A^2}{\langle v^2 \rangle} \tau_c^{-1}. \tag{20}$$

For a large scale magnetic field in equipartition with the turbulent cascade this implies that magnetic helicity is transferred to the large scale field in one eddy turn over time. This suggests that unless the large scale field is very weak it is reasonable to take

$$H \approx \langle \mathbf{A} \rangle \cdot \langle \mathbf{B} \rangle. \tag{21}$$

Then combining (19) and (21) we see that

$$2\langle \mathbf{B} \rangle \cdot \langle \mathbf{v} \times \mathbf{b} \rangle = -\nabla \cdot [\langle \mathbf{a} \times (\mathbf{v} \times \mathbf{B} + \nabla \phi) \rangle] \equiv -\nabla \cdot \mathbf{J}_H, \tag{22}$$

where J_H is defined as the magnetic helicity current carried by small-scale structures, or the anomalous magnetic helicity current. That is, the component of the electromotive force parallel to the large scale magnetic field is given by the divergence of the magnetic helicity current carried by eddy scale structures. If $J_H \approx 0$, then it follows from (22) that mean-field dynamos are impossible. This argument was advanced by Gruzinov and Diamond [34,35] who pointed out that magnetic helicity conservation combined with the assumption of stationary statistics for small-scale structure implied almost complete suppression of the kinematic dynamo. We note also that the form of the parallel component of the electromotive force given in (22) has been suggested before [6], although the interpretation that the relevant current is a magnetic helicity current appeared somewhat later [42,46]. Here we will follow the treatment in [83], where the magnetic helicity current was derived for the first time for homogeneous turbulence.

Equations (12) and (22) yield

$$\partial_t \langle \mathbf{B} \rangle = \nabla \times \left[\frac{-\langle \mathbf{B} \rangle}{2\langle B \rangle^2} \nabla \cdot \mathbf{J}_H + \langle \mathbf{v} \times \mathbf{b} \rangle_\perp \right], \tag{23}$$

where the second term on the right hand side is the component of the electromotive force perpendicular to the large scale field direction. Evaluating $\mathbf{J}_H$ is necessary to understand the dynamo process. By contrast, attempts to estimate the kinetic helicity only tell us about the dynamo process when the large scale magnetic field is so weak that the transfer of magnetic helicity between scales is unaffected by the extremely limited capacity of the turbulent eddies to store magnetic helicity.

The most direct way to estimate the anomalous magnetic helicity current is to write $\mathbf{a}$ in terms of the action of the turbulent velocity field on the large scale magnetic field, or

$$\mathbf{a} \approx (\mathbf{v} \times \mathbf{B} - \nabla \phi) \tau_c, \tag{24}$$

where

$$\nabla^2 \phi = \nabla \cdot (\mathbf{v} \times \mathbf{B}). \tag{25}$$

If we substitute this into the definition of the magnetic helicity current we find, after some manipulation, that

$$\mathbf{J}_H = -\tau_c \int \frac{d^3\mathbf{r}}{4\pi r} \epsilon_{lmn} \langle B_k \rangle \langle B_l \rangle \langle \partial_k \partial_m \langle v_i(\mathbf{x}) v_n(\mathbf{x}+\mathbf{r}) \rangle. \tag{26}$$

We see that $\mathbf{J}_H$ is parity-invariant, unlike α. In completely isotropic turbulence it will also vanish, but the degree of symmetry breaking necessary for a dynamo effect is smaller than in the conventional picture. There will also be contributions to J_H driven by the effects of background structure, but for strongly rotating systems these will be smaller than the expression given here.

Equation (26) is not a particularly enlightening expression, but we can gain somewhat more insight by rewriting it as

$$\mathbf{J}_H \approx -\lambda_c^2 \tau_c \langle \langle \mathbf{B} \rangle \cdot \omega (\langle \mathbf{B} \rangle \cdot \nabla) \mathbf{v} \rangle, \tag{27}$$

where λ_c is some suitably averaged eddy size and $\omega \equiv \nabla \times \mathbf{v}$ is the fluid vorticity. This corresponds to twisting a field line in both directions, but then systematically moving right and left handed spiral segments in opposite directions.

In this model, we generate left (or right) handed spirals by separating segments of the same field line with different helicities, moving them in opposite directions, and then reconnecting them (along two dimensional) surfaces, into new field lines. If the flow of magnetic helicity has a non-zero divergence, then the new field lines will have a preferred sense of twisting, and the first step in the usual scheme for the dynamo process will have been completed without violating magnetic helicity conservation. We illustrate this modified version of mean-field dynamo action in Fig. 3.

In some sense (22) gives the *minimal* change in dynamo theory which respects conservation of magnetic helicity, since it leaves $\mathcal{E}_{mf\perp}$ unchanged. That is, we assume that the backreaction from small scales affects only the component of the electromotive force along the direction of the large scale mean field. While this may seem unduly optimistic, we note that in the absence of any other large scale vector quantity, symmetry considerations alone should be sufficient grounds for this assumption. However, we are often concerned with circumstances where other large scale vectors are present, for example systems with differential rotation. In this case we are not guaranteed that $\langle \mathbf{v} \times \mathbf{b} \rangle_\perp$ is unaffected by the inverse cascade. Obviously this is an important direction for future work.

There is another way in which (26) may fail to give the full magnetic helicity current. Equation (24) takes into account the perturbations in the large scale magnetic field driven by small-scale velocities. However, fluctuations in the magnetic field also drive the velocity field. In conventional mean-field dynamo theory these considerations lead to replacing (14) with

$$\alpha_{il} \equiv \epsilon_{ijk} \left[\langle v_j \partial_l v_k \rangle - \frac{\langle b_j \partial_l b_k \rangle}{4\pi\rho} \right] \tau_c. \tag{28}$$

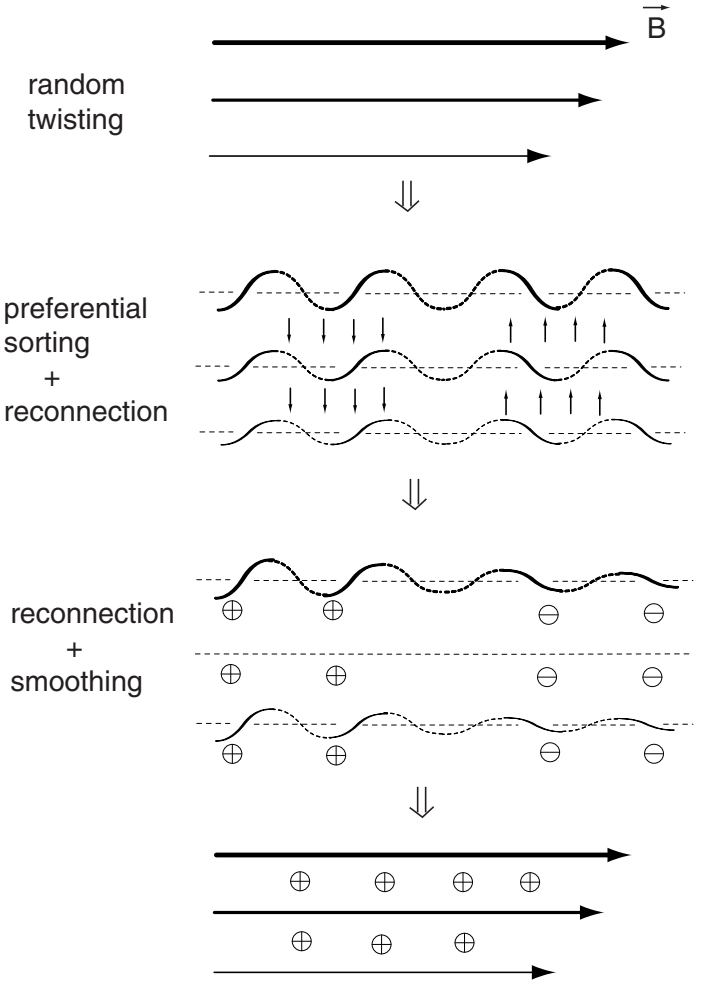


Fig. 3. A new version of the mean-field dynamo. Turbulence twists the field lines into spirals, with each field line accumulating regions of right and left handed twisting. The turbulence is biased so that left-handed segments move down and right-handed segments move up. Reconnection, and a gradient in the strength of the spirals gives us new field lines with a net left-handed twist. The last step involves averaging over each field line, which is trivial but not part of the original picture, and reconnection to produce a new orthogonal field component.

In our case (26) can be replaced by a symmetrized version, that is

$$\mathbf{J}_H = -\tau_c \int \frac{d^3\mathbf{r}}{4\pi r} \epsilon_{lmn} \langle B_k \rangle \langle B_l \langle \partial_k \partial_m \left[\langle v_i(\mathbf{x}) v_n(\mathbf{x}+\mathbf{r}) - \frac{b_i(\mathbf{x}) b_n(\mathbf{x}+\mathbf{r})}{4\pi\rho} \rangle \right]. \tag{29}$$

This form fails to take into account the effects of shear from a large scale velocity field (like differential rotation). Given perfect symmetry between the dynamics of

the velocity and magnetic fields, this term will be zero, regardless of any spatial symmetry-breaking effects. MHD turbulence tends to evolve towards this kind of symmetry. However, on the scale of the largest energy containing eddies, in realistic systems, we expect that differential shear, background gradients, and specific dynamical instabilities may all play a role, and all these effects will not respect the symmetry between magnetic and velocity fields. An example of an instability which will necessarily produce such an asymmetry is the magneto-rotational (or Balbus-Hawley) instability in accretion disks [81,21,2], which is discussed below.

Here we summarize some of the more important conclusions from these arguments.

1. The fluid helicity is largely irrelevant to flux generation, except when the field amplitude is very small, although it does affect flux transport.
2. This prescription eliminates turbulent dissipation for currents aligned with the large scale magnetic field, but continues to damp other current components. This is not a qualitative change in the role of turbulent damping for most field configurations, but does imply that force-free large scale fields are protected against turbulent dissipation.
3. The anomalous magnetic helicity current, j_H, depends on $\langle([\mathbf{B}\cdot\nabla]\mathbf{v})(\mathbf{B}\cdot[\nabla\times\mathbf{v}])\rangle$, which has no particular relationship to the fluid helicity and is parity invariant. Rather than violating spatial symmetry along all three principal axes, a successful dynamo can result from a situation where only two out of three directions have broken symmetries. An example of this is differential rotation in a vertically uniform cylinder.
4. The '$\alpha-\Omega$' dynamo has an analog in this new theory, which gives a similar growth rate, but which does not depend on any background vertical structure. (The analogous effect is provided by vertical gradients in $\langle\mathbf{B}\rangle$.) In an accretion disk, the success of the dynamo is tied to the outward transport of angular momentum [83].
5. The analog of the 'α^2' dynamo of conventional theory has the difficulty that the turbulent dissipation term is of the same order as the driving term resulting from magnetic helicity transport. This does not imply that this kind of dynamo is impossible, but configurations with force-free, or nearly force-free, fields are strongly favored.
6. The analog of the '$\alpha-\Omega$ dynamo has a strong vertical magnetic helicity current, which has the same sign as $-\partial_r\Omega(r)$, i.e. towards $\hat{z}$ for an accretion disk and $-\hat{z}$ for a star like the Sun. This implies the necessity for magnetic helicity ejection at the system boundaries. While the energy budget for this is small compared to the energy budget of the dynamo in an accretion disk, it represents an unsolved aspect of the dynamo physics.

We note that the ejection of magnetic helicity from rotating systems as a necessary part of the dynamo process has been suggested by other authors [10,11], although the terms of the discussion were somewhat different. In particular, they were concerned with removing the magnetic helicity constraint by removing the

magnetic helicity. Although realistic systems can work this way, there is no fundamental reason why the magnetic helicity can't circulate within a closed system and produce a dynamo effect by being carried by large scale fields in one part of the system and eddy scale fields in another.

4 Applying and Testing the Theory

Both reconnection and the dynamo are the subject of intensive experimental research. Magnetic reconnection is being studied on several dedicated experiments around the world (MRX, TS3/4, SSX, VTF). In each experiment, magnetized loops are generated and merged. At present, the physical scales of such experiments are 0.1 to 1 m and the Lundquist number is about 1000. Sophisticated diagnostics are used to get plasma and magnetic field parameters in these experiments. This enables testing theoretical predictions. The direct relation between those experiments and astrophysics is complicated by the fact that some of the effects that are important in laboratory, e.g. anomalous resistivity, may not be important in conditions of astrophysical plasma, e.g. interstellar gas.

Dynamo experiments, e.g. using liquid sodium, are mostly focused on the reproduction of the dynamo effect for the low Lundquist numbers. In this regime, neither reconnection nor magnetic helicity are expected to provide strong constraints on the evolution of the experiments.

On the other hand, numerical simulations are a valuable source of information for mean-field dynamo theory. We have already discussed their role in undermining the conventional approach to this topic. It is also important to note that there are a large number of simulations which seem to show the operation of a successful dynamo, in the sense that they demonstrate the growth of a magnetic field with a significant component at large spatial scales. These are the simulations of magnetized, ionized accretion disks (e.g. [36,39,40,73,14,15] see also [3,38] for a review) which are subject to the magneto-rotational instability. These simulations include extremely simplified physics and cover a limited set of spatial scales and geometries. Nevertheless, they agree in a number of important aspects, namely:

1. Any magnetic seed field undergoes substantial amplification to a final state which is (apparently) independent of initial conditions. In this state the magnetic field pressure is a few percent of the gas pressure and the dimensionless 'viscosity', α_{SS} is about half of this ratio. A large fraction of the magnetic energy is contained in a large scale field with a domain size which is a large fraction of the simulation box size. This large scale field is not static, but varies on time scales of tens of shearing times, a feature which seen in all of the simulations cited above.
2. The growth in field strength is rapid, i.e. a significant fraction of the shear rate, even when the field is weak. The growth is several times slower for an initially azimuthal field, where the amplification depends on dynamo action, as for an initially vertical field, where the amplification can simply reflect the linear growth of the instability, which generates azimuthal field, but the point remains true in either case.

3. The magnetic field pressure in the saturated state is not a constant fraction of the ambient pressure, but varies from a small fraction at the midplane (for models with vertical structure) to a value comparable to the gas pressure a few scale heights away from the midplane. This effect is particularly dramatic in the recent simulations of Miller and Stone [61]. This distribution does not seem to be due to magnetic buoyancy [73], that is, the fields are mostly generated locally. The time averaged magnetic stress ($\langle B_r B_\theta \rangle$) is somewhat more uniform.
4. The dynamo persists when vertical gravity is turned off, although the simulations are slightly different in this case [37,39].

A possible interpretation of these results is that the simulations are showing a chaotic dynamo [5,44], in which turbulent stretching of embedded magnetic field lines results in a runaway amplification of the magnetic field. One problem with this is that the accretion disk simulations are unique in generating substantial magnetic field energy on scales much larger than the typical eddy size. Simulations of MHD turbulence in a box typically produce magnetic field structure whose energy spectrum peaks on scales slightly *smaller* than a typical eddy scale and with a total energy density which is a (large) fraction of the kinetic energy density. At longer wavelengths the magnetic energy density falls, although slowly. Another point is that the vertical distribution of magnetic energy is not simply a reflection of local conditions, but seems to show some sort of global field evolution. The obvious conclusion is that some sort of large scale dynamo effect is being produced in the simulations, as a consequence of the Balbus-Hawley instability. Since there has been no attempt to look specifically at the flow of magnetic helicity, it is difficult to know whether to ascribe the dynamo effect to a locally produced fluid helicity (in which case the dynamo should slow down as the resolution is increased) or to a turbulently driven magnetic helicity current. The dynamo growth rate does not *appear* to slow down in the higher resolution studies, but this has not been examined critically. Further study of these simulations should allow testing of the notion that the turbulently driven magnetic helicity current is playing a critical role in these simulations.

There is one indirect test which has already been performed. Equation (27) can be used to show, via integration by parts, that the sign of the magnetic helicity current depends on the direction of angular momentum transport. Reversing the sign of the magnetic helicity current has the effect of turning off the dynamo. One simple numerical experiment is to conduct a simulation in which the angular momentum current flows in the opposite direction. This has been done [40] by turning off the centrifugal force term, so that the turbulence is driven only by a kind of magnetized Kelvin-Helmholtz instability. The dynamo effect was suppressed and the magnetic field decayed away, after an initial burst of growth. This simulation had a limited dynamic range, so that all the eddies were dominated by the local shear. Consequently, the elimination of the dynamo effect led to a complete suppression of the magnetic field through azimuthal stretching and radial mixing.

There has been a recent attempt to combine shear with an asymmetrically driven turbulence [17] to produce a non-heliacal dynamo. The results were disappointing. The expected correlation between the magnetic helicity flux and the velocity correlation seen in (27) was found, but the magnetic helicity flux was largely divergenceless and there was no clear correlation between its small divergence and the electromotive force. This may have been due to the boundary conditions, which forced a return loop of magnetic flux within the box. Clearly further numerical experiments would be useful.

Assuming that we can understand the conceptual basis of accretion disk dynamos, it should be possible to construct a useful mean-field theory that incorporates transport effects and allows us to predict the dynamics of accretion disk fields. This model will need to incorporate the effects of fluctuations in the electromotive force [82]. In the conventional mean-field dynamo theory such fluctuations have been shown to be capable of driving a mean-field dynamo in the absence of any average helicity. Their role in the modified version of mean-field dynamo theory is not yet understood. Such a model would be useful for building models of disks that incorporate both realistic local physics and MHD turbulence. A similar effort should be made for stellar dynamos, although there has been, as yet, no progress in this direction. There has been work on the galactic dynamo [47] which incorporates the notion of magnetic helicity current, including the term given in (26).

Much less progress has been made in numerical simulations of stochastic reconnection. This is particularly unfortunate since magnetic reconnection is one of the most fundamental properties of the magnetic field dynamics in the conducting fluid, and its applications are not limited to its consequences for astrophysical dynamos. In fact, reconnection is likely to be extremely important for the dynamics of the advection dominated flows, star formation, propagation (see [26]) and acceleration (see [28,29]) of cosmic rays, dynamics of charged dust [55]. Direct study of the reconnection layer is difficult as both very small scales (turbulent microscales comparable to the current sheet thickness) and large scales (the contact region scale) are present in the problem. The requirement that we evolve structures at all scales over the whole broad reconnection region suppress any hope that adaptive mesh codes would be very helpful. Nevertheless, simple diagnostics may be used to distinguish fast stochastic reconnection from the Sweet-Parker model. For instance, using MHD simulations we can measure currents J, magnetic fields B and velocities v. If we divide $\langle J^2[(B\cdot\nabla)B\cdot v]^2\rangle$ by $\langle J^2(\nabla\times B)^2\rangle$ then we have a measure of the rms magnetic field across a typical current sheet times the speed with which it is expelled. An approximate measure of reconnection speed can then be obtained by dividing the result by $\langle(J\times B)^2\rangle/\langle J^2\rangle$ and taking the square root. Although reconnection rates for low Lundquist numbers are not so different for the Sweet-Parker and stochastic reconnection models, the scaling of the reconnection rates with the Alfven velocity are very different. This gives some hope that the stochastic reconnection model can be tested before long.

5 Discussion and Summary

It is not possible to understand the astrophysical dynamo and dynamics of magnetized astrophysical plasmas without understanding how magnetic fields evade the topological constraints imposed by flux-freezing. This obviously includes the problem of reconnection, but also the more subtle difficulty posed by magnetic helicity conservation. Here we have compared traditional approaches to the problem of magnetic reconnection and the mean-field dynamo with new approaches based on an explicit recognition of the role geometry plays in both these problems. In fact, one of the more striking aspects of stochastic reconnection model [53] is that the global reconnection speed is relatively insensitive to the actual physics of reconnection. Equation (8) only depends on the nature of the turbulent cascade. In the end, reconnection can be fast because if we consider any particular flux element inside the contact volume, assumed to be of order L_x^3, the fraction of the flux element that actually undergoes microscopic reconnection vanishes as the resistivity goes to zero. This is turn implies that reconnection is not tightly coupled to electron heating. More generally, the results presented here suggest that, in most cases, microphysics is irrelevant to the dynamo process.

Although objections to conventional dynamo theory tend to conflate the issues of reconnection and magnetic helicity conservation, these are, in fact, two separate problems, for which we have proposed two separate resolutions. Taken together, they imply that astrophysical dynamos are capable of operating in a broad range of circumstances. However, it is important to remember that they stand separately. Conventional dynamo theory is not rescued by assuming rapid reconnection, although it requires it. Conversely, the use of (22) and (26) to describe dynamo activity do not require stochastic reconnection, but only that *some* model of fast reconnection work.

Our main conclusions are as follows:

- The rate of magnetic reconnection is increased dramatically in the presence of a stochastic component to the magnetic field. Even when the turbulent cascade is weak the resulting reconnection speed is independent of the Ohmic resistivity. However, it is extremely sensitive to the level of noise. This may explain the variable rates of magnetic reconnection seen in the solar corona. It also implies that laminar flow patterns that drive a magnetic helicity current may still require some level of local turbulence in order to drive a large scale dynamo.
- The argument that the rapid rise of random magnetic field associated with dynamo action results in the suppression of dynamo [50] is untenable since the increase of the random component of the magnetic field increases the reconnection rate. We conclude that dynamo is a self-regulating process.
- Conventional mean-field dynamo theory, which does not account for the conservation of magnetic helicity is ill-founded. The suggested modification of the mean-field dynamo equations allow us to account for results of numerical sumulations and make the theory, for the first time ever, self-consistent.

Acknowledgements

AL and JC acknowledge the support of NSF grant NSF AST-0125544. ETV acknowledges the support of NSF grant AST-0098615. AL thanks the organizers for the financial support.

References

1. J. W. Armstrong, B.J. Rickett, S.R. Spangler: Astrophys. J. **443**, 209 (1995)
2. S. A. Balbus, J. F. Hawley: Astrophys. J. **376**, 214 (1991)
3. S. A. Balbus, J. F. Hawley: Rev. Mod. Phys. **70**, 1 (1998)
4. D. Balsara: Rev.Mex.A.A. **9**, 9 (2000)
5. G. K. Batchelor: Proc. R. Soc. Lond. **A201**, 405 (1950)
6. A. Bhattacharjee, E. Hameiri: Phys. Rev. Lett. **57**, 206 (1986)
7. D. Biskamp: Phys. Lett. **A 105**, 124 (1984)
8. D. Biskamp: Phys. Fluids **29**, 1520 (1986)
9. D. Biskamp: Astrophys. & Sp. Sci. **242**, 165 (1996)
10. E. G. Blackman, G. B. Field: Astrophys. J. **534**, 984 (2000)
11. E. G. Blackman, G. B. Field: Mon. Not. R. A. S. **318**, 724 (2000)
12. D. Bohm, E. H. S. Burhop, H. S. W. Massey: In: *The Characteristics of Electrical Discharges in Magnetic Fields*, ed. by A. Guthrie & R.K. Wakerling (New York: McGraw Hill 1949) pp. 77-86
13. A. Brandenburg: Astrophys. J. **550**, 824 (2001)
14. A. Brandenburg, Å. Nordlund, R. F. Stein, U. Torkelsson: Astrophys. J. **446**, 741 (1995)
15. A. Brandenburg, Å. Nordlund, R. F. Stein, U. Torkelsson: Astrophys. J. Lett. **458**, 45 (1996)
16. A. Brandenburg, E. Zweibel: Astrophys. J. **448**, 734 (1995)
17. A. Brandenburg, A. Bigazzi, K. Subramanian: Mon. Not. R. Astron. Soc. **325(2)**, 685 (2001)
18. S. V. Bulanov, J. Sakai, S. I. Syrovatskii: Sov. J. Plasma Phys. **5(2)**, 157 (1979)
19. F. Cattaneo, D. W. Hughes: Phys. Rev. **E**, **54**, 4532 (1996)
20. F. Cattaneo, S. I. Vainshtein: Astrophys. J. Lett. **376**, 21 (1991)
21. S. Chandrasekhar: *Hydrodynamic and Magnetohydrodynamic Stability*(Oxford: Oxford University Press 1961)
22. J. Cho, E. T. Vishniac: Astrophys. J. **538**, 217 (2000a)
23. J. Cho, E. T. Vishniac: Astrophys. J. **539**, 273 (2000b)
24. J. Cho, A. Lazarian, E. T. Vishniac: Astrophys. J. **564**, 291 (2002)
25. J. Cho, A. Lazarian, E. T. Vishniac: Astrophys. J. Lett. **566**, 49 (2002)
26. J. Cho, A. Lazarian, H. Yan: In: *Seeing through dust*, eds, ASP conf. ser. Russ Taylor, Tom Landecker, and Tony Willis, (2002)
27. K. P. Dere: Astrophys. J. **472**, 864 (1996)
28. E. M. de Gouveia Dal Pino, A. Lazarian: Astrophys. J. **536**, 31 (2000)
29. E. M. de Gouveia Dal Pino, A. Lazarian: Astrophys. J. **560**, 358 (2001)
30. U. Frisch, A. Pouquet, J. Leorat, A. Mazure: J. Fluid Mech. **68**, 769 (1975)
31. H. P. Furth, J. Killeen, M. N. Rosenbluth: Phys. Fluids **6**, 459 (1963)
32. G. A. Glatzmaier, P. H. Roberts: Nature **377**, 203 (1995)
33. P. Goldreich, S. Sridhar: Astrophys. J. **438**, 763 (1995)
34. A. Gruzinov, P. H. Diamond: Phys. Rev. Lett. **72**, 1651 (1994)

35. A. Gruzinov, P. H. Diamond: Phys. Plasmas **3**, 1853 (1996)
36. J. F. Hawley, S. A. Balbus: Astrophys. J. **376**, 223 (1991)
37. J. F. Hawley, S. A. Balbus: Astrophys. J. **400**, 610 (1992)
38. J. F. Hawley, S. A. Balbus: Phys. Plasmas **6(12)**, 4444 (1999)
39. J. F. Hawley, C. F. Gammie, S. A. Balbus: Astrophys. J. **440**, 742 (1995)
40. J. F. Hawley, C. F. Gammie, S. A. Balbus: Astrophys. J. **464**, 690 (1996)
41. P. Iroshnikov: Astron. Zh. **40**, 742 (1963) [Sov. Astron. **7**, 566 (1963)]
42. H. Ji: Phys. Rev. Lett. **83**, 3198 (1999)
43. S. R. Kane, K. Hurley, J. M. McTiernan, M. Boer, M. Niel, T. Kosugi, M. Yoshimori: Astrophys. J. **500**, 1003 (1998)
44. A. P. Kazantsev: JETP **53**, 1806 (1967)
45. E.-J. Kim, P. H. Diamond: Astrophys. J. **556**, 1052 (2001)
46. N. Kleeorin, D. Moss, I. Rogachevskii, D. Sokoloff: Astro. & Astrophys. Lett. **361**, 5 (2000)
47. N. Kleeorin, D. Moss, I. Rogachevskii, D. Sokoloff: Astro. & Astrophys. **387**, 453 (2002)
48. R. Kraichnan: Phys. Fluids **8**, 1385 (1965)
49. F. Krause, K. H. Radler: *Mean-Field Magnetohydrodynamics and Dynamo Theory* (Oxford: Pergamon Press 1980)
50. R. M. Kulsrud, S. W. Anderson: Astrophys. J. **396**, 606 (1992)
51. A. Lazarian: Astron. & Astrophys. **264**, 326 (1992)
52. A. Lazarian, D. Pogosyan: Astrophys. J. **537**, 720 (2000)
53. A. Lazarian, E. T. Vishniac: Astrophys. J. **517**, 700 (1999)
54. A. Lazarian, E. T. Vishniac: Revista Mexicana de Astronomia y Astrofisica **9**, 55 (2000)
55. A. Lazarian, H. Yan: Astrophys. J. **566**, L105, (2002)
56. A. Lazarian, E. T. Vishniac, J. Cho: in preparation (2002)
57. R. J. Leamon, C. W. Smith, N. F. Ness, W. H. Matthaeus: J. Geophys. Res. **103**, 4775 (1998)
58. Z. W. Ma, A. Bhattacharjee: J. Geophys. Res. **101**, 2641 (1996)
59. W. H. Matthaeus, S. L. Lamkin: Phys. Fluids **28**, 303 (1985)
60. M. Meneguzzi, U. Frisch, A. Pouquet: Phys. Rev. Lett. **47**, 1060 (1981)
61. K. A. Miller, J. M. Stone: Astrophys. J. **534**, 398 (2000)
62. H. K. Moffatt: *Magnetic Field Generation in Electrically Conducting Fluids*, (Cambridge: Cambridge University Press 1978)
63. K. Naidu, J. F. McKenzie, W. I. Axford: Ann. Geophysicae **10**, 827 (1992)
64. E. N. Parker: J. Geophys. Res. **62**, 509 (1957)
65. E. N. Parker:*Cosmical Magnetic Fields* (Oxford: Clarendon Press 1979)
66. E. N. Parker: Astrophys. J. **401**, 137 (1992)
67. H. E. Petschek: 'Magnetic Field Annihilation' In: *The Physics of Solar Flares*, ed. by W.H. Hess (Washington, DC, NASA Special Publications 50) pp. 425-439
68. E. Priest, T. Forbes: *Magnetic Reconnection: MHD Theory and Applications* (Cambridge: Cambridge University Press 2000)
69. M. Rees: Q. J. Roy. Astr. Soc. **28**, 197 (1988)
70. M. A. Shay, J. F. Drake: Geophys. Res. Lett. **25(20)**, 3759 (1998)
71. M. A. Shay, J. F. Drake, R. E. Denton, D. Biskamp: J. Geophys. Res. **103**, 9165 (1998)
72. M. A. Shay, J. F. Drake, B. N. Rogers, R. E. Denton: Geophys. Res. Lett. **26(14)**, 2163 (1999)
73. J. M. Stone, J. F. Hawley, C. F. Gammie, S. A. Balbus: Astrophys. J. **463**, 656 (1996)

74. S. Stanimirovic, A. Lazarian: Astrophys. J. Lett. **551**, 53 (2002)
75. H. R. Strauss: Astrophys. J. **326**, 412 (1988)
76. W. T. Stribling, W. H. Matthaeus, S. Ghosh: J. Geophys. Res. **99**, 2567 (1994)
77. W. T. Stribling, W. H. Matthaeus, S. Oughton: Phys. Plasmas **2**, 1437 (1995)
78. P. A. Sweet: 'The Neutral Point Theory of Solar Flares'. In: *IAU Symp. 6, Electromagnetic Phenomena in Cosmical Plasma*. ed. by B. Lehnert (New York: Cambridge Univ. Press 1958) pp.123-134
79. J. B. Taylor: Phys. Rev. Lett. **33**, 1139 (1974)
80. S. I. Vainshtein, E. N. Parker, R. Rosner: Astrophys. J. **404**, 773 (1993)
81. E. P. Velikhov: Sov. Phys. JETP Lett. **35**, 1398 (1959)
82. E. T. Vishniac, A. Brandenburg: Astrophys. J. **475**, 263 (1997)
83. E. T. Vishniac, J. Cho: Astrophys. J. **550**, 752 (2001)
84. E. T. Vishniac, A. Lazarian: Astrophys. J. **511**, 193 (1999)
85. X. Wang, Z. W. Ma, A. Bhattacharjee: Phys. Plasmas **3(5)**, 2129 (1996)
86. M. Yamada, H. Ji, S. Hsu, T. Carter, R. Kulsrud, F. Trintchouk: Phys. Plasmas **7(5)**, 1781 (2000)
87. E. Zweibel: Phys. Plasmas **5**, 247 (1998)
88. E. Zweibel, A. Brandenburg: Astrophys. J. **478**, 563 (1997)

The Helicity Issue in Large Scale Dynamos

Axel Brandenburg

NORDITA, Blegdamsvej 17, DK-2100 Copenhagen Ø, Denmark

Abstract. The connection between helically isotropic MHD turbulence and mean-field dynamo theory is reviewed. The nonlinearity in the mean-field theory is not yet well established, but detailed comparison with simulations begin to help select viable forms of the nonlinearity. The crucial discriminant is the magnetic helicity, which is known to evolve only on a slow resistive time scale in the limit of large magnetic Reynolds number. Particular emphasis is put on the possibility of memory effects, which means that an additional explicitly time-dependent equation for the nonlinearity is solved simultaneously with the mean-field equations. This approach leads to better agreement with the simulations, while it would also produce more favorable agreement between models and stellar dynamos.

1 Introduction

In an early paper Parker [1] identified cyclonic convection as a key process for converting large scale toroidal magnetic field into poloidal fields that have coherence over about half a hemisphere. This process is now generally referred to as the α-effect, although it may arise not only from thermal buoyancy [2], but also from magnetic buoyancy [3,4,5], the magneto-rotational instability [6,7], or some other magnetic instability [8]. In each case the effect of the Coriolis force together with some kind of radial stratification is crucial for making the motions helical [9]. Upward moving fluid expands, and the Coriolis force makes it rotate retrograde, causing negative (positive) kinetic helicity in the northern (southern) hemisphere. Downward moving fluid contracts, rotates in the prograde direction and contributes in the same sense to negative (positive) kinetic helicity on the northern (southern) hemisphere. This causes a positive α-effect in the northern hemisphere, but if magnetic stresses and shear become strong (for example in accretion discs) the sign may reverse [10,11].

When combined with differential rotation, the main outcome of α-effect models is the possibility of cyclic magnetic fields associated with latitudinal migration. The first global (two-dimensional) models were presented by Steenbeck & Krause [12], but similar models, with different physics, are still being studied today [13,14,15]. The migratory behavior is best seen in contours of the longitudinally averaged mean magnetic field versus latitude and time, which should show tilted structures converging to the equator. Such plots can be compared with the solar butterfly diagram of sunspot numbers (so called because the structures resemble a sequence of butterflies).

E. Falgarone and T. Passot (Eds.): LNP 614, pp. 402–431, 2003.
© Springer-Verlag Berlin Heidelberg 2003

A key property of all these models is that not only the motions are helical, but the large scale magnetic field itself is also helical. Of course, not all dynamos require helicity, but nonhelical dynamos tend to generate preferentially small-scale fields [16]. In a recent attempt, Vishniac & Cho [17] proposed a mechanism relevant in particular to accretion discs where shear is strong. Their mechanism would not lead to the production of net magnetic helicity, but numerical simulations [18] failed so far to show convincingly *large scale* dynamo action based on the proposed mechanism. Shear does produce large scale fields, but only in the toroidal direction. It does not explain the latitudinal coherence of the field over several tens of degrees (corresponding to several hundred megameters). On the other hand, there is direct observational evidence that the solar magnetic field is indeed helical. (We shall return to observations in Sect. 2.)

The trouble with helical fields is that magnetic helicity is conserved by the induction equation in the ideal limit and can only change on a resistive time scale, provided there is no significant loss through boundaries (at the surface or the equator, for example). This approximate magnetic helicity conservation leads to magnetic field saturation on a resistive time scale [19]. Depending on how effectively the boundaries transmit magnetic energy and helicity, the final saturation amplitude will be lowered if losses occur preferentially on large scales while the (linear) growth rate of magnetic energy (past initial saturation) remains *unchanged* [20]. In this sense final saturation can be achieved earlier. The above results are particularly clear when the flow is nearly fully helical, i.e. when the normalized kinetic helicity, $\epsilon_{\rm f} \equiv \langle \boldsymbol{\omega} \cdot \mathbf{u} \rangle / (\omega_{\rm rms} u_{\rm rms})$, where $\boldsymbol{\omega} = \boldsymbol{\nabla} \times \mathbf{u}$ is the vorticity, is large. In the sun, and probably in all other celestial bodies with rotating turbulence, the relative kinetic helicity is small; $\epsilon_{\rm f} \sim 5\%$. It is however this small helical fraction of the turbulence that is responsible for the a finite but small α-effect, so a proper understanding of its dynamics is crucial if one wants to build models based on the α-effect. Below we shall also discuss the alternative that astrophysical dynamos may shed preferentially small-scale helical fields through the boundaries. This could theoretically enhance large-scale dynamo action [21,22,23].

2 Magnetic Helicity Production

Before we begin discussing the magnetic helicity problem and possible remedies we need to be sure that the solar dynamo is indeed likely to involve significant amounts of magnetic helicity. There is direct observational evidence that the field of the sun is actually helical. Firstly, active regions are known to have systematically different signs of current helicity in the two hemispheres [24,25,26,27]; preferentially negative (positive) in the northern (southern) hemisphere. Secondly, magnetic helicity flux from the solar surface has also been inferred and this confirms the same sign as that of the current helicity. The magnetic helicity flux driven by the surface differential rotation has been estimated by Berger & Ruzmaikin [28], who find a total magnetic helicity flux on the order of $4 \times 10^{46}\,{\rm Mx}^2$ over the 22 year solar cycle. Similar values were also found by DeVore [29].

Finally, Chae [30] estimated the magnetic helicity flux based on counting the crossings of pairs of flux tubes. Combined with the assumption that two nearly aligned flux tubes are nearly parallel (rather than antiparallel) his results again suggest that the magnetic helicity is preferentially negative (positive) in the northern (southern) hemisphere.

The magnetic helicity is noisy, i.e. the sign can fluctuate and has only on average systematic behavior. This reflects the fact that only a fraction of the turbulence is helical. Thus, detailed understanding and measurements of the departures from systematic behavior is just as important as understanding and recording the systematic behavior.

In astrophysical flows, kinetic helicity can be generated in rotating stratified turbulence [9]. Such flows are intrinsically anisotropic. Whilst this is not a problem for numerical simulations, it definitely complicates the theoretical understanding and one should not be surprised if some fundamental aspects of mean-field theory (e.g. α proportional to $-\langle\boldsymbol{\omega}\cdot\mathbf{u}\rangle$) are not recovered. We just mention that under certain conditions, stratified rotating flows exhibit an α-tensor whose vertical (z-z) component has the opposite sign as the horizontal (x-x and y-y) components [32,33,34,35].)

In the following we concentrate on the isotropic helical aspects of the turbulence. This is accomplished by driving the flow with random polarized waves in a periodic domain. In most of the cases we use fully helical turbulence, but in many estimates the fraction of helicity enters only as an additional scaling factor [23]. The main goal here is a better understanding of the α-effect and the nonlinear feedback when the field becomes dynamically important. We therefore discuss in detail the perhaps simplest possible system: the α^2-dynamo in a periodic box.

In spherical geometry, the term α^2-dynamo refers to the fact that both large-scale poloidal and toroidal fields are maintained against ohmic decay by the α-effect. By contrast, the $\alpha\Omega$-dynamo is one where the large-scale toroidal field is generated mostly by differential rotation (the Ω-effect) and the α-effect can be neglected by comparison. If the α-effect is not neglected one speaks of an $\alpha^2\Omega$-dynamo. We stress that the α^2-dynamo has nothing to do with the so-called small-scale dynamo. These are turbulent dynamos operating only on scales less than the energy-carrying scale of the turbulence. They are quite common if the flows are non-helical. By contrast, both α^2 and $\alpha\Omega$-dynamos also generate fields on large scales, but they are necessarily accompanied by some level of small scale fields as well.

In its simplest form the α^2-dynamo equations for isotropic α and turbulent diffusivity $\eta_{\rm t}$ can be written as

$$\frac{\partial\overline{\mathbf{B}}}{\partial t}=\boldsymbol{\nabla}\times\left(\alpha\overline{\mathbf{B}}-\eta_{\rm T}\mu_0\overline{\mathbf{J}}\right), \tag{1}$$

where $\overline{\mathbf{J}}=\boldsymbol{\nabla}\times\overline{\mathbf{B}}/\mu_0$ is the mean current density and μ_0 is the magnetic permeability. This equation permits plane wave solutions of the form $\overline{\mathbf{B}}=\hat{\mathbf{B}}\exp(\lambda t+{\rm i}\boldsymbol{k}\cdot\boldsymbol{x})$, with the dispersion relation $\lambda=\pm|\alpha|k-\eta_{\rm T}k^2$, where $k=|\boldsymbol{k}|$.

The maximum of λ is where $\mathrm{d}\lambda/\mathrm{d}k = 0$, which yields

$$k = k_{\max} = \alpha/(2\eta_{\mathrm{T}}). \tag{2}$$

For a periodic box of size L^3, the most easily excited mode has $k = k_1 \equiv 2\pi/L$, and the $\boldsymbol{k}$ vector can point in any of the three coordinate directions. For $\alpha < 0$ (the case considered in [19]), the three possible eigenfunctions are

$$\mathbf{B}^{(z)} = B_0 \begin{pmatrix} \cos k_1 z \\ \sin k_1 z \\ 0 \end{pmatrix}, \quad \mathbf{B}^{(x)} = B_0 \begin{pmatrix} 0 \\ \cos k_1 x \\ \sin k_1 x \end{pmatrix}, \quad \mathbf{B}^{(y)} = B_0 \begin{pmatrix} \sin k_1 y \\ 0 \\ \cos k_1 y \end{pmatrix}, \tag{3}$$

where we have ignored arbitrary phase shifts in any of the three directions. All three solutions have been found in the simulations [19].

In the simulations there is of course no explicit α-effect in the usual sense, because we just solve the primitive MHD equations. The turbulence does, however, display *collective behavior* – just as it is expected based on mean-field α^2-dynamo theory, as explained above. We begin with the simulations.

3 Helical Turbulence: Prototype of an α^2-Dynamo

We consider a compressible isothermal gas with constant sound speed c_{s}, constant dynamical viscosity μ, and constant magnetic diffusivity η. To make sure the magnetic field stays solenoidal, i.e. $\boldsymbol{\nabla} \cdot \mathbf{B} = 0$, we express $\mathbf{B}$ in terms of the magnetic vector potential $\mathbf{A}$, so the field is written as $\mathbf{B} = \boldsymbol{\nabla} \times \mathbf{A}$. The governing equations for density ρ, velocity $\mathbf{u}$, and magnetic vector potential $\mathbf{A}$, are given by

$$\frac{\mathrm{D} \ln \rho}{\mathrm{D}t} = -\boldsymbol{\nabla} \cdot \mathbf{u}, \tag{4}$$

$$\frac{\mathrm{D}\mathbf{u}}{\mathrm{D}t} = -c_{\mathrm{s}}^2 \boldsymbol{\nabla} \ln \rho + \frac{\mathbf{J} \times \mathbf{B}}{\rho} + \frac{\mu}{\rho}(\nabla^2 \mathbf{u} + \tfrac{1}{3} \boldsymbol{\nabla} \boldsymbol{\nabla} \cdot \mathbf{u}) + \boldsymbol{f}, \tag{5}$$

$$\frac{\partial \mathbf{A}}{\partial t} = \mathbf{u} \times \mathbf{B} - \eta \mu_0 \mathbf{J} - \boldsymbol{\nabla} \phi, \tag{6}$$

where $\mathrm{D}/\mathrm{D}t = \partial/\partial t + \mathbf{u} \cdot \boldsymbol{\nabla}$ is the advective derivative. The current density, $\mathbf{J} = \boldsymbol{\nabla} \times \mathbf{B}/\mu_0$, is obtained in the form $\mu_0 \mathbf{J} = -\nabla^2 \mathbf{A} + \boldsymbol{\nabla} \boldsymbol{\nabla} \cdot \mathbf{A}$. We often use $\phi = 0$ as a convenient gauge for the electrostatic potential. Other frequent choices are $\phi = -\eta \boldsymbol{\nabla} \cdot \mathbf{A}$, $\phi = \mathbf{u} \cdot \mathbf{A}$, or combinations of these [36]. The Coulomb gauge, $\boldsymbol{\nabla} \cdot \mathbf{A} = 0$, corresponds to $\phi = -\boldsymbol{\nabla} \cdot \mathbf{E}$, where $\mathbf{E} = -\mathbf{u} \times \mathbf{B} + \eta \mu_0 \mathbf{J}$, but the original reason for solving for $\mathbf{A}$ instead of $\mathbf{B}$ was just to get rid of the solenoidality condition, so the Coulomb condition has computational disadvantages.

For the following it is useful to recall that each vector field can be decomposed into a solenoidal and two vortical parts with positive and negative helicity, respectively. These are also referred to as Chandrasekhar-Kendall functions. Although it is often useful to decompose the magnetic field into positive and negative helical parts, we also use the helical fields (with positive helicity) as forcing

function $\boldsymbol{f}$ of the flow. We restrict ourselves to functions selected from a finite band of wavenumbers around the wavenumber $k_{\rm f}$, but direction and amplitude are chosen randomly at each timestep. Details can be found in [19]. Similar work was first carried out by Meneguzzi et al. [37], but at the time one was barely able to run even until saturation. Throughout the nineties, work has been done on forced ABC flows [38,39,40]. In none of these investigations, however, the issue of resistively slow magnetic helicity evolution past initial saturation has been noted. It is exactly this aspect that has now become so crucial in understanding the saturation behavior of nonlinear dynamos. We begin by discussing first the linear (kinematic) evolution of the magnetic field.

3.1 Linear Behavior

If the magnetic Reynolds number, defined here as $R_{\rm m} = u_{\rm rms}/(\eta k_{\rm f})$, exceeds a certain critical value, $R_{\rm m}^{\rm (crit)}$, there is dynamo action. For helical flows, the so defined $R_{\rm m}^{\rm (crit)}$ is between 1 and 2 (see Table 1 of [19]). In the supercritical case, $R_{\rm m} > R_{\rm m}^{\rm (crit)}$, the field grows exponentially with growth rate λ, which scales with the inverse turnover time, $u_{\rm rms}/k_{\rm f}$. The resistively limited saturation behavior that will be discussed below in full detail has no obvious correspondence in the kinematic stage when the field is too weak to affect the motions by the Lorentz force [41]. Nevertheless, there is actually a subtle effect on the shape of the eigenfunction as $R_{\rm m}$ increases. Before we can appreciate this, we need to discuss the effect the kinetic helicity has on the field.

A helical velocity tends to drive helicity in the magnetic field as well, but in the nonresistive limit magnetic helicity conservation dictates that $\langle \mathbf{A} \cdot \mathbf{B} \rangle = {\rm const} = 0$ if the initial field (or at least its helicity) was infinitesimally weak. Thus, there must be some kind of magnetic helicity cancelation. Under homogeneous isotropic conditions there cannot be a spatial segregation in positive and negative helical parts. Instead, there is a spectral segregation: there is a bump at the forcing wavenumber and another 'secondary' bump at somewhat smaller wavenumber. The two bumps have opposite sign of magnetic helicity such that the net magnetic helicity is zero. At the forcing wavenumber, the sign of magnetic helicity agrees with that of the kinetic helicity, but at smaller wavenumber the sign of magnetic helicity is opposite. At small $R_{\rm m}$, this secondary peak can be identified with the wavenumber where the corresponding α^2-dynamo has maximum growth rate; see (2). Simulations show that as $R_{\rm m}$ increases, $k_{\rm max}$ approaches $\frac{1}{2}k_{\rm f}$ [23]. This agrees qualitatively with earlier results [41,42] which suggested that the magnetic helicity approaches zero in the high-$R_{\rm m}$ limit.

3.2 Nonlinear Behavior

Eventually, the magnetic energy stops increasing exponentially. This is due to the nonlinear terms, in particular the Lorentz force $\mathbf{J} \times \mathbf{B}$ in (5), which begins to affect the velocity field. The temporal growth of the power spectra saturates, but only partially; see Fig. 1, where we show data from a run with forcing at

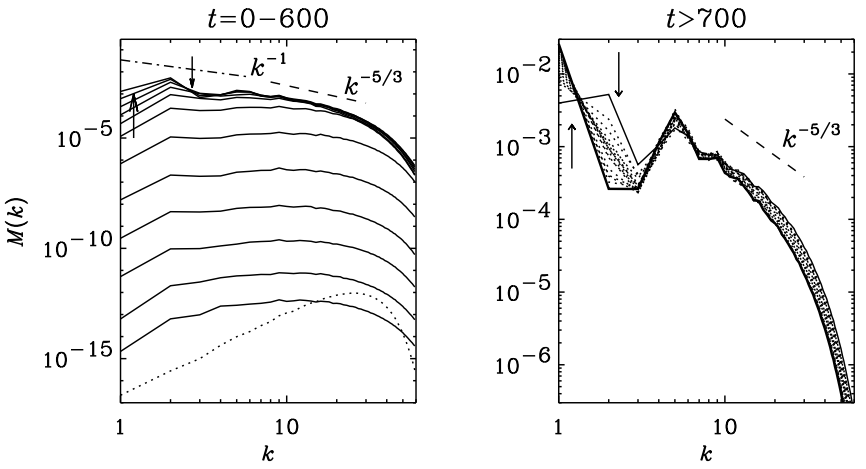


Fig. 1. Power spectra of magnetic energy of Run 3 of [19]. During the initial growth phase the field saturates at small scales first and only later at large scales (left hand panel). Later, when also the large scale field saturates, the field at intermediate scales ($k = 2$, 3, and 4) becomes suppressed. In the second panel, intermediate times are shown as dotted lines, $t = 700$ is shown in solid and $t = 1600$ is shown as a thick solid line. The forcing wavenumber is $k_\mathrm{f} = 5$.

wavenumber $k_\mathrm{f} = 5$. In the left hand panel we see that by the time $t = 600$ the power spectra have saturated at larger wavenumbers, $k \gtrsim 3$. It takes until $t \simeq 1600$ for the power spectra to be saturated also at $k = 1$ (right hand panel of Fig. 1). In order to see more clearly the behavior at large scales, we show in Fig. 2 data from a run with $k_\mathrm{f} = 27$ and compare spectra in the linear and nonlinear regimes. In the linear regime, all spectra are just shifted along the ordinate, so the spectra have been compensated by the factor $M_\mathrm{ini} \exp(2\lambda t)$, where λ is the growth rate and M_ini the initial magnetic energy. In the nonlinear regime the bump on the right stays at approximately the same wavenumber (the forcing wavenumber), while the bump on the left propagates gradually further to the left. As it does so, and since the amplitude of the secondary peak even increases slightly, the net magnetic helicity inevitably increases (or rather becomes more negative in the present case). But because of the asymptotic magnetic helicity conservation, this can only happen on a slow resistive time scale. This leads to the appearance of a (resistively) slow saturation phase past the initial saturation; see Fig. 3.

3.3 The Final Saturation Value

In a periodic box with helically driven turbulence, the final saturation value of the magnetic field is determined by the ratio of the size of the domain to the scale

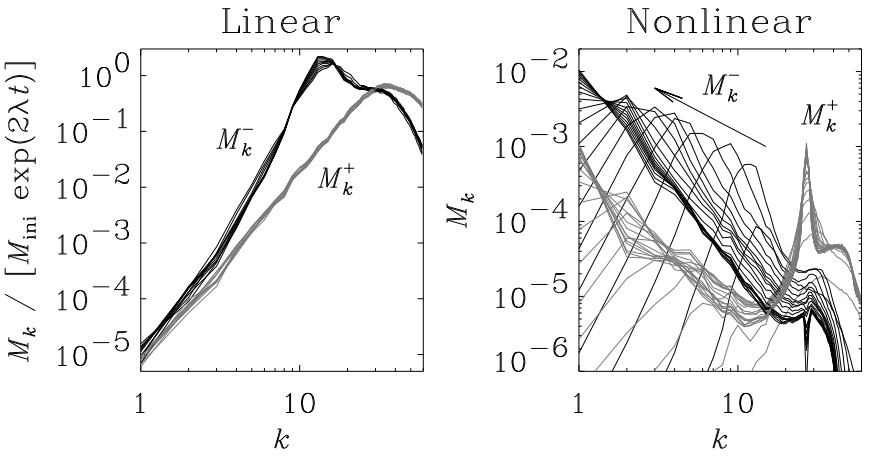


Fig. 2. Power spectra of magnetic energy of positively and negatively polarized parts (M_k^+ and M_k^-) in the linear and nonlinear regimes. The spectra in the linear regime have been compensated by the exponential growth factor to make them collapse on top of each other. Here the forcing wavenumber is in the dissipative subrange, $k_{\rm f} = 27$, but this allows enough scale separation to see the inverse transfer of magnetic energy to smaller k. The data are from Run B of [44].

of the forcing. This is best seen by considering the magnetic helicity equation

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle \mathbf{A}\cdot\mathbf{B}\rangle = -2\eta\mu_0\langle \mathbf{J}\cdot\mathbf{B}\rangle. \tag{7}$$

If $\langle \mathbf{A}\cdot\mathbf{B}\rangle$ were not gauge invariant (for example if there are open boundaries), (7) would be useless. In particular, $\langle \mathbf{A}\cdot\mathbf{B}\rangle$ will in general not be constant in the steady state (see Fig. 2 of [23], for an example). One therefore has to go to the gauge-invariant relative magnetic helicity of Berger & Field [43]. This has been done in [20,23]. In the present case of periodic boundaries, however, $\langle \mathbf{A}\cdot\mathbf{B}\rangle$ is automatically gauge invariant and therefore a physically meaningful quantity, so it must be constant in the steady state. Equation (7) says that then the current helicity, $\langle \mathbf{J}\cdot\mathbf{B}\rangle$, must vanish. At first glance this seems to be in conflict with the idea of building up a helical large scale field. The solution is that there must then be an equal amount of small scale current helicity so that

$$\langle \mathbf{J}\cdot\mathbf{B}\rangle = \langle \overline{\mathbf{J}}\cdot\overline{\mathbf{B}}\rangle + \langle \mathbf{j}\cdot\mathbf{b}\rangle = 0 \quad \text{(in the steady state)}. \tag{8}$$

If the field is fully helical, we have $\mu_0\langle \overline{\mathbf{J}}\cdot\overline{\mathbf{B}}\rangle = \mp k_1\langle \overline{\mathbf{B}}^2\rangle$ and $\mu_0\langle \mathbf{j}\cdot\mathbf{b}\rangle = \pm k_{\rm f}\langle \mathbf{b}^2\rangle$, where the upper (lower) sign applies to the case where the small scale helicity at the forcing scale is positive (negative). We then have from (8)

$$\langle \overline{\mathbf{B}}^2\rangle = (k_{\rm f}/k_1)\,\langle \mathbf{b}^2\rangle. \tag{9}$$

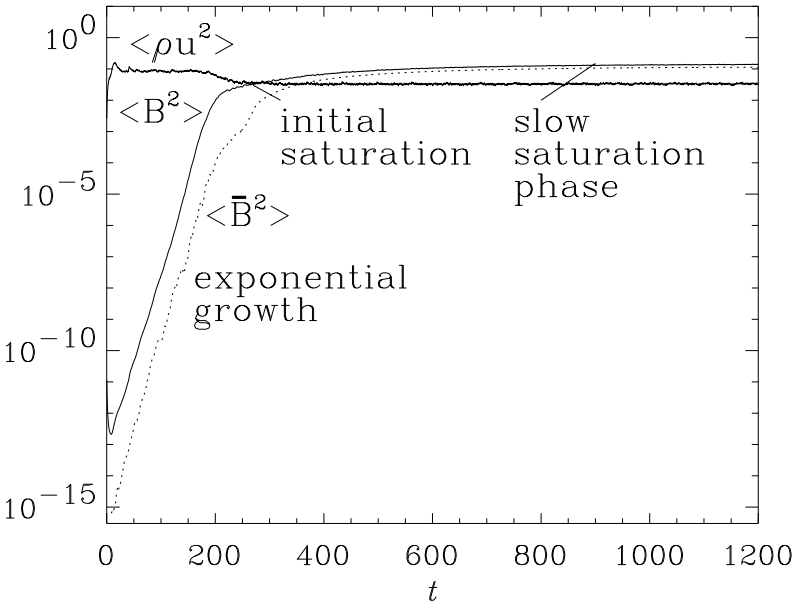


Fig. 3. The three stages of the magnetic field growth: exponential growth until initial saturation (when $\langle \mathbf{B}^2 \rangle / \mu_0 = \langle \rho \mathbf{u}^2 \rangle$), followed by a (resistively) slow saturation phase. In this plot we have used $\mu_0 = 1$. The energy of the large scale magnetic field, $\langle \overline{\mathbf{B}}^2 \rangle$, is shown for comparison. The data are from Run 3 of [19].

To a good approximation, $\langle \mathbf{b}^2 \rangle^{1/2}$ will be close to the equipartition field strength, $B_{\rm eq}$, so

$$\langle \overline{\mathbf{B}}^2 \rangle / B_{\rm eq}^2 \approx k_{\rm f}/k_1 > 1. \tag{10}$$

This means that the energy of the large scale field must, in the final state, be in super-equipartition by a factor approximately equal to the scale separation. We recall that this applies to the case of periodic boundaries. For closed (e.g. perfectly conducting) boundaries, $\langle \overline{\mathbf{B}}^2 \rangle$ can be even larger than $k_{\rm f}/k_1$ times the equipartition value [23]. This is because the large scale field is no longer fully helical, while the small scale field still is. In the presence of *open* boundaries, on the other hand, the large scale field will generally be smaller than suggested by (10), unless the boundaries transmit preferentially small scale fields (Sect. 4).

3.4 Sensitivity to Using Hyperdiffusivity

The above statements can readily be generalized to the case where the usual magnetic diffusion operator, $\eta \nabla^2$, is replaced by hyperdiffusion, $(-1)^{n-1} \eta_n \nabla^{2n}$ with $n = 2$. This implies that the diffusion has now become more strongly

wavenumber dependent; from ηk^2 to $\eta_2 k^4$. If the diffusion is the same at small scales, then the diffusion at large scales must be significantly smaller in the hyperdiffusive runs. This leads to a dramatic *increase* of the resistive saturation time. At the same time the final saturation field strength is no longer given by (10). The rate of magnetic helicity dissipation is now no longer proportional to k, but to k^3. Therefore the final saturation field strength is given by

$$\langle\overline{\mathbf{B}}^2\rangle/B_{\rm eq}^2 \approx (k_{\rm f}/k_1)^3 \gg 1. \tag{11}$$

This result was confirmed also numerically [44]. Again, this applies to periodic boundaries. Hyperdiffusion has been used in the past in connection with open boundaries [6,45], but it is not clear how serious the possible artifacts from hyperdiffusion would be in such cases with open boundaries.

3.5 The Magnetic Helicity Constraint

The case of periodic boundaries is particularly useful as a benchmark to all dynamos exhibiting large scale dynamo action due to the helicity effect. Here we discuss the functional form $\overline{\mathbf{B}}^2(t)$ for the late saturation phase.

Equation (8) allows us not only to determine the final saturation strength, but also the approximate time evolution near saturation. As before, we make the assumption of fully helical fields. However, given that prior to saturation $|\langle\mathbf{j}\cdot\mathbf{b}\rangle| \approx |\langle\overline{\mathbf{J}}\cdot\overline{\mathbf{B}}\rangle|$, we must have $|\langle\mathbf{a}\cdot\mathbf{b}\rangle| \ll |\langle\overline{\mathbf{A}}\cdot\overline{\mathbf{B}}\rangle|$, so we can set

$$\langle\mathbf{A}\cdot\mathbf{B}\rangle \approx \langle\overline{\mathbf{A}}\cdot\overline{\mathbf{B}}\rangle = \mp k_1^{-1}\langle\overline{\mathbf{B}}^2\rangle, \tag{12}$$

where the upper (lower) sign refers to positive (negative) small scale (kinetic and magnetic) helicity. Inserting this into (7) we have

$$k_1^{-1}\frac{\mathrm{d}}{\mathrm{d}t}\langle\overline{\mathbf{B}}^2\rangle = -2\eta k_1\langle\overline{\mathbf{B}}^2\rangle + 2\eta k_{\rm f}\langle\mathbf{b}^2\rangle. \tag{13}$$

The small scale field saturates first, so (13) can then be integrated to get the *subsequent* evolution of $\langle\overline{\mathbf{B}}^2\rangle$ toward full saturation [19]

$$\langle\overline{\mathbf{B}}^2\rangle = \frac{k_{\rm f}}{k_1}\langle\mathbf{b}^2\rangle\left[1-e^{-2\eta k_1^2(t-t_{\rm sat})}\right] \quad (\text{for } t>t_{\rm sat}), \tag{14}$$

where $t_{\rm sat}$ is the time when the small scale field has reached saturation. Equation (14) is what we usually mean by the *magnetic helicity constraint*.

3.6 Inverse Cascade versus α-Effect

The process outlined above can be interpreted in two different ways: inverse cascade of magnetic helicity and/or α-effect. The two are similar in that they tend to produce magnetic energy at scales larger than the energy-carrying scale of the turbulence. As can be seen from Figs 1 and 2, the present simulations

support the notion of *nonlocal* inverse transfer [19]. This is not really an inverse cascade in the usual sense, because there is no sustained flux of energy through wavenumber space as in the direct Kolmogorov cascade. Instead, there is just a bump traveling to smaller k in wavenumber space. In that respect, the present simulations seem to differ from the Eddy Damped Quasi-Normal Markovian (EDQNM) closure approximation [46].

The other interpretation is in terms of the α-effect. We recall that there is a wavenumber $k_{\max}$ where the growth of the large scale field is largest; see (2). For reasonable estimates, $k_{\max}$ coincides with the position of the secondary bump in the spectrum ([19], Sect. 3.5). This can be taken as evidence in favor of the α-effect. In the nonlinear regime, the secondary bump travels to the left in the spectrum (i.e. toward smaller k). In the EDQNM picture this has to do with the equilibration of kinetic and current helicities at progressively smaller wavenumbers, which then saturates further growth at that wavenumber, but permits further growth at smaller wavenumbers until equilibration occurs. Another interpretation is simply that if α is quenched to a smaller value, $k_{\max} = \alpha/(2\eta_{\rm T})$ peaks at smaller wavenumbers where the growth has not yet saturated, until equilibration is attained also at that scale.

3.7 Implications

The growth of the large scale magnetic field can be interpreted as being due to the α-effect. Consequently, a slow-down in the final saturation phase must have to do with a suppression of α. According to closure models [46], the α-effect is really the residual between two competing effects: a kinematic helicity effect (which itself decreases somewhat for strong magnetic fields), and a current helicity effect of opposite sign. In Sect. 7 we shall use this phenomenology in a mean-field model where an explicitly time-dependent equation for the current helicity is solved. It turns out that in this model the late saturation phase is resistively limited–just like in the simulations.

It is somewhat worrisome that in the nonlinear regime the value of α depends on the microscopic magnetic diffusivity, which is very small in most astrophysical situations. One might therefore be concerned that in astrophysical dynamos the saturation would be unacceptably slow. In order to say more about α and also the turbulent magnetic diffusivity $\eta_{\rm t}$, we need to determine how α and $\eta_{\rm t}$ depend on $\mathbf{B}$. This can either be done directly [47], which is difficult and the results are noisy, or we can compare with models that incorporate the effects of α and $\eta_{\rm t}$ quenching. Before we do this (Sects 6 and 7), we first want to assess the effects of boundaries.

4 Open Boundaries: Good or Bad?

Boundaries generally lead to a loss of magnetic field both on small and large scales. Losses at large scale tend to lower the saturation field strength of the mean field. Losses of small scale fields can, at least in principle, enhance the

large scale field [21,22]. This has been demonstrated in an idealized numerical experiment [23], where the magnetic field at the forcing wavenumber and beyond had been removed in regular time intervals; see Fig. 4. We discuss this now in more detail.

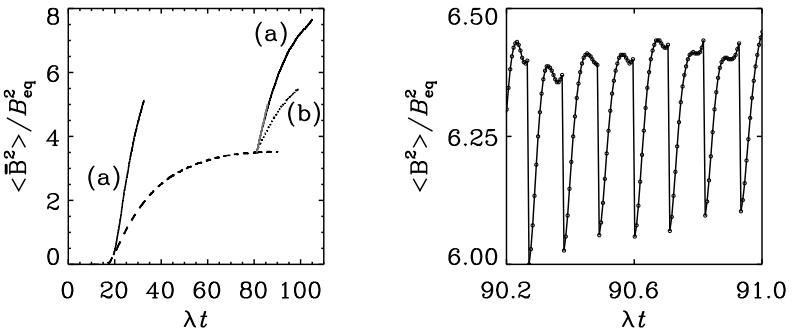


Fig. 4. The effect of removing small scale magnetic energy in regular time intervals Δt on the evolution of the large scale field (solid lines). The dashed line gives the evolution of $\langle\overline{\mathbf{B}}^2\rangle$ for Run 3 of [19] (where no such energy removal was included) in units of $B_{\rm eq}^2=\mu_0\rho_0\langle\mathbf{u}^2\rangle$. The two solid lines show the evolution of $\langle\overline{\mathbf{B}}^2\rangle$ after restarting the simulation from Run 3 of [19] at $\lambda t=20$ and $\lambda t=80$. Time is scaled with the kinematic growth rate λ. The curves labeled (a) give the result for $\Delta t=0.12\lambda^{-1}$ and those labeled (b) for $\Delta t=0.4\lambda^{-1}$. The second panel shows, for a short time interval, the sudden drop and subsequent recovery of the total (small and large scale) magnetic energy in regular time intervals. (Adapted from [23].)

4.1 Enhancement through Losses at Small Scales

The enhancement of large scale field by losses at small scales may seem somewhat mysterious. One way to interpret this result is by saying that the slow growth occurred because the energy of the small scale magnetic field has already reached the level of the kinetic energy, and only the large scale field has not yet saturated. After small scale magnetic fields have been removed, the field is for a short time interval in sub-equipartition at small scales and so the overall field (both at small and large scales) can then grow further during the short time interval during which the small scale field has not yet fully recovered to the equipartition value. The effect of a single such event is small, but many such events can produce a significant effect. This is exactly what is seen. Another way of interpreting this result is in terms of mean-field theory where the α-effect is saturated by a cancelation of kinetic and current helicities. If small scale magnetic fields are removed, the residual α-effect can be larger for some time interval, which then allows the field to grow further. In the following we illuminate this

result further by considering a modified magnetic helicity constraint for the case of open boundaries.

4.2 The Modified Magnetic Helicity Constraint

In Sect. 3.5 we have discussed an equation for the evolution of the magnetic energy of the large scale field at late times. Here we have assumed that there is no loss of magnetic energy and magnetic helicity through the boundaries. This equation has been generalized to account for losses of *large scale* magnetic helicity [20,23]. The idea is that there will be a magnetic helicity flux that is proportional to the gradient of the large scale magnetic helicity density (in a fixed gauge), and hence to the gradient of the magnetic energy density. This gives rise to an extra diffusion term, and hence to a renormalized, *effective* magnetic diffusivity, $\eta_{\rm eff}^{(1)}$, i.e. the term $2\eta k_1\langle\overline{\mathbf{B}}^2\rangle$ has to be replaced by $2\eta_{\rm eff}^{(1)}k_1\langle\overline{\mathbf{B}}^2\rangle$. Therefore, (13) takes the form [23]

$$k_1^{-1}\frac{\mathrm{d}}{\mathrm{d}t}\langle\overline{\mathbf{B}}^2\rangle = -2\eta_{\rm eff}^{(1)}k_1\langle\overline{\mathbf{B}}^2\rangle + 2\eta k_{\rm f}\langle\mathbf{b}^2\rangle, \tag{15}$$

which has the solution

$$\langle\overline{\mathbf{B}}^2\rangle = \frac{\eta k_{\rm f}}{\eta_{\rm eff}^{(1)}k_1}\langle\mathbf{b}^2\rangle\left[1-e^{-2\eta_{\rm eff}^{(1)}k_1^2(t-t_{\rm sat})}\right] \quad (\text{for } t>t_{\rm sat}), \tag{16}$$

Note that the saturation amplitude is decreased by a factor $\eta/\eta_{\rm eff}^{(1)}$ compared with (14), but at the same time the e-folding time has decreased to $(2\eta_{\rm eff}^{(1)}k_1^2)^{-1}$. We return to this behavior in the next subsection.

When the losses through the surface occur preferentially at small scales, an effective diffusivity would instead affect the small scale helicity flux. Therefore, (14) takes then the form

$$k_1^{-1}\frac{\mathrm{d}}{\mathrm{d}t}\langle\overline{\mathbf{B}}^2\rangle = -2\eta k_1\langle\overline{\mathbf{B}}^2\rangle + 2\eta_{\rm eff}^{(\rm f)}k_{\rm f}\langle\mathbf{b}^2\rangle, \tag{17}$$

which has the solution

$$\langle\overline{\mathbf{B}}^2\rangle = \frac{\eta_{\rm eff}^{(\rm f)}k_{\rm f}}{\eta k_1}\langle\mathbf{b}^2\rangle\left[1-e^{-2\eta k_1^2(t-t_{\rm sat})}\right] \quad (\text{for } t>t_{\rm sat}), \tag{18}$$

Note that the saturation amplitude is now increased by a factor $\eta_{\rm eff}^{(\rm f)}/\eta$ compared with (14), but the e-folding time, $(2\eta k_1^2)^{-1}$, is still resistively limited. This is in good agreement with what is seen in the simulations; see Fig. 4. In reality, there will be both small and large-scale losses, so the large scale magnetic energy is expected to evolve according to

$$\langle\overline{\mathbf{B}}^2\rangle = \frac{\eta_{\rm eff}^{(\rm f)}k_{\rm f}}{\eta_{\rm eff}^{(1)}k_1}\langle\mathbf{b}^2\rangle\left[1-e^{-2\eta_{\rm eff}^{(1)}k_1^2(t-t_{\rm sat})}\right] \quad (\text{for } t>t_{\rm sat}). \tag{19}$$

This equation allows time scales and saturation amplitudes that are not resistively limited.

4.3 Simulations with Open Boundaries

So far, simulations have not yet shown that the losses of small-scale magnetic fields are actually stronger than those of large-scale fields. Simulations with a vertical field (pseudo-vacuum) boundary condition have shown that most of the magnetic energy is lost at small scales [20]. The way this affects the slow resistively limited saturation process discussed earlier is by simply cutting off the saturation process at an earlier time, without changing the approximately linear slope past the initial saturation; cf. (16). In Fig. 5 we demonstrate a very similar behavior in another system which is actually periodic, but the helicity of the forcing is modulated in the z-direction such that the sign of the kinetic helicity changes in the middle. One can therefore view this system as two subsystems with a boundary in between. This boundary would correspond to the equator in a star or the midplane in a disc.

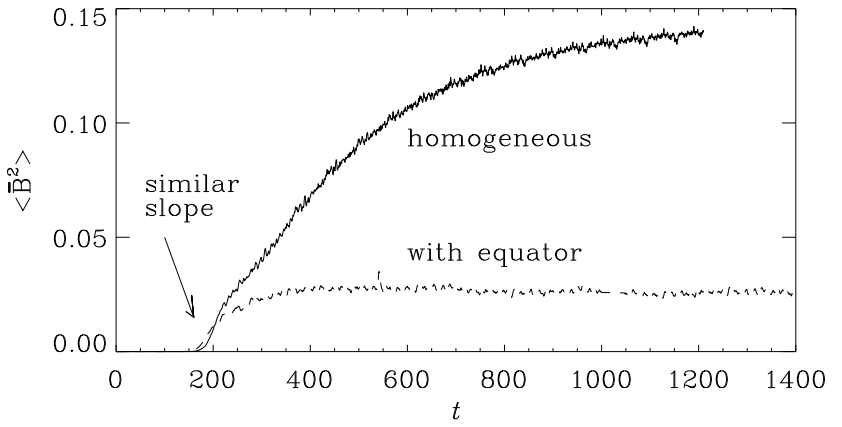


Fig. 5. Evolution of the magnetic energy for a run with homogeneous forcing function (solid line) and a forcing function whose helicity varies sinusoidally throughout the domain (dotted line) simulating the effects of equators at the two nodes of the sinusoidal helicity profile.

A somewhat surprising property of the models with variation of helicity in the z-direction is the fact that the mean field varies mostly in the x-direction, i.e. it follows the variation of the background model only weakly; see Fig. 6. Therefore, the mean field must be allowed to be two-dimensional, i.e.

$$\overline{\mathbf{B}}(x,z,t) = \int \mathbf{B} \mathrm{d}y \,/ \int \mathrm{d}y. \tag{20}$$

Similar behavior was also found in simulations with boundaries, especially when the aspect ratio was large [23]. In the present context we were able to confirm, using a two-dimensional mean field dynamo in periodic geometry, that for $\alpha \propto \sin k_1 z$ the most easily excited mode varies indeed both in x and z; see Fig. 7.

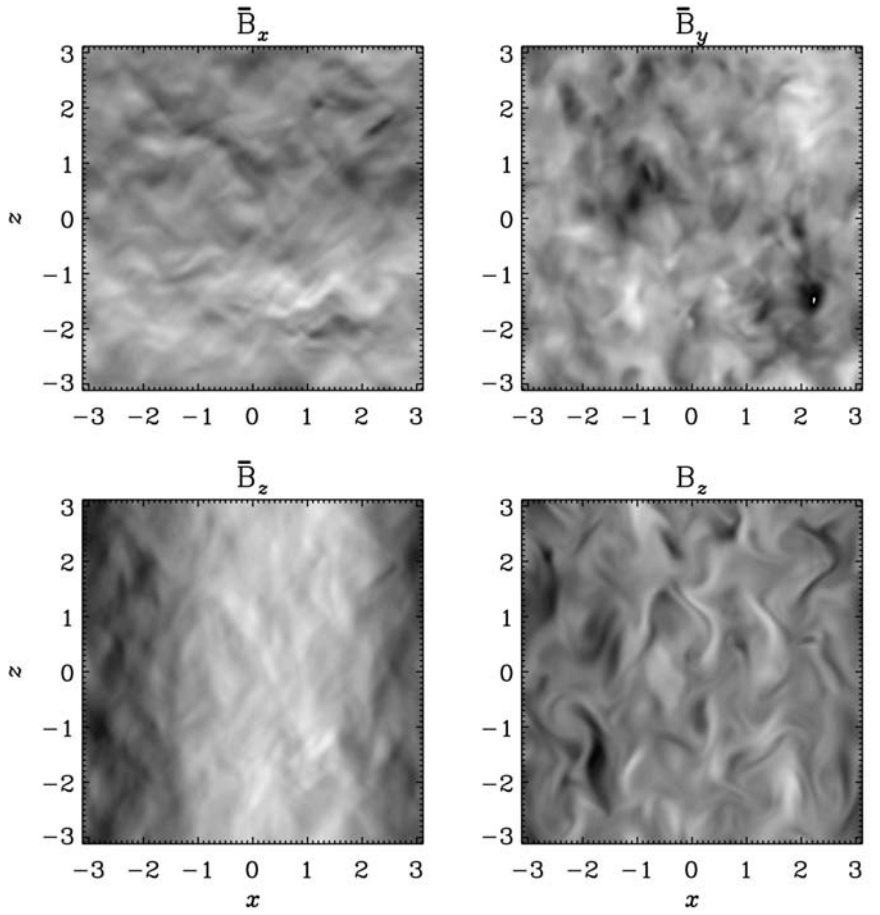


Fig. 6. Images of the three components of the mean field (averaged over the y-direction) for a run with sinusoidally varying helicity in the z-direction. Note that the most pronounced component of the mean field is actually $\overline{B}_z(x)$. The large-scale field is also visible in a y-slice of $\overline{B}_z$ (last panel).

In the simulations presented so far, boundaries merely tend to *reduce* the final saturation field strength. Thus, the idea to enhance the large scale field by small scale losses is not currently supported by simulations. It is quite possible, however, that this is simply a consequence of too simple a representation of the physical boundary. In the sun, coronal mass ejections are quite vigorous events that are known to shed large amounts of helical magnetic fields [28,29,30,31]. This kind of physics is not at all represented by adopting vacuum or pseudo-vacuum (vertical field) boundary conditions, as was done in [23].

We may then conclude that in simulations of large scale dynamos with relatively simple boundary conditions, open boundaries tend to be more important for large scale fields than for small scale fields. Although more realistic bound-

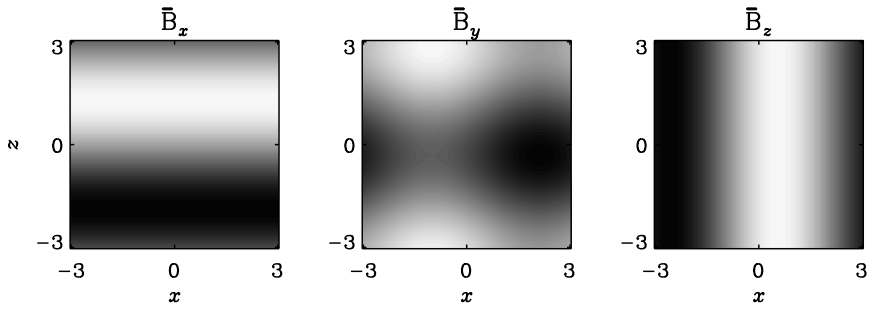


Fig. 7. Images of the three components of the mean field for an α^2 dynamo with sinusoidally varying α-effect in the z-direction. The $\overline{B}_x$ and $\overline{B}_z$ components resemble those in the direct simulation shown in Fig. 6.

ary conditions still need to be considered, it is useful to study more carefully whether, on observational grounds, a resistively limited dynamo can indeed be clearly excluded.

5 How Long Is Long? – Or What the Skin Depth Has to Do with the Solar Cycle

In this section we want to estimate the amount of magnetic helicity that is to be expected for a model of the solar dynamo. We also need to know which fraction of the magnetic field takes part in the 11-year cycle. Following an approach similar to that of Berger [48], we can bound the rate of change of magnetic helicity in terms of the rate of Joule dissipation, $Q_{\rm Joule}$, and magnetic energy, M. For an oscillatory dynamo, all three variables, H, M, and $Q_{\rm Joule}$ vary in an oscillatory fashion with a cycle frequency ω of magnetic energy (corresponding to 11 years for the sun – not 22 years), so we estimate $|{\rm d}H/{\rm d}t| \lesssim \omega|H|$ and $Q_{\rm Joule} \lesssim \omega M$, which leads to the inequality [23,44]

$$|H|/(2\mu_0 M) \leq \ell_{\rm skin}, \tag{21}$$

where $\ell_{\rm skin} = (2\eta/\omega)^{1/2}$ is the skin depth, here associated with the oscillation frequency ω. Thus, the maximum magnetic helicity that can be generated and dissipated during one cycle is characterized by the length scale $|H|/(2\mu_0 M)$, which has to be less than the skin depth $\ell_{\rm skin}$.

For η we have to use the Spitzer resistivity which is proportional to $T^{-3/2}$ (T is temperature), so η varies between $10^4\,{\rm cm}^2/{\rm s}$ at the base of the convection zone to about $10^7\,{\rm cm}^2/\,{\rm s}$ near the surface layers and decreases again in the solar atmosphere. Using $\omega = 2\pi/(11\,{\rm yr}) = 2\times10^{-8}\,{\rm s}^{-1}$ for the relevant frequency at which H and M vary we have $\ell_{\rm skin} \approx 10\,{\rm km}$ at the bottom of the convection zone and $\ell_{\rm skin} \approx 300\,{\rm km}$ at the top.

This needs to be compared with the value $|H|/(2\mu_0 M)$ obtained from dynamo models. Although mean-field theory has been around for several decades,

the helicity aspect has only recently attracted significant attention. In the proceedings of a meeting devoted specifically to this topic [49], magnetic helicity was discussed extensively also in the context of mean-field theory. However, the precise amount of magnetic helicity relative to the magnetic energy, and the possibility of helicity reversals at some length scale were not addressed at the time, although the evolution of the current helicity has already been investigated in the context of a mean field model [50].

For a sphere (or a half-sphere) with open boundary conditions and volume V (for example the northern hemisphere), one has to use the gauge-invariant relative magnetic helicity of Berger & Field [43],

$$H = \int_V (\mathbf{A} + \mathbf{A}_\mathrm{P}) \cdot (\mathbf{B} - \mathbf{B}_\mathrm{P}) \, \mathrm{d}V, \tag{22}$$

where $\mathbf{B}_\mathrm{P} = \boldsymbol{\nabla} \times \mathbf{A}_\mathrm{P}$ is a potential field used as reference field that has on the boundaries the same normal component as $\mathbf{B}$. Any additional gradient terms, $\boldsymbol{\nabla}\phi$, in $\mathbf{A}$ or $\mathbf{A}_\mathrm{P}$ yield only a surface term,

$$\int_{\partial V} \phi(\mathbf{B} - \mathbf{B}_\mathrm{P}) \cdot \mathrm{d}\mathbf{S}, \tag{23}$$

which vanishes because $\mathbf{B}_\mathrm{P} \cdot \boldsymbol{n} = \mathbf{B} \cdot \boldsymbol{n}$. When applied to an axisymmetric mean field, which can be written as $\overline{\mathbf{B}} = b\hat{\boldsymbol{\phi}} + \boldsymbol{\nabla} \times a\hat{\boldsymbol{\phi}}$, it turns out that the relative magnetic helicity integral is simply [23]

$$H = 2 \int_V ab \, \mathrm{d}V. \tag{24}$$

In order to see how the condition (21) is met by mean-field models, we have calculated a typical $\alpha\Omega$ model relevant to the sun and evaluated (24) over the volume of the northern hemisphere [23].

We recall that in the Babcock-Leighton approach it is mainly the latitudinal differential rotation that enters. We also note that, although the latitudinal migration could be explained by radial differential rotation, meridional circulation is in principle able to drive meridional migration even when the sense of radial differential rotation would otherwise be wrong for driving meridional migration [15,51,52]. Therefore, we have considered the simple models in spherical geometry. The results of such a model show that [23], once $B_\mathrm{pole}/B_\mathrm{belt}$ is in the range consistent with observations, $B_\mathrm{pole}/B_\mathrm{belt} = (1...3) \times 10^{-4}$, $H_\mathrm{N}/(2\mu_0 M_\mathrm{N} R)$ is around $(2-5) \times 10^{-4}$ for models with latitudinal shear. (Here, the subscript 'N' refers to the northern hemisphere.) This confirms the scaling

$$H_\mathrm{N}/(2\mu_0 M_\mathrm{N} R) = \mathcal{O}(B_\mathrm{pol}/B_\mathrm{tor}) \gtrsim B_\mathrm{pole}/B_\mathrm{belt}. \tag{25}$$

Given that $R = 700\,\mathrm{Mm}$ this means that $H_\mathrm{N}/(2\mu_0 M_\mathrm{N}) \approx 70...200\,\mathrm{km}$, which would be comparable with the value of ℓ_skin near the upper parts of the solar convection zone, or for models with only latitudinal shear.

The surprising conclusion is that the amount of mean field helicity that needs to be generated in order to explain the large scale solar magnetic fields is so small,

that it may be plausible that microscopic magnetic diffusion could still play a role in the solar dynamo. In other words, open boundary effects may well be important for understanding the time scale of the dynamo, but the effect does not need to be extremely strong.

6 Connection with α-Quenching

Nonlinear helical dynamos in a periodic domain are particularly simple. They provide therefore an ideal benchmark for models of α-quenching. Before applying models to more complicated situations, they should pass the test of predicting the right behavior in the simple case of an α^2-dynamo. By "right behavior" we mean that the large-scale magnetic field saturates as seen in (14). While this is already a relatively stringent test that allows us to eliminate some earlier models (see Sect. 6.2), we point out that the complete time evolution (including the early kinematic exponential growth phase) can only be described correctly by an explicitly time-dependent evolution equation for α that ensures that the magnetic helicity equation is obeyed exactly at all times.

6.1 The Lorentzian Quenching Formula

We first assume that both α-effect and the turbulent magnetic diffusivity, $\eta_{\rm t}$, are being affected by the magnetic field. In the first class of models we assume

$$\alpha = \alpha_0\, q(\overline{\mathbf{B}}), \quad \eta_{\rm t} = \eta_{\rm t0}\, q(\overline{\mathbf{B}}), \tag{26}$$

i.e. we postulate the existence of an algebraic quenching formula for both α and $\eta_{\rm t}$. The models that work best are those with a lorentzian quenching formula,

$$q(\overline{\mathbf{B}}) = \frac{1}{1 + a\overline{\mathbf{B}}^2/B_{\rm eq}^2} \quad \text{(lorentzian formula)}, \tag{27}$$

where a is a dimensionless coefficient and $B_{\rm eq}$ is the equipartition field strength with $B_{\rm eq}^2/\mu_0 = \langle\rho\mathbf{u}^2\rangle$.

For an α^2-dynamo in a periodic box, the coefficient a can be readily determined. This is because here we actually know the final saturation field strength; see (10). In a mean-field model, on the other hand, where quenching is the only nonlinearity, the saturation field strength is proportional to $a^{-1/2}$. In order to work out the coefficients, we consider the α^2 mean-field equation for a Beltrami field with wavenumber k_1. The evolution of the mean-squared field strength is then governed by [19]

$$\tfrac{1}{2}\frac{\rm d}{{\rm d}t}\ln\langle\overline{\mathbf{B}}^2\rangle = \frac{\alpha_0 k_1 - \eta_{\rm t0}k_1^2}{1 + a\langle\overline{\mathbf{B}}^2\rangle/B_{\rm eq}^2} - \eta k_1^2. \tag{28}$$

In the steady state, the right hand side of (28) has to vanish, so

$$\frac{\alpha_0 k_1 - \eta_{\rm t0}k_1^2}{1 + a\langle\overline{\mathbf{B}}^2\rangle/B_{\rm eq}^2} - \eta k_1^2 = 0, \tag{29}$$

or, using (10),

$$a = \frac{\alpha_0 - \eta_{\rm t0} k_1}{\eta k_{\rm f}} \equiv \frac{\lambda}{\eta k_{\rm f} k_1} \tag{30}$$

(see [19]). The main point here is that the parameter a scales with the magnetic Reynolds number, so it is very large under astrophysical circumstances. Moreover, if the growth rate λ scales with the inverse turnover time on the forcing scale, then the $k_{\rm f}$ in the denominator cancels and a is proportional to the magnetic Reynolds number [53,54,55]. This is also consistent with the analysis of [56].

A consequence of the large value of a is a very slow saturation phase and not necessarily a very low saturation level, as is usually believed [54]. That the saturation phase is slow can be seen, for example, by considering the magnitude of the right hand side of (28) near saturation, so we put

$$\langle \overline{\mathbf{B}}^2 \rangle / B_{\rm eq}^2 = (1-\epsilon) k_{\rm f}/k_1, \tag{31}$$

which means that we are a fraction ϵ away from full equipartition. We then have

$$\text{rhs of (28)} = \frac{\alpha_0 k_1 - \eta_{\rm t0} k_1^2}{1 + \frac{\alpha_0 k_1 - \eta_{\rm t0} k_1^2}{\eta k_{\rm f}} (1-\epsilon) \frac{k_{\rm f}}{k_1}} - \eta k_1^2. \tag{32}$$

Since the field is already strong, the '1+' in the denominator may be neglected, so we obtain

$$\text{rhs of (28)} = \left[\frac{\alpha_0 k_1 - \eta_{\rm t0} k_1^2}{(\alpha_0 k_1 - \eta_{\rm t0} k_1^2)(1-\epsilon)} - 1 \right] \eta k_1^2 = \frac{\epsilon}{1-\epsilon} \eta k_1^2 \approx \epsilon \eta k_1^2. \tag{33}$$

The rhs of (28) can be regarded as a local growth rate if the field was unchanged. The extrapolated saturation time would therefore be $(\epsilon \eta k_1^2)^{-1}$, which is more than a resistive time!

6.2 Other Quenching Formulae

By contrast, consider a quenching of the form

$$q(\overline{\mathbf{B}}) = 1 - a \overline{\mathbf{B}}^2 / B_{\rm eq}^2 \quad \text{(quadratic formula)}. \tag{34}$$

To match the right saturation field strength we have to have

$$(\alpha_0 k_1 - \eta_{\rm t0} k_1^2)(1 - a k_{\rm f}/k_1) - \eta k_1^2 = 0. \tag{35}$$

This gives

$$a = \frac{\alpha_0 k_1 - \eta_{\rm T0} k_1^2}{\alpha_0 k_1 - \eta_{\rm t0} k_1^2} \frac{k_1}{k_{\rm f}} \to \frac{k_1}{k_{\rm f}} \quad \text{for small } \eta. \tag{36}$$

Again, we calculate the instantaneous growth rate for a field that was close to final saturation, i.e. we put $\langle \overline{\mathbf{B}}^2 \rangle / B_{\rm eq}^2 = (1-\epsilon) k_{\rm f}/k_1$. This gives

$$\text{rhs of (28)} = (\alpha_0 k_1 - \eta_{\rm t0} k_1^2) \left[1 - \frac{\alpha_0 k_1 - \eta_{\rm T0} k_1^2}{\alpha_0 k_1 - \eta_{\rm t0} k_1^2} (1-\epsilon) \right] - \eta k_1^2 \tag{37}$$

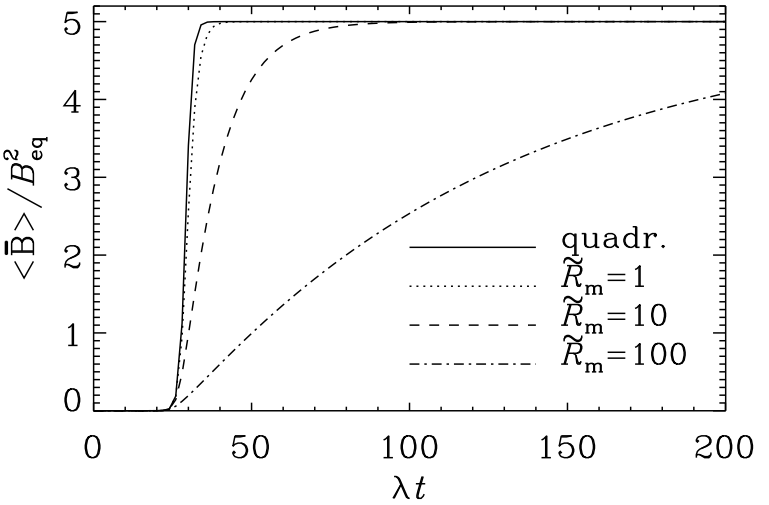


Fig. 8. Saturation behavior for the quadratic quenching formula (solid line) compared with different lorentzian quenching formulae with different values of $\tilde{R}_{\rm m} \equiv \lambda/(\eta k_1^2)$ (broken lines). Note the slow saturation behavior, consistent with (14). The results for the quadratic quenching formula are independent of the value of $\tilde{R}_{\rm m}$, which is therefore inconsistent with (14).

or, after some simplifications,

$$\text{rhs of (28)} = \epsilon(\alpha_0 k_1 - \eta_{\rm T0} k_1^2). \tag{38}$$

This means that it will only take a few dynamical time scales before the extrapolated field will reach final saturation, which is of course incompatible with the magnetic helicity constraint (14). Figure 8 shows the great discrepancy between these two quenching formulae. The quadratic quenching formula gives, for different values of $\tilde{R}_{\rm m} \equiv \lambda/(\eta k_1^2)$, always the same saturation behavior, while the lorentzian formula gives a more prolonged saturation phase as $\tilde{R}_{\rm m}$ is increased.

One may still be tempted to expect that there could be many other quenching formulae that might work as well. One clear counter example is shown in Fig. 9 where we compare the results from a cubic quenching formula with the magnetic helicity constraint in (14). For the cubic quenching we just used the very simple formula $q = 1/(1+\beta^3)$, where $\beta^2 = a\langle\overline{\mathbf{B}}^2\rangle/B_{\rm eq}^2$. The departure between cubic is not very strong, but clearly noticeable. We note that if we replaced $\eta \to \frac{3}{2}\eta$, the helicity constraint would actually fit, but of course η is an input parameter, so we cannot just adopt a different value in the analysis. We may therefore conclude that cubic quenching can be ruled out.

Finally, we consider a quenching formula that has the correct strong field asymptotics, $q \to \beta^{-2}$, and also the same weak field Taylor expansion as the

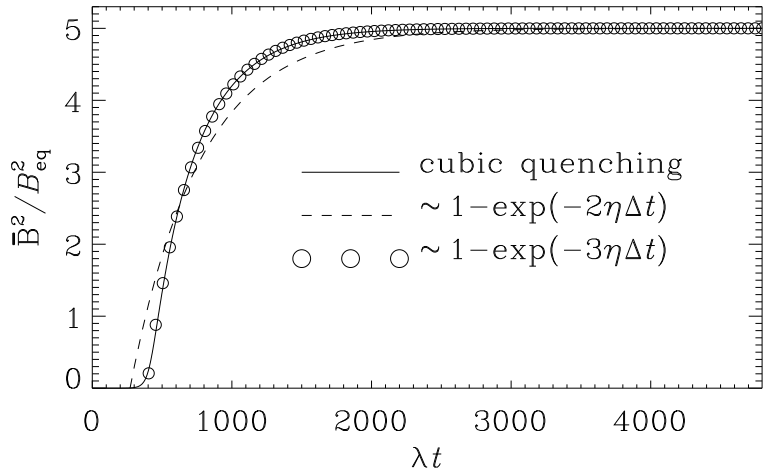


Fig. 9. Saturation behavior for cubic quenching (solid line) compared with the magnetic helicity constraint (dashed line). Note also that if we used the *wrong* microscopic diffusivity, $\eta \to \frac{3}{2}\eta$, the helicity constraint would actually fit.

lorentzian formula, $q \approx 1 + \beta^2$,

$$q(\overline{\mathbf{B}}) = \frac{1}{\beta^2}\left(1 - \frac{\tan^{-1}\sqrt{3\beta^2}}{\sqrt{3\beta^2}}\right), \quad \text{where} \;\; \beta^2 = a\langle\overline{\mathbf{B}}^2\rangle/B_{\rm eq}^2. \tag{39}$$

This expression entered in the quenching formula derived by Field, Blackman, & Cho [57]. In Fig. 10 we compare the saturation behavior for the atan formula with the lorentzian one. Clearly, the atan formula is much better than any of the other formulae considered in this subsection, but the lorentzian formula is still considerably closer to the magnetic helicity constraint than the atan formula.

6.3 Non-universality of the Lorentzian Quenching Formula

We may now be under the impression that the lorentzian formula is probably the correct quenching expression. While is does indeed provide a good description of what is going on in the simulations, we note that there are also a few problems. Firstly, when applied to other models where the field is in general no longer isotropic and force-free (e.g., if there is shear or if there are boundaries) the best fit value of a is no longer equal to the value calculated for the α^2-dynamo; see (30).

The other problem is rather an uncomfortable prediction from the quenching model with the lorentzian formula. Since both α and $\eta_{\rm t}$ are quenched by equal amounts, $\eta_{\rm t}$ will become comparable with η near full saturation. As a consequence, the cycle period of $\alpha\Omega$-dynamo models becomes comparable to the resistive time which is rather long. So, mean-field theory of the solar dynamo may face a very serious problem, unless open boundary effects (Sect. 4) play an

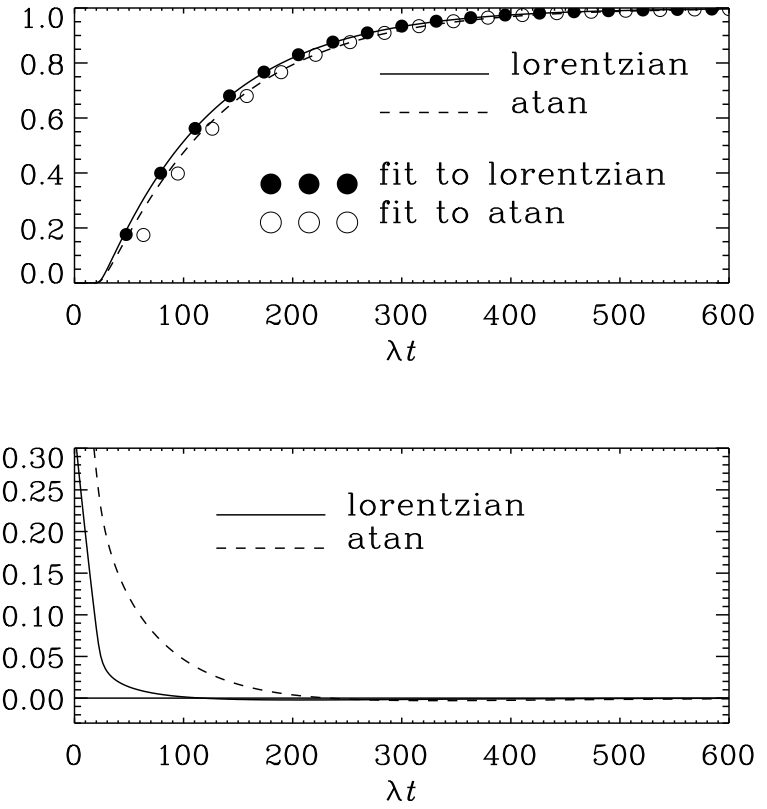


Fig. 10. Comparison of the saturation behavior for the lorentzian and atan formulae, together with the corresponding fits obtained from the helicity constraint.

important role. Simulations perhaps seem to point into this direction as well: the cycle period found in simulations with shear [58] was already rather long, and with a further reduction of the magnetic diffusivity by a factor of 2.5, the cycles disappeared altogether [23]. On the other hand, the absence of cycles could have been for other reasons, for example due to too restricted a geometry (with sinusoidal shear flow on a scale only 5 times larger than the energy carrying scale of the turbulence). Thus, these two simulations are perhaps not yet fully conclusive.

In the following we point out that there is yet another possibility that is at least equally well in agreement with our α^2-dynamo benchmark result and theoretically more theoretically more appealing because it satisfies the magnetic helicity equation exactly at all times.

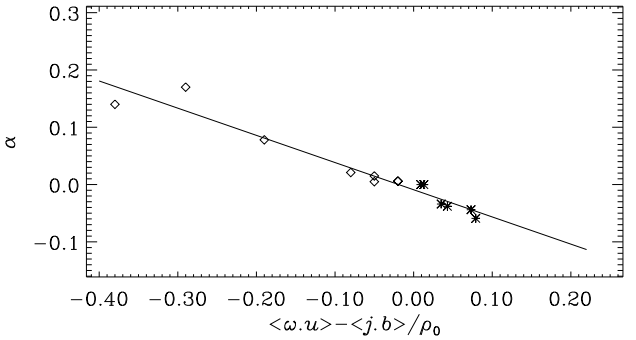


Fig. 11. Dependence of $\alpha = \overline{\mathcal{E}}_y/B_{0y}$ on the residual helicity, obtained by imposing a uniform magnetic field $\mathbf{B}_0 = (0,\ B_{0y},\ 0)$ and driving the turbulence either through the momentum equation (diamonds) or by an extra forcing term in the induction equation (asterisks).

7 Dynamical Quenching

A quenching formula that we have not yet discussed is the formula for the *residual* α-effect,

$$\alpha = \alpha_{\rm K} + \alpha_{\rm M} \quad \text{with} \quad \alpha_{\rm K} = -\frac{\tau}{3}\langle \boldsymbol{\omega}\cdot\mathbf{u}\rangle, \quad \alpha_{\rm M} = +\frac{\tau}{3\rho_0}\langle \mathbf{j}\cdot\mathbf{b}\rangle, \tag{40}$$

which is due to Pouquet, Frisch, and Léorat [46], and has frequently been used in connection with α-quenching [59,60,61,62]. The result that the α-effect is proportional to the *residual* helicity, $\langle \boldsymbol{\omega}\cdot\mathbf{u}\rangle - \langle \mathbf{j}\cdot\mathbf{b}\rangle/\rho_0$ has also been confirmed numerically [63] by imposing a uniform magnetic field and driving the turbulence either through the momentum equation (as is done in the rest of the paper) or through a forcing term in the induction equation; see Fig. 11.

The question now is how to use (40) in a mean field model, which only knows about the mean field, $\overline{\mathbf{B}}$. If we were to approximate $\langle \mathbf{j}\cdot\mathbf{b}\rangle$ in a direct manner by $\overline{\mathbf{B}}^2$, this would correspond to the quenching formula (34), which is clearly ruled out because it violates the magnetic helicity equation. On the other hand, we can explicitly make sure that the magnetic helicity equation (7) is obeyed. The contribution $\langle \overline{\mathbf{A}}\cdot\overline{\mathbf{B}}\rangle$ from the large scale fields to the magnetic helicity equation is automatically taken into account by the mean-field equation, so we only need to solve for the missing contribution from small scales, $\langle \mathbf{a}\cdot\mathbf{b}\rangle$, and the two equations for $\langle \overline{\mathbf{A}}\cdot\overline{\mathbf{B}}\rangle$ and $\langle \mathbf{a}\cdot\mathbf{b}\rangle$ must be fully coupled.

Another way of seeing this [64] is that, while $\langle \mathbf{A}\cdot\mathbf{B}\rangle$ stays close to zero on short enough time scales, any increase of the magnetic field by the α-effect leads to an increase of the large scale magnetic helicity, $\langle \overline{\mathbf{A}}\cdot\overline{\mathbf{B}}\rangle$. This can only be consistent with an almost unchanged $\langle \mathbf{A}\cdot\mathbf{B}\rangle$ if there is a simultaneous generation of small scale magnetic helicity, $\langle \mathbf{a}\cdot\mathbf{b}\rangle$, of opposite sign, so that

$$\langle \mathbf{A}\cdot\mathbf{B}\rangle = \langle \overline{\mathbf{A}}\cdot\overline{\mathbf{B}}\rangle + \langle \mathbf{a}\cdot\mathbf{b}\rangle \tag{41}$$

stays close to zero. The price to pay for this is that the small scale magnetic helicity can also produce an α-effect, $\alpha_{\rm M}=\frac{1}{3}\tau\langle\mathbf{j}\cdot\mathbf{b}\rangle/\rho_0$, but it has the opposite sign than $\alpha_{\rm K}$, so the residual α-effect becomes quenched. Mathematically, this quenching of α can be described by the magnetic helicity equation. The contribution of the large scale field to the magnetic helicity equation follows from the mean-field equation. The small scale contribution is exactly such that the sum of these two equations gives (7). Thus, we have a pair of two equations [64]

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle\overline{\mathbf{A}}\cdot\overline{\mathbf{B}}\rangle=2\langle\overline{\boldsymbol{\mathcal{E}}}\cdot\overline{\mathbf{B}}\rangle-2\eta\mu_0\langle\overline{\mathbf{J}}\cdot\overline{\mathbf{B}}\rangle, \tag{42}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle\mathbf{a}\cdot\mathbf{b}\rangle=-2\langle\overline{\boldsymbol{\mathcal{E}}}\cdot\overline{\mathbf{B}}\rangle-2\eta\mu_0\langle\mathbf{j}\cdot\mathbf{b}\rangle. \tag{43}$$

where $\overline{\boldsymbol{\mathcal{E}}}=\overline{\mathbf{u}\times\mathbf{b}}$ is the mean turbulent electromotive force, for which we adopt the usual mean-field closure in terms of α-effect and turbulent magnetic diffusivity, i.e.

$$\overline{\boldsymbol{\mathcal{E}}}=\alpha\overline{\mathbf{B}}-\eta_{\rm t}\mu_0\overline{\mathbf{J}}. \tag{44}$$

Note that (42) follows directly from the usual mean-field dynamo equation (1). Making use of the relation $\mu_0\langle\mathbf{j}\cdot\mathbf{b}\rangle=k_{\rm f}^2\langle\mathbf{a}\cdot\mathbf{b}\rangle$ in (40), (43) becomes [56]

$$\frac{\mathrm{d}\alpha_{\rm M}}{\mathrm{d}t}=-2\eta k_{\rm f}^2\left(R_{\rm m}\frac{\langle\overline{\boldsymbol{\mathcal{E}}}\cdot\overline{\mathbf{B}}\rangle}{B_{\rm eq}^2}+\alpha_{\rm M}\right), \tag{45}$$

where $R_{\rm m}=\eta_{\rm t}/\eta$ is the appropriate definition in the present context.

An evolution equation for α was already proposed twenty years ago [65]; see also [66,67]. Nevertheless, dynamical quenching was usually ignored, although it has sometimes been used in mean-field models with the main motivation to promote and study chaotic behavior in stellar dynamos [68,69,70,71]. Kleeorin et al. [22] were the first to point out that the catastrophic quenching of Vainshtein & Cattaneo [54] is just a special case of dynamical quenching.

7.1 Adiabatic Approximation and Force-Free Degeneracy

Near the saturated state the explicit time derivative in (45), $\mathrm{d}\alpha_{\rm M}/\mathrm{d}t$, can be neglected and the value of $\alpha_{\rm M}$ adjusts 'adiabatically' as the field saturates. Thus, we have [56]

$$0=R_{\rm m}\frac{\langle\overline{\boldsymbol{\mathcal{E}}}\cdot\overline{\mathbf{B}}\rangle}{B_{\rm eq}^2}+\alpha_{\rm M}, \tag{46}$$

or, after substituting $\alpha_{\rm M}=\alpha-\alpha_{\rm K}$,

$$R_{\rm m}\left(\alpha\langle\overline{\mathbf{B}}^2\rangle-\eta_{\rm t}\langle\overline{\mathbf{J}}\cdot\overline{\mathbf{B}}\rangle\right)+(\alpha-\alpha_{\rm K})=0, \tag{47}$$

which yields

$$\alpha=\frac{\alpha_{\rm K}+R_{\rm m}\eta_{\rm t}\langle\overline{\mathbf{J}}\cdot\overline{\mathbf{B}}\rangle}{1+R_{\rm m}\langle\overline{\mathbf{B}}^2\rangle}. \tag{48}$$

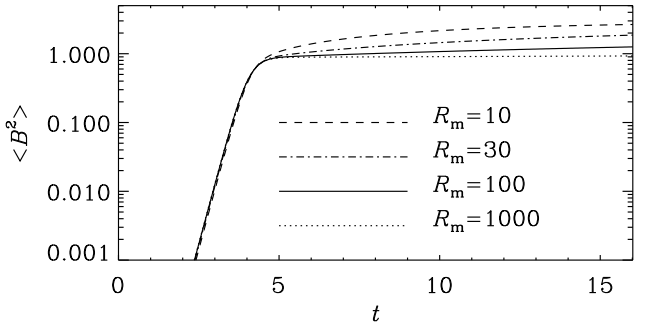


Fig. 12. The early saturation phase of the energy of the mean field in the dynamical quenching model for four different values of the magnetic Reynolds number.

This equation was already obtained by Gruzinov & Diamond [59,67]. The late saturation phase of α^2-dynamos is well described by (48). This is because near saturation time dependence is governed by the slow resistive adjustment in the mean field equation for the large-scale field, whilst the α equation is quickly adjusting to whatever the large-scale field is at any time.

The reason why the lorentzian quenching formula describes the resistively limited quenching behavior so well is because in the case of a nearly force free large scale magnetic field the lorentzian quenching formula and the adiabatic approximation become identical. Indeed, if the large-scale magnetic field is force-free, we have $\langle\overline{\mathbf{J}}\cdot\overline{\mathbf{B}}\rangle\overline{\mathbf{B}} = \langle\overline{\mathbf{B}}^2\rangle\overline{\mathbf{J}}$, which allows us to write the full electromotive force, $\overline{\boldsymbol{\mathcal{E}}} = \alpha\overline{\mathbf{B}} - \eta_{\rm t}\mu_0\overline{\mathbf{J}}$, in the form

$$\begin{aligned}\overline{\boldsymbol{\mathcal{E}}} &= \frac{\alpha_{\rm K} + R_{\rm m}\eta_{\rm t}\mu_0\langle\overline{\mathbf{J}}\cdot\overline{\mathbf{B}}\rangle/B_{\rm eq}^2}{1+R_{\rm m}\langle\overline{\mathbf{B}}^2\rangle/B_{\rm eq}^2}\,\overline{\mathbf{B}} - \eta_{\rm t0}\mu_0\overline{\mathbf{J}} \\ &= \frac{\alpha_{\rm K}\overline{\mathbf{B}}}{1+R_{\rm m}\langle\overline{\mathbf{B}}^2\rangle/B_{\rm eq}^2} - \frac{\eta_{\rm t0}\mu_0\overline{\mathbf{J}}}{1+R_{\rm m}\langle\overline{\mathbf{B}}^2\rangle/B_{\rm eq}^2},\end{aligned} \tag{49}$$

which shows that in the force-free case the adiabatic approximation together with constant (unquenched) turbulent magnetic diffusivity becomes equal to the pair of expressions where both α and $\eta_{\rm t}$ are catastrophically quenched. This is called the force-free degeneracy [56]. This degeneracy is lifted in cases with shear or when the turbulence is no longer fully helical.

7.2 α^2-Dynamos

When applied to an α^2-dynamo model with dynamical quenching, the helicity constraint is well satisfied [56,64] and the difference between the solutions with dynamical and algebraic quenching turns out to be small if $R_{\rm m}$ is less than about 1000. The difference in the evolution of magnetic energy with dynamical

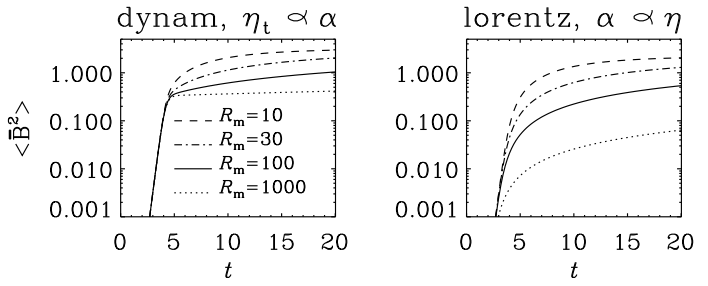


Fig. 13. The effects of assuming $\eta_t \propto \alpha$ in the dynamical and the lorentzian quenching models for four different values of the magnetic Reynolds number.

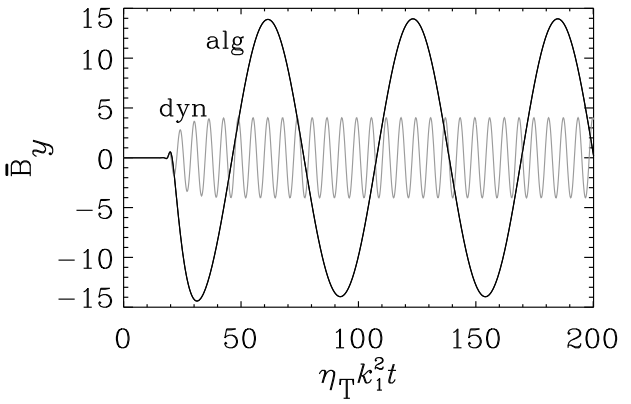


Fig. 14. Comparison of $\alpha\Omega$-dynamo models with algebraic (solid line) and dynamical quenching (grey line). $\Omega'/(\eta_T k_1^2) = 100$, $\alpha_K/(\eta_T k_1) = 0.1$, $R_m \equiv \eta_t/\eta = 10$.

quenching and with algebraic (or lorentzian) quenching, is only a few percent if the magnetic Reynolds number is small. The difference increases as the magnetic Reynolds number increases; see Fig. 13, where we plot the evolution of $\langle \overline{\mathbf{B}}^2 \rangle$ around the time when the kinematic exponential growth turns into the resistively limited saturation phase which was already described in Sect. 3.5. Conclusive agreement with simulations is at this point not possible, mostly because the magnetic Reynolds numbers are not large enough.

7.3 $\alpha\Omega$-Dynamos

When shear is included, toroidal field can be regenerated solely by the shear term. This is where dynamical and algebraic quenching lead to very different behaviors. With algebraic quenching, the reduction of η_t leads to resistively long cycle periods in the nonlinear regime. With dynamical quenching, η_t is

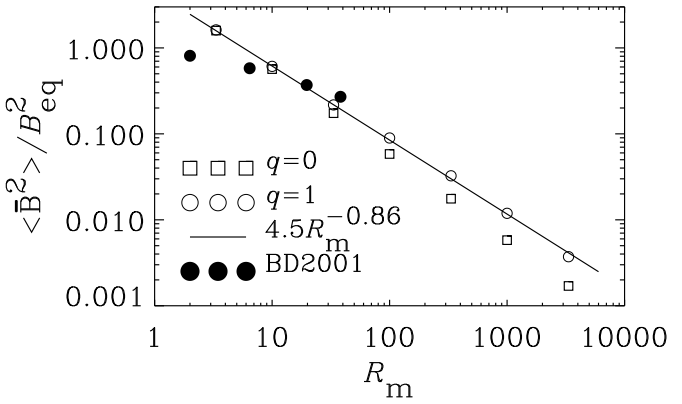


Fig. 15. Saturation energy versus $R_{\rm m}$ for α^2-dynamos with boundaries. Models with ($q = 1$) and without ($q = 0$) loss term are indicated by open circles and open squares, respectively. The line gives the fit $\langle\overline{\mathbf{B}}^2\rangle \sim R_{\rm m}^{-0.86}$. Simulations of [20] are shown as full circles.

constant and the cycle frequency remains of order unity; see Fig. 14. Here, $R_{\rm m}$ is only 10, but for larger values the final field amplitude and the cycle period are considerably enhanced.

7.4 Open Boundaries

Finally we consider a model with vertical field boundary conditions ($\overline{B}_x = \overline{B}_y = 0$ on $z = \pm\pi$) and solve the α^2-dynamo equation. The resulting mean-squared field strength is plotted in Fig. 15 versus $R_{\rm m}$. We find that $\langle\overline{\mathbf{B}}^2\rangle \sim R_{\rm m}^{-1}$, which is consistent with (16), but steeper than what was obtained in the simulations [20]. Kleeorin et al. [22,72] pointed out that in (45) there should be an additional loss term on the right hand side. By making this loss term suitably $R_{\rm m}$-dependent, one could in principle make the $R_{\rm m}$-dependence of $\langle\overline{\mathbf{B}}^2\rangle$ less steep, but some reduction is already obtained by allowing for a diffusion-like loss term of the form $q\eta_{\rm T}\nabla^2\alpha_{\rm M}$ on the right hand side; see the open circles in Fig. 15. The parameter q is used to regulate the efficiency of this loss. The simulations of [20] are shown as full circles.

7.5 Generalization to Nonuniform α

In all astrophysical bodies there are opposite signs of kinetic helicity, $\langle\boldsymbol{\omega}\cdot\mathbf{u}\rangle$, in the northern and southern hemispheres. In the approach of Kleeorin and collaborators [65,66,67] the dynamical α-quenching framework was always used as a theory for nonuniform α, which does not seem to be well justified. Most importantly, the connection between $\langle\mathbf{j}\cdot\mathbf{b}\rangle$ (in the expression for $\alpha_{\rm M}$) and $\langle\mathbf{a}\cdot\mathbf{b}\rangle$

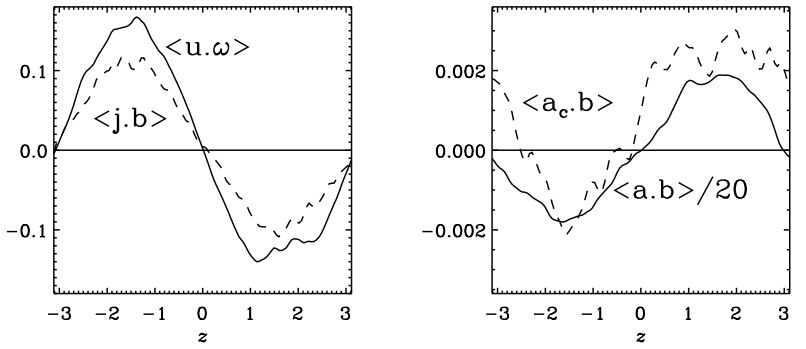


Fig. 16. Comparison of the horizontally averaged kinetic and current helicity densities, $\langle \boldsymbol{\omega} \cdot \mathbf{u} \rangle$ and $\langle \mathbf{j} \cdot \mathbf{b} \rangle$, respectively (left hand panel), and the horizontally averaged magnetic helicity densities in the $\phi = 0$ gauge $\langle \mathbf{a} \cdot \mathbf{b} \rangle$ and the Coulomb gauge $\langle \mathbf{a}_{\rm c} \cdot \mathbf{b} \rangle$.

is no longer straightforward when the angular brackets denote ensemble averages (which are not really of practical interest) or averages over one or two periodic coordinate directions. Indeed, $\langle \mathbf{a} \cdot \mathbf{b} \rangle$ is no longer gauge-invariant, so one has to fix the gauge. The Coulomb gauge is the most common one, but one should realize that going to another gauge can make a major difference. In Fig. 16 we show the magnetic helicity in the $\phi = 0$ and the Coulomb gauges, $\langle \mathbf{a} \cdot \mathbf{b} \rangle$ and $\langle \mathbf{a}_{\rm c} \cdot \mathbf{b} \rangle$, respectively (see Sect. 3). Note that the Coulomb gauged magnetic helicity is about twenty times smaller than that in the $\phi = 0$ gauge. More importantly, the magnetic helicity is *not* a positive multiple of the current helicity, $\langle \mathbf{j} \cdot \mathbf{b} \rangle$. (We recall that for homogeneous turbulence, $\langle \mathbf{j} \cdot \mathbf{b} \rangle = k_{\rm f}^2 \langle \mathbf{a} \cdot \mathbf{b} \rangle$.) Another problem is that when solving numerically the evolution equation for a space-dependent α-effect one needs for reasons of numerical stability a diffusion term [72,73]. However, it is now clear that a loss or exchange of small-scale helicity leads to an enhancement of the large-scale field [56], but from simulations we know that the presence of an equator rather lowers the energy of the mean field. Perhaps this could be fixed by adopting an additional loss term in the mean-field equation for the large-scale field, but this procedure would be completely ad hoc. One may hope however that some kind of a generalization of the dynamical α-quenching is at least in principle possible.

8 Conclusions

In this review we have outlined some of the main results of isotropic MHD simulations in the presence of helicity. We have focussed on the connection with the α-effect in mean-field dynamo theory. We should emphasize that in the case where the magnetic energy density is uniform in space, the agreement between simulations and theory is now well established. In all other cases, things are immediately more complicated. Moreover, dynamical quenching cannot readily be

generalized to the case where $\alpha_{\rm M}$ varies in space. In that case the equation for the magnetic helicity density would not be gauge-invariant. Another problem arises when $\alpha_{\rm K}$ varies in space and if it changes sign across the equator, for example. These are very important aspects requiring clarification. It is quite possible that significant improvement in the theory will soon be possible. Without a corresponding generalization of dynamic α-quenching, if it is ever possible, it would be difficult to use dynamo theory in astrophysically interesting circumstances.

References

1. E. N. Parker: Astrophys. J. 122, 293 (1955)
2. M. Steenbeck, F. Krause, K.-H. Rädler: Z. Naturforsch. 21a, 369 (1966) See also the translation in Roberts & Stix, "The turbulent dynamo" Tech. Note 60, NCAR, Boulder, Colorado (1971)
3. R. B. Leighton: Astrophys. J. 156, 1 (1969)
4. A. Ferriz-Mas, D. Schmitt, M. Schüssler: Astron. Astrophys. 289, 949 (1994)
5. A. Brandenburg, D. Schmitt: Astron. Astrophys. 338, L55 (1998)
6. A. Brandenburg, Å. Nordlund, R. F. Stein, U. Torkelsson: Astrophys. J. 446, 741 (1995)
7. A. Brandenburg, K. J. Donner: Month. Not. Roy. Astron. Soc. 288, L29 (1997)
8. P. A. Gilman, P. A. Fox: Astrophys. J. 484, 439 (1997)
9. F. Krause, K.-H. Rädler: *Mean-Field Magnetohydrodynamics and Dynamo Theory* (Akademie-Verlag, Berlin; also Pergamon Press, Oxford 1980)
10. A. Brandenburg: 'Disc Turbulence and Viscosity'. In: *Theory of Black Hole Accretion Discs*, ed. by M. A. Abramowicz, G. Björnsson & J. E. Pringle (Cambridge University Press 1998) pp. 61-86
11. G. Rüdiger, V. V. Pipin: Astron. Astrophys. 362, 756 (2000)
12. M. Steenbeck, F. Krause: Astron. Nachr. 291, 49 (1969)
13. G. Rüdiger, A. Brandenburg: Astron. Astrophys. 296, 557 (1995)
14. M. Dikpati, P. Charbonneau: Astrophys. J. 518, 508 (1999)
15. M. Küker, G. Rüdiger, M. Schultz: Astron. Astrophys. 374, 301 (2001)
16. F. Cattaneo: Astrophys. J. 515, L39 (1999)
17. E. T. Vishniac, J. Cho: Astrophys. J. 550, 752 (2001)
18. R. Arlt, A. Brandenburg: Astron. Astrophys. 380, 359 (2001)
19. A. Brandenburg: Astrophys. J. 550, 824 (2001a)
20. A. Brandenburg, W. Dobler: Astron. Astrophys. 369, 329 (2001)
21. E. G. Blackman, G. F. Field: Astrophys. J. 534, 984 (2000)
22. N. I. Kleeorin, D. Moss, I. Rogachevskii, D. Sokoloff: Astron. Astrophys. 361, L5 (2000)
23. A. Brandenburg, W. Dobler, K. Subramanian: Astron. Nachr. 323, 99 (2002)
24. N. Seehafer: Solar Phys. 125, 219 (1990)
25. A. A. Pevtsov, R. C. Canfield, T. R. Metcalf: Astrophys. J. 440, L109 (1995)
26. S. D. Bao, H. Q. Zhang, G. X. Ai, M. Zhang: Astron. Astrophys. 139, 311 (1999)
27. A. A. Pevtsov, S. M. Latushko: Astrophys. J. 528, 999 (2000)
28. M. A. Berger, A. Ruzmaikin: J. Geophys. Res. 105, 10481 (2000)
29. C. R. DeVore: Astrophys. J. 539, 944 (2000)
30. J. Chae: Astrophys. J. 540, L115 (2000)
31. B. C. Low: J. Geophys. Res. 106, 25,141 (2001)

32. A. Brandenburg, Å. Nordlund, P. Pulkkinen, R. F. Stein, I. Tuominen: Astron. Astrophys. 232, 277 (1990)
33. K. Ferrière: Astrophys. J. 389, 286 (1992)
34. G. Rüdiger, L. L. Kitchatinov: Astron. Astrophys. 269, 581 (1993)
35. M. Ossendrijver, M. Stix, A. Brandenburg: Astron. Astrophys. 376, 713 (2001)
36. A. Brandenburg: astro-ph/0109497 (2001)
37. M. Meneguzzi, U. Frisch, A. Pouquet: Phys. Rev. Lett. 47, 1060 (1981)
38. B. Galanti, P.-L. Sulem: Phys. Fluids A 3, 1778 (1991)
39. B. Galanti, P.-L. Sulem, A. D. Gilbert: Physica D 47, 416 (1991)
40. D. Balsara, A. Pouquet: Phys. Plasmas 6, 89 (1999)
41. A. D. Gilbert: Geophys. Astrophys. Fluid Dyn. 96, 135 (2002)
42. D. W. Hughes, F. Cattaneo, E. J. Kim: Phys. Lett. 223, 167 (1996)
43. M. Berger, G. B. Field: J. Fluid Mech. 147, 133 (1984)
44. A. Brandenburg, G. R. Sarson: Phys. Rev. Lett. 88, 055003 (2002)
45. G. A. Glatzmaier, P. H. Roberts: Nature 377, 203 (1995)
46. A. Pouquet, U. Frisch, J. Léorat: J. Fluid Mech. 77, 321 (1976)
47. A. Brandenburg, D. Sokoloff: Geophys. Astrophys. Fluid Dyn., in press (2002), astro-ph/0111568
48. M. Berger: Geophys. Astrophys. Fluid Dyn. 30, 79 (1984)
49. M. R. Brown, R. C. Canfield, A. A. Pevtsov: *Magnetic Helicity in Space and Laboratory Plasmas* (Geophys. Monograph **111**, American Geophysical Union, Florida 1999)
50. M. Dikpati, P. A. Gilman: Astrophys. J. 559, 428 (2001)
51. B. R. Durney: Solar Phys. 166, 231 (1995)
52. A. R. Choudhuri, M. Schüssler, M. Dikpati: Astron. Astrophys. 303, L29 (1995)
53. F. Cattaneo, S. I. Vainshtein: Astrophys. J. 376, L21 (1991)
54. S. I. Vainshtein, F. Cattaneo: Astrophys. J. 393, 165 (1992)
55. F. Cattaneo, D. W. Hughes: Phys. Rev. E 54, R4532 (1996)
56. E. G. Blackman, A. Brandenburg: Astrophys. J., in press (2002), astro-ph/0204497
57. G. B. Field, E. G. Blackman, H. Chou: Astrophys. J. 513, 638 (1999)
58. A. Brandenburg, A. Bigazzi, K. Subramanian: Month. Not. Roy. Astron. Soc. 325, 685 (2001)
59. A. V. Gruzinov, P. H. Diamond: Phys. Rev. Lett. 72, 1651 (1994)
60. A. V. Gruzinov, P. H. Diamond: Phys. Plasmas 2, 1941 (1995)
61. A. V. Gruzinov, P. H. Diamond: Phys. Plasmas 3, 1853 (1996)
62. A. Bhattacharjee, Y. Yuan: Astrophys. J. 449, 739 (1995)
63. A. Brandenburg: 'Helicity in large-scale dynamo simulations'. In: *Magnetic Helicity in Space and Laboratory Plasmas*, ed. by M. R. Brown, R. C. Canfield, A. A. Pevtsov (Geophys. Monograph **111**, American Geophysical Union, Florida 1999) pp. 65-73
64. G. B. Field, E. G. Blackman: Astrophys. J. 572, 685 (2002)
65. N. I. Kleeorin, A. A. Ruzmaikin: Magnetohydrodynamics 18, 116 (1982); translation from Magnitnaya Gidrodinamika, 2, pp. 17–24 (1982)
66. Ya. B. Zeldovich, A. A. Ruzmaikin, D. D. Sokoloff: *Magnetic Fields in Astrophysics* (Gordon & Breach, New York 1983)
67. N. I. Kleeorin, I. Rogachevskii, A. Ruzmaikin: Astron. Astrophys. 297, 159 (1995)
68. A. A. Ruzmaikin: Comments Astrophys. 9, 85 (1981)
69. S. Schmalz, M. Stix: Astron. Astrophys. 245, 654 (1991)
70. U. Feudel, W. Jansen, J. Kurths: J. Bifurcation & Chaos 3, 131 (1993)
71. E. Covas, A. Tworkowski, A., Brandenburg, R. Tavakol: Astron. Astrophys. 317, 610 (1997)

72. N. I Kleeorin, D. Moss, I. Rogachevskii, D. Sokoloff: Astron. Astrophys. 387, 453 (2002)
73. E. Covas, R. Tavakol, A. Tworkowski, A. Brandenburg: Astron. Astrophys. 329, 350 (1998)

Recent Developments in Magnetic Dynamo Theory

Eric G. Blackman

Department of Physics and Astronomy, and Laboratory for Laser Energetics, University of Rochester, Rochester NY 14627, USA

Abstract. Two spectral regimes of magnetic field amplification in magnetohydrodynamic (MHD) flows can be distinguished by the scale on which fields are amplified relative to the primary forcing scale of the turbulence. For field amplification at or below the forcing scale, the amplification can be called a "small-scale dynamo." For amplification at and above the forcing scale the process can be called a "large-scale dynamo." $Non-local$ (in wave number) effects play a key role in both the growth of the small-scale field in non-helical turbulence and the growth of large and small-scale fields in helical turbulence. Mean field dynamo (MFD) theory represents a simple semi-analytic way to get a handle on large-scale field amplification in MHD turbulence. Helicity has long been known to be important for large scale, flux generating, externally forced MFDs. The extent to which such MFDs operate "slow" or "fast" (dependent or independent on magnetic Reynolds number) has been controversial, but there has been recent progress. Simulations of α^2 dynamos in a periodic box dynamo and their quenching can now be largely understood within a simplified dynamical non-linear paradigm in which the MFD growth equation is supplemented by the total magnetic helicity evolution equation. For α^2 dynamos, the large-scale field growth is directly related to the large-scale magnetic helicity growth. Magnetic helicity conservation then implies that growth of the large-scale magnetic helicity induces growth of small-scale magnetic (and current) helicity of the opposite sign, which eventually suppresses the α effect driving the MFD growth. Although the α^2 MFD then becomes slow in the long time limit, substantial large-scale field growth proceeds in a kinematic, "fast" phase before non-linear asymptotic quenching of the "slow" phase applies. Ultimately, the MFD emerges as a process that transfers magnetic helicity between small and large scales. How these concepts apply to more general dynamos with shear, and open boundary dynamos is a topic of ongoing research. Some unresolved issues are identified. Overall, the following summarizes the most recent progress in mean-field dynamo theory:

For a closed turbulent flow,
the non-linear mean field dynamo,
is first fast and kinematic, then slow and dynamic,
and magnetic helicity transfer makes it so.

1 Small-Scale vs. Large-Scale Field Amplification

A dynamo is a process which exponentially amplifies or sustains magnetic energy in the presence of finite dissipation. In this paper I will focus on magnetohydrodynamic (MHD) dynamos, where the only flux dissipating term in Ohm's law for the total magnetic field is the resistive term.

E. Falgarone and T. Passot (Eds.): LNP 614, pp. 432–463, 2003.
© Springer-Verlag Berlin Heidelberg 2003

The simple definition of a dynamo given above does not distinguish the scale on which the magnetic energy is sustained against turbulent forcing, whether a net flux is produced in the spatial region of interest, or the nature of the forcing (e.g. shear driven or isotropically forced). It is helpful to distinguish between "small-scale dynamos" which describe field generation at or below the turbulent forcing scale, and "large-scale dynamos" which describe field generation on scales larger than the forcing scale. Both present their own set of problems.

For the Galaxy, supernovae dominate the nearly isotropic turbulent forcing, and although there is a range of forcing scales, typically the dominant scale is $\sim 50 - 100$pc [1]. A cascade leads to a nearly Kolmogorov turbulent kinetic energy spectrum. Faraday rotation and synchrotron polarization observations reveal the presence of a random component of the Galactic field, also with a dominant scale of $\sim 50 - 100$ parsecs, and an ordered toroidal field on the scale of $\gtrsim$ 1kpc

An important point about large-scale field growth is that regardless of long standing debates about whether the large-scale fields of the Galaxy is primordial or produced in situ, [2,3,4] and regardless of similar debates about the origin of large-scale, jet producing, poloidal fields in accretion disks (e.g. [5]) one thing we do know is that the Sun, at least *must* have a large-scale dynamo operating because the mean flux reverses sign every 11 years. If the flux were simply flux frozen into the sun's formation from the protostellar gas, we would not expect such reversals.

For both small and large-scale dynamos, the richness and the complication of dynamo theory is the non-linearity of the MHD equations. To really understand the theory, we need to understand the backreaction of the growing magnetic field on the turbulence, the saturated spectra of the magnetic field, and the spectral evolution time scales. Do theoretical and numerical calculations make predictions which are consistent with what is observed in astrophysical systems or not? What are the limitations of these predictions?

1.1 Small-Scale Dynamo

The "small-scale dynamo" describes magnetic field amplification on and below the turbulent forcing scale (e.g. [6,7,8]). Field energy first builds up to equipartition with the kinetic energy on the smallest scales [9] because the growth time is the turnover time. The turnover times are shorter for smaller scales, and the equipartition level is lower for the small scales. The fast growth, seen in non-linear simulations, can be predicted analytically [6,7,8,9]. The approach to near equipartition is not controversial but the shape of the saturated small-scale spectrum needs more discussion, particularly for the Galaxy.

Recent simulations of forced non-helical turbulence for magnetic Prandtl number ($Pr \equiv \nu_v/\nu_M$, where ν_v is the viscosity and ν_M is the magnetic diffusivity) satisfying $Pr \geq 1$ in a periodic box have shown that the field does not build up to anywhere near equipartition on the input scale of the turbulence [10,11]. Rather, the field piles up on the smallest scales. The interpretation is that the forcing scale velocity directly shears the field into folds or filaments with

length of order the forcing scale but with cross field scale of order the resistive scale, accounting for the spectral power at small scales. (see [12] for a careful semi-analytic study of the $Pr \gg 1$ case). Since the small-scale field is amplified by shear directly from the input scale, one can argue that there is a $non-local$ direct cascade.

How does the observed small-scale field of the Galaxy compare with the above results? Since observations show near equipartition of the field energy with the kinetic energy at the forcing scale [13], there is a discrepancy between the observations and the numerical simulation results for non-helical turbulence.

A possible resolution might arise from the idealized problem, explored in [14]. When the turbulence is forced with sufficient kinetic helicity $\langle \mathbf{v} \cdot \nabla \times \mathbf{v} \rangle$, the spectrum changes in an important way. Figures 1 and 2 below come from 3-D MHD simulations in which kinetic energy is forced in a box, and the helical fraction of the kinetic forcing is varied ($f_h = 1$ corresponds to maximal helical forcing). The forcing wave number is $k/2\pi = 4.5$. For sufficiently large f_h, (determined by that for which the kinematic α^2 dynamo can grow) the magnetic spectrum grows two peaks, one at $k/2\pi = 1$ and one at the forcing scale. The kinetic helicity thus influences both the small and large-scale field growth.

Whether these results apply to the Galaxy or protogalaxy is unclear since those systems have shear (see also [15]), boundaries, and stratification unlike the simulations of [14]. But the principle that the large *and* small-scale fields are both influenced by helicity has been demonstrated.

The emergence of two peaks motivates a two-scale approach which is a great simplification that will be exploited later. The shift of the small-scale peak to the forcing-scale is less understood than the generation of the large-scale field. I now discuss the latter.

1.2 Large-Scale Dynamo

The large-scale field of the Galaxy [13,3] seems to be of quadrupole mode and therefore the planar component of the large-scale field has the same sign across the mid-plane. The toroidal field reverses on scales of a few kpc in radius. In these annuli, there appears to be a net toroidal flux when integrated over the full height of the Galaxy. If this inference continues to survive future observations, the mechanism for field production must produce a net toroidal flux in annuli that extend the full vertical disk thickness, not just a net magnetic energy.

A leading framework for understanding the in situ origin of large-scale magnetic field energy and flux growth in galaxies and stars, and even for the peak in the large-scale magnetic energy in the helically forced case of Fig. 1 (see also [16]) has been the mean-field dynamo (MFD) theory [1,7,8,17,18]. The theory appeals to some combination of helical turbulence (leading to the α effect), differential rotation (the Ω effect), and turbulent diffusion (the β effect) to exponentiate an initial seed of mean magnetic field. Reference [19] developed a formalism for describing the concept [7] that helical turbulence can twist toroidal (ϕ) fields into the poloidal (r, z) direction, where they can be acted upon by differential

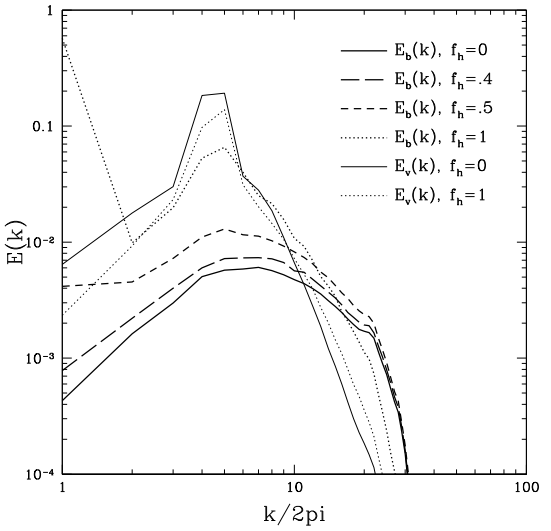


Fig. 1. Saturated kinetic and magnetic energy spectra for successive values of fractional helicity f_h.

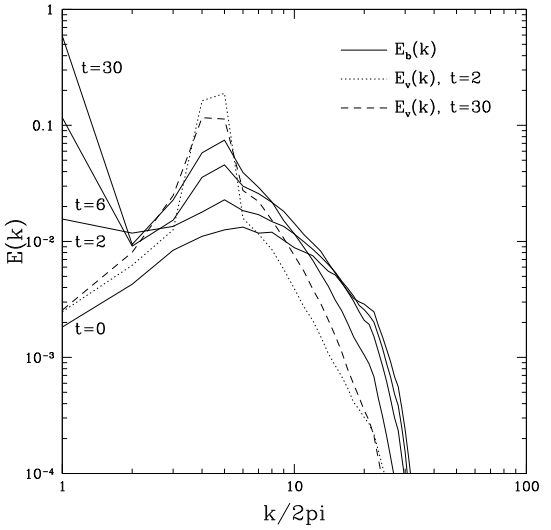


Fig. 2. Time sequence of kinetic and magnetic energy spectra for $f_h = 1$

rotation to regenerate a powerful large-scale toroidal magnetic field. The turbulent diffusion serves to redistribute the flux so that inside the bounded volume of interest, a net flux can grow.

The formalism separates the total magnetic field into a mean component $\overline{\mathbf{B}}$ and a fluctuating component $\mathbf{b}$, and similarly for the velocity field $\mathbf{V}$. The mean

can be a spatial or ensemble average. The ensemble average is approximately equal to the spatial average when there is a scale separation between the mean and fluctuating scales. In reality, the scale separation is often weaker than the dynamo theorist desires, though a weak separation is also helpful given the limited dynamic range of simulations. I proceed to consider spatial averages to simplify the discussion. (Please also note that non-helical large-scale field generation exists from the magnetorotational instability (MRI) [20] but I do not consider this here and focus on externally forced systems. Actually the helical MFD dynamo may be operating even in systems with the MRI. Also I do not consider the model of [15] here as that will be covered elsewhere in this volume.)

The mean field $\overline{\mathbf{B}}$ satisfies the induction equation [17,19]

$$\frac{\partial \overline{\mathbf{B}}}{\partial t} = -c\nabla \times \overline{\mathbf{E}}, \tag{1}$$

where

$$\overline{\mathbf{E}} = -\left(\overline{\mathbf{V}} \times \overline{\mathbf{B}}\right)/c - \langle \mathbf{v} \times \mathbf{b} \rangle / c + \nu_M \nabla \times \overline{\mathbf{B}}, \tag{2}$$

$$\langle \mathbf{v} \times \mathbf{b} \rangle_i = \alpha_{ij} \overline{B}_j - \beta_{ijk} \partial_j \overline{B}_k \tag{3}$$

is the turbulent electromotive force (EMF), and $\nu_M = \eta c^2/4\pi$ is the magnetic diffusivity defined with the resistivity η. Here α_{ij} contains Parker's twisting (the α effect) and β_{ijk} contains the turbulent diffusivity. Reference [19] calculated $\overline{\mathbf{E}}$ to first order in $\overline{B}$ for isotropic α_{ij} and β_{ijk} and hence the pseudo-scalar and scalar dynamo coefficients α and β respectively to zeroth order in $\overline{B}$ from the statistics of the turbulence, ignoring the Navier-Stokes equation. When the Navier-Stokes equation is not used, we speak of the "kinematic theory." Using the equation for the fluctuating field $\partial_t \mathbf{b}$, plugging it into $\langle \mathbf{v} \times \mathbf{b} \rangle$, the standard approach gives a kinematic $\alpha \propto -\int \langle \mathbf{v}(t) \cdot \nabla \times \mathbf{v}(t') \rangle dt' \sim -\tau_c \langle \mathbf{v} \cdot \nabla \times \mathbf{v} \rangle$ and $\beta \propto \int \langle \mathbf{v}(t) \cdot \mathbf{v}(t') \rangle dt' \sim \tau_c \langle \mathbf{v} \cdot \mathbf{v} \rangle$. I will come back to correcting the form for α in Sects. 2.3 and 2.4. because we really want a fully dynamic theory, that accounts for the dynamo coefficients' dependence on $\mathbf{b}$ and $\overline{\mathbf{B}}$. Only then can one fully address the fundamental problem of mean field dynamo theory: how does the growing magnetic field affect the rate and saturation level of growth?

Substituting (2) into (1), gives the MFD equation

$$\frac{\partial \overline{\mathbf{B}}}{\partial t} = \nabla \times \left(\overline{\mathbf{V}} \times \overline{\mathbf{B}}\right) + \nabla \times (\alpha \overline{\mathbf{B}}) - \nabla \times (\beta + \lambda) \nabla \times \overline{\mathbf{B}}. \tag{4}$$

The first term on the right is the non-controversial "$\Omega -$ effect." If one assumes $\overline{\mathbf{V}}$, α and β to be independent of $\overline{B}$, then (4) can be solved as a linear eigenvalue problem for the growing modes of $\overline{\mathbf{B}}$ in the Sun and other bodies. However a rapid growth of the fluctuating field necessarily accompanies the MFD in a turbulent medium. Its impact upon the growth of the mean field, and the impact of the mean field itself on its own growth have been controversial.

The controversy arises because Lorentz forces from the growing magnetic field react back on the turbulent motions driving the field growth [9,21,22,23,24,25,27].

It is useful to distinguish between fast MFD action (also called "rapid" MFD action) and, slow MFD action (also called "resistively limited" MFD action). Fast MFD action proceeds at growth rates which do not go to zero as the magnetic Reynolds number $R_M \to \infty$, and maintain this property even when the non-linear backreaction from the magnetic field is included. Slow MFD action proceeds at rates that vanish as $R_M \to \infty$. I sometimes use "resistively limited" rather than "slow" because the former more explicitly describes the reduced action.

For galaxies and stars, conventional wisdom (which could be challenged [28] and see Sect. 2.3) presumes that rapid MFD action is necessary if the observed large-scale fields are to be wholly produced and sustained by the MFD, given observed cycle periods and available time scales. That this may not be the case is a separate issue from understanding what the theory can actually provide. The latter is the focus herein. If dynamos in stars and galaxies do operate fast, then we would like to understand how, in light of recent numerical and theoretical evidence for slow α^2 dynamos in periodic boxes [16,30]. What are the differences between these dynamos and real systems? where does the theory fail and where does the theory succeed?

There are several issues which must be disentangled. First is the role of magnetic helicity conservation in constraining dynamo theory. In the steady state, such constraints are strongly influenced by boundary conditions, so one must be careful to understand the differences when applying idealized equations for closed systems to real systems with boundaries. Time-dependent dynamical constraints require helicity conservation to be supplemented in some way by the Navier-Stokes equation. Then the fully dynamical evolution of α and the non-linear backreaction can be studied.

Two directions emerge. One is to produce a time dependent dynamical nonlinear theory that fully agrees with numerical simulations in periodic boxes. This has recently been done [30]. The second is to recognize that in real systems, there is physics which has not yet been comprehensively studied. This includes shear [15,29], boundary terms which alleviate magnetic helicity conservation constraints, gravity, and a vertical variation of the dynamo coefficients. It has proven difficult in the simplest generalizations, to make the dynamo asymptotically fast [28,29], but this may not be needed in some applications (e.g. Galaxy) if a kinematic phase lasts long enough. This will be discussed in detail.

In Sect. 2, I discuss the role of magnetic helicity conservation in dynamo theory. I first show that the escape of magnetic helicity through the boundaries might play an important role in maintaining fast MFD action during the steady sustenance phase of dynamos, and thus in interpreting quenching studies for this regime. I discuss the direct observational implications of the boundary terms and the fact that stars and disks harbor active coronae. I then show that the time dependent mean field dynamo is really a process by which magnetic helicity gets transferred from small to large scales by a non-local inverse cascade and that the time-dependent dynamical quenching in recent periodic box simulations can be understood in this framework [30]. At late times, these box dynamos

are resistively limited, depending sensitively on the magnetic Reynolds number, however the kinematic phase lasts quite a while and this fact has very important implications. In the last part of Sect. 2, I discuss an unsolved puzzle that arises in the derivation of the successful dynamical quenching model. In Sect. 3, I summarize some key conclusions and pose open questions.

2 Magnetic Helicity Conservation and Dynamo Quenching

Although the MFD theory predates detailed studies of magnetohydrodynamic (MHD) turbulence, the MFD may be viewed as a framework for studying the inverse cascade of magnetic helicity. Whether this inverse cascade is primarily local (proceeding by interactions of eddies/waves of nearby wave numbers) or non-local (proceeding with a direct conversion of power from large to small wave numbers) is important to understand. The simple MFD seems to be consistent with the latter [16].

From the numerical solution of approximate equations describing the spectra of energy and helicity in MHD turbulence, [31] showed that the α effect conserves magnetic helicity, $H_M = \int (\mathbf{A} \cdot \mathbf{B}) d^3x$, by pumping a positive (negative) amount to scales $> l$ (the outer scale of the turbulence) while pumping a negative (positive) amount to scales $\ll l$. Magnetic energy at the large scale was identified with the $\overline{\mathbf{B}}$ of [19]. Thus, dynamo action leading to an ever larger $\overline{B}$, hence the creation of ever more large-scale helicity, can proceed as long as helicity of the opposite sign can be removed or dissipated. More on this in Sect. 2.3. Here I first derive the general magnetic helicity conservation equations used to constrain the turbulent EMF, and investigate the implications of two steady state cases in detail before considering the time dependent case and interpretation of dynamo simulations.

Using Ohm's law for the electric field,

$$\mathbf{E} = -c^{-1}\mathbf{V} \times \mathbf{B} + \eta \mathbf{J} \tag{5}$$

and averaging its dot product with $\mathbf{B}$, gives

$$\langle \mathbf{E} \cdot \mathbf{B} \rangle = \overline{\mathbf{E}} \cdot \overline{\mathbf{B}} + \langle \mathbf{e} \cdot \mathbf{b} \rangle = -c^{-1} \langle \mathbf{v} \times \mathbf{b} \rangle \cdot \overline{\mathbf{B}} + \eta \overline{\mathbf{J}} \cdot \overline{\mathbf{B}} + \langle \mathbf{e} \cdot \mathbf{b} \rangle \tag{6}$$

where $\mathbf{J}$ is the current density. A second expression for $\langle \mathbf{E} \cdot \mathbf{B} \rangle$ also follows from Ohm's law without first splitting into mean and fluctuating components, that is

$$\langle \mathbf{E} \cdot \mathbf{B} \rangle = \eta \langle \mathbf{J} \cdot \mathbf{B} \rangle = \eta \overline{\mathbf{J}} \cdot \overline{\mathbf{B}} + \eta \langle \mathbf{j} \cdot \mathbf{b} \rangle = \eta \overline{\mathbf{J}} \cdot \overline{\mathbf{B}} + c^{-1} \nu_M \langle \mathbf{b} \cdot \nabla \times \mathbf{b} \rangle . \tag{7}$$

Using (7) and (6), we have

$$-c^{-1} \langle \mathbf{v} \times \mathbf{b} \rangle \cdot \overline{\mathbf{B}} = c^{-1} \lambda \langle \mathbf{b} \cdot \nabla \times \mathbf{b} \rangle - \langle \mathbf{e} \cdot \mathbf{b} \rangle, \tag{8}$$

which can be used to constrain $\langle \mathbf{v} \times \mathbf{b} \rangle$ in the mean field theory.

Now consider $\mathbf{E}$ in terms of the vector and scalar potentials $\mathbf{A}$ and Φ:

$$\mathbf{E} = -\nabla\Phi - (1/c)\partial_t\mathbf{A}. \tag{9}$$

Dotting with $\mathbf{B} = \nabla \times \mathbf{A}$ we have

$$\mathbf{E}\cdot\mathbf{B} = -\nabla\Phi\cdot\mathbf{B} - (1/c)\mathbf{B}\cdot\partial_t\mathbf{A}. \tag{10}$$

After straightforward algebraic manipulation and application of Maxwell's equations, this equation implies

$$\begin{aligned}\mathbf{E}\cdot\mathbf{B} &= -(1/2)\nabla\cdot\Phi\mathbf{B} + (1/2)\nabla\cdot(\mathbf{A}\times\mathbf{E})\\ -(1/2c)\partial_t(\mathbf{A}\cdot\mathbf{B}) &= (-1/2c)\partial_\mu H^\mu(\mathbf{B}) = \eta\mathbf{J}\cdot\mathbf{B},\end{aligned} \tag{11}$$

where

$$H^\mu(\mathbf{B}) = (H_0, H_i) = [\mathbf{A}\cdot\mathbf{B}, c\Phi\mathbf{B} - c\mathbf{A}\times\mathbf{E}] \tag{12}$$

is the magnetic helicity density 4-vector [32], and the contraction has been done with the 4 x 4 matrix $\eta_{\mu\nu}$ where $\eta_{\mu\nu} = 0$ for $\mu \neq \nu$, $\eta_{\mu\nu} = 1$ for $\mu = \nu = 0$ and $\eta_{\mu\nu} = -1$ for $\mu = \nu > 0$. Taking the average of (11) gives

$$\partial_\mu\overline{H}^\mu(\mathbf{B}) = -2c\langle\mathbf{E}\cdot\mathbf{B}\rangle = -2c\overline{\mathbf{E}}\cdot\overline{\mathbf{B}} - 2c\langle\mathbf{e}\cdot\mathbf{b}\rangle = -2c\eta\langle\mathbf{J}\cdot\mathbf{B}\rangle. \tag{13}$$

If, instead of starting with the total $\mathbf{E}$ as in (9), I start with $\mathbf{e}$ and then dot with $\mathbf{b}$ and average, the analagous derivation replaces (13) by

$$\partial_\mu\overline{H}^\mu(\mathbf{b}) = -2c\langle\mathbf{e}\cdot\mathbf{b}\rangle, \tag{14}$$

where $\overline{H}^\mu(\mathbf{b})$ indicates the average of $H^\mu(\mathbf{b})$. The latter is defined like (12) but with the corresponding fluctuating quantities replacing the total quantities. Similarly, starting with $\overline{\mathbf{E}}$ and dotting with $\overline{\mathbf{B}}$, gives

$$\partial_\mu H^\mu(\overline{\mathbf{B}}) = -2c\overline{\mathbf{E}}\cdot\overline{\mathbf{B}}, \tag{15}$$

where $H^\mu(\overline{\mathbf{B}})$ is defined as in (12) but with the corresponding mean quantities replacing the total quantities.

In the next three subsections, I consider the implications of magnetic helicity conservation for 3 separate cases. In the first I assume a steady state and ignore boundary terms. In the second, I include boundary terms but still demand a steady state. The third is the case in which boundaries are ignored, but a full time evolution of α is considered and a fully dynamical theory is obtained that agrees with recent numerical simulations.

2.1 Case 1: Homogeneous, Stationary Turbulence in Periodic Box

Consider statistically stationary turbulence, where the averaging is over periodic boundaries. Then, the spatial divergence terms on the left of (13) become surface integrals and vanish. Discarding the spatial divergenece terms in (14) gives

$$\langle\mathbf{e}\cdot\mathbf{b}\rangle = -\frac{1}{2c}\partial_t\langle\mathbf{a}\cdot\mathbf{b}\rangle. \tag{16}$$

In the steady state, (16) vanishes and (8) shows that $\langle\mathbf{v}\times\mathbf{b}\rangle$ is resistively limited.

If a uniform $\overline{\mathbf{B}}$ is imposed over the periodic box, then $\overline{\mathbf{B}}$ cannot change with time, and has no gradients. This is the case of [25], which measures α but cannot test for MFD action. No mean quantity varies on long time scales. In this case, (8) then implies

$$-c^{-1}\langle \mathbf{v}\times\mathbf{b}\rangle\cdot\overline{\mathbf{B}} = \alpha\overline{\mathbf{B}}^2/c = c^{-1}\lambda\langle \mathbf{b}\cdot\nabla\times\mathbf{b}\rangle\,, \tag{17}$$

where $\alpha=\alpha_{33}$ for a uniform field in the z direction. Rearranging gives

$$|\alpha| = \left|\frac{\lambda\langle \mathbf{b}\cdot\nabla\times\mathbf{b}\rangle}{\overline{\mathbf{B}}^2}\right|. \tag{18}$$

We then have

$$|\alpha| \lesssim \left|\frac{k_2 b_2^2\tau_2}{R_{M,2}^n\overline{\mathbf{B}}^2/v_2^2}\right| \sim \left|\frac{\alpha_0}{R_{M,2}^n\overline{\mathbf{B}}^2/b_2^2}\right| \tag{19}$$

where b_2 and v_2 are the fluctuating field magnetic energy of the dominant energy containing eddies (which is the forcing scale for both v and b for $f_h=1$ turbulence as described in Sect. 1), k_2 is the wavenumber for that scale (= the forcing scale), and $n=3/4$ if the current helicity is dominated by large wavenumber and $n=1$ if it is dominated by small wavenumbers. The quantity τ_2 is the associated eddy turnover time and $R_{M,2}=v_2/k_2\nu_M$ is the magnetic Reynolds number associated with the forcing scale. Assuming a steady state, I used $v_2\sim b_2$ which is roughly consistent with numerical results. Then assuming forcing with maximal kinetic helicty, I replaced the numerator with α_0, the maximum possible value of a helical quantity of that dimension.

Pouquet Correction and Connection to Previous Studies: If we now take α to be of the form first derived by [31], (discussed further in Sect. 2.3) in the context of a maximally helical, force free α^2 dynamo in a periodic box, we have

$$\alpha = -(1/3)\tau(\langle \mathbf{v}\cdot\nabla\times\mathbf{v}\rangle - \langle \mathbf{b}\cdot\nabla\times\mathbf{b}\rangle), \tag{20}$$

where τ is the correlation time of the turbulence at k_2. Consider the implications of this formula in the steady state. If we use (16) in (8), and allow for a non-uniform mean field it is straightforward to show that α can be written

$$\alpha = \frac{\alpha_0 + R_{M,2}\beta\langle\overline{\mathbf{B}}\cdot\nabla\times\overline{\mathbf{B}}\rangle/B_{eq}^2}{1+R_{M,2}\overline{B}^2/B_{eq}^2}. \tag{21}$$

Note that brackets around the current helicity are present because we allow for the fact that the mean field can have a scale smaller than the overall system (or box) scale. For example, the growth of the mean field at wavenumber $k=1$ in a periodic box can occur when the $k=0$ field is zero. Indeed, the limit of

this equation for a mean field of zero curl produces exactly the result obtained numerically by [25], result for uniform $\overline{\mathbf{B}}$. However, as is clear from this formula, it only emerges when $\overline{\mathbf{B}}$ is uniform and the system is in a steady state.

For steady-state but non-uniform $\overline{\mathbf{B}}$ in the simplest case of a maximally helical α^2 dynamo in periodic box with no shear, the field energy and large-scale helicity sustenance depend on $\overline{\mathbf{E}} \cdot \overline{\mathbf{B}}$. If we assume for example that $\beta = \beta_0$ =constant (unquenched, one extreme limit) and ignore mean velocities, and use (21)

$$-\langle \overline{\mathbf{E}} \cdot \overline{\mathbf{B}} \rangle = \alpha \langle \overline{B}^2 \rangle - \beta \langle \overline{\mathbf{B}} \cdot \nabla \times \overline{\mathbf{B}} \rangle = \frac{\alpha_0 + R_{M,2}\beta_0 \langle \overline{\mathbf{B}} \cdot \nabla \times \overline{\mathbf{B}} \rangle / B_{eq}^2}{1 + R_{M,2}\overline{B}^2/B_{eq}^2} \langle \overline{B}^2 \rangle - \beta_0 \langle \overline{\mathbf{B}} \cdot \nabla \times \overline{\mathbf{B}} \rangle$$
$$= \frac{\alpha_0}{1 + R_{M,2}\langle \overline{B}^2 \rangle / B_{eq}^2} \langle \overline{B}^2 \rangle - \frac{\beta_0}{1 + R_{M,2}\langle \overline{B}^2 \rangle / B_{eq}^2} \langle \overline{\mathbf{B}} \cdot \nabla \times \overline{\mathbf{B}} \rangle, \tag{22}$$

where $B_{eq} = v_2$ in velocity units. Notice that choosing a constant $\beta = \beta_0$ in the steady state, leads to an EMF that is resistively quenched, and is the same as that with an artificially imposed symmetric, resisitive quenching of α and β. These forms of α and β are however misleading in the sense that although their combination is consistent with helicity conservation in the EMF, the division of α and β in this way is not the division which was consistent with our initial assumption of $\beta = \beta_0$ and so their forms are mere artifacts [26]. Actually, in the saturated state, the current helicities of the large and small scales must be equal and opposite [16,30,26] (this follows from (13) with no divergence terms). We then have $\langle \overline{\mathbf{J}} \cdot \overline{\mathbf{B}} \rangle \propto \alpha - \alpha_0$. This implies, for our assumed $\beta = \beta_0$, and for large $R_{M,2}$ that

$$\alpha = \frac{\alpha_0}{1 + \langle \overline{B}^2 \rangle / B_{eq}^2}. \tag{23}$$

This is the actual steady state form of α when β is unquenched. Note that there is no $R_{M,2}$ depednence in the separate forms of α or β even though $\overline{\mathbf{E}} \cdot \overline{\mathbf{B}}$ is resitively limited: in the steady state $\langle \overline{\mathbf{E}} \cdot \overline{\mathbf{B}} \rangle$ must satisfy $\langle \overline{\mathbf{E}} \cdot \overline{\mathbf{B}} \rangle = \nu_M \overline{\mathbf{J}} \cdot \overline{\mathbf{B}} \rangle$. Ultimately, one can solve for $\overline{B}$ to see this demonstrated.

If we were to instead assume $\alpha \propto \beta$ from the outset, one finds

$$\alpha = \frac{\alpha_0}{1 + R_{M,2}(\alpha/\alpha_0 + \langle \overline{B}^2 \rangle / B_{eq}^2 - 1)}. \tag{24}$$

Later we will see a deeper implication of the appearance of the current helicity in (17).

2.2 Case 2: Inhomogeneous Turbulence, Finite Boundary Terms and Implications for Coronal Activity in Steady State

Now consider a system (e.g. Galaxy or Sun) to have a volume much smaller than the universal volume. Integrating (11) over all of space, U, then gives

$$\int_U \mathbf{E} \cdot \mathbf{B} \, d^3x = -(1/2) \int_U \nabla \cdot \Phi \mathbf{B} \, d^3x + (1/2) \int_U \nabla \cdot (\mathbf{A} \times \mathbf{E}) \, d^3x$$
$$-(1/2c)\partial_t \int_U \mathbf{A} \cdot \mathbf{B} \, d^3x = -(1/2c)\partial_t \mathcal{H}(\mathbf{B}) = \int_U \eta \mathbf{J} \cdot \mathbf{B} d^3x, \tag{25}$$

where the divergence integrals vanish when converted to surface terms at infinity. I have defined the global magnetic helicity

$$\mathcal{H}(\mathbf{B}) \equiv \int_U \mathbf{A}\cdot\mathbf{B}\; d^3x, \tag{26}$$

where U allows for scales much larger than the mean field scales. It is easy to show that a parallel argument for the mean and fluctuating fields leads to

$$\partial_t \mathcal{H}(\overline{\mathbf{B}}) = \partial_t \int_U \overline{\mathbf{A}}\cdot\overline{\mathbf{B}}\; d^3x = -2c\int_U \overline{\mathbf{E}}\cdot\overline{\mathbf{B}}\; d^3x \tag{27}$$

and

$$\partial_t \overline{\mathcal{H}}(\mathbf{b}) = \partial_t \int_U \langle \mathbf{a}\cdot\mathbf{b}\rangle\; d^3x = -2c\int_U \langle \mathbf{e}\cdot\mathbf{b}\rangle\; d^3x = -2c\int_U \mathbf{e}\cdot\mathbf{b}\; d^3x = \partial_t \mathcal{H}(\mathbf{b}), \tag{28}$$

where the penultimate equality in (28) follows from redundancy of averages.

I now split (27) and (28) into contributions from inside and outside the rotator. One must exercise caution in doing so because $\mathcal{H}$ is gauge invariant and physically meaningful only if the volume U over which $\mathcal{H}$ is integrated is bounded by a magnetic surface (i.e. normal component of $\mathbf{B}$ vanishes at the surface), whereas the surface separating the outside from the inside of the rotator is not magnetic in general. Reference [34] shows how to construct a revised gauge invariant quantity called the relative magnetic helicity,

$$\mathcal{H}_{R,i}(\mathbf{B}_i) = \mathcal{H}(\mathbf{B}_i, \mathbf{P}_o) - \mathcal{H}(\mathbf{P}_i, \mathbf{P}_o) \tag{29}$$

where the two arguments represent inside and outside the body respectively, and $\mathbf{P}$ indicates a potential field. The relative helicity of the inside region is thus the difference between the actual helicity and the helicity associated with a potential field inside that boundary. The use of $\mathbf{P}_i$ is not arbitrary in (29), and is in fact the field configuration of lowest energy. While (29) is insensitive to the choice of external field [34], it is most convenient to take it to be a potential field as is done in (29) symbolized by $\mathbf{P}_o$. The relative helicity of the outer region, $\mathcal{H}_{R,o}$, is of the form (29) but with the o's and i's reversed. The $\mathcal{H}_R$ is invariant even if the boundary is not a magnetic surface.

The total global helicity, in a magnetically bounded volume divided into the sum of internal and external regions, $U = U_i + U_e$, satisfies [34]

$$\mathcal{H}(\mathbf{B}) = \mathcal{H}_{R,o}(\mathbf{B}) + \mathcal{H}_{R,i}(\mathbf{B}), \tag{30}$$

when the boundary surfaces are planar or spherical. This latter statement results from the vanishing of an additional term associated with potential fields that would otherwise appear in (30). Similar equations apply for $\overline{\mathbf{B}}$ and $\mathbf{b}$, so (27) and (28) can be written

$$\partial_t \mathcal{H}(\overline{\mathbf{B}}) = \partial_t \mathcal{H}_{R,o}(\overline{\mathbf{B}}) + \partial_t \mathcal{H}_{R,i}(\overline{\mathbf{B}}), \tag{31}$$

and

$$\partial_t \mathcal{H}(\mathbf{b}) = \partial_t \mathcal{H}_{R,o}(\mathbf{b}) + \partial_t \mathcal{H}_{R,i}(\mathbf{b}) \tag{32}$$

respectively. According to equation (62) of [25],

$$\partial_t \mathcal{H}_{R,i}(\mathbf{B}) = -2c \int_{U_i} \mathbf{E} \cdot \mathbf{B} d^3x + 2c \int_{DU_i} (\mathbf{A}_p \times \mathbf{E}) \cdot d\mathbf{S}, \tag{33}$$

where $\mathbf{A}_p$ is the vector potential corresponding to a potential field $\mathbf{P}$ in U_e, and DU_i indicates surface integration. Similarly,

$$\partial_t \mathcal{H}_{R,i}(\overline{\mathbf{B}}) = -2c \int_{U_i} \overline{\mathbf{E}} \cdot \overline{\mathbf{B}} d^3x + 2c \int_{DU_i} (\overline{\mathbf{A}}_p \times \overline{\mathbf{E}}) \cdot d\mathbf{S} \tag{34}$$

and

$$\partial_t \mathcal{H}_{R,i}(\mathbf{b}) = -2c \int_{U_i} \mathbf{e} \cdot \mathbf{b} d^3x + 2c \int_{DU_i} \mathbf{a}_p \times \mathbf{e} \cdot d\mathbf{S}. \tag{35}$$

Note again that the above internal relative helicity time derivatives are both gauge invariant and independent of the field assumed in the external region. If we were considering the relative helicity of the external region, that would be independent of the actual field in the internal region.

Now if I take the average over a region $\leq U_i$, I then replace (35) by

$$\langle \mathbf{e} \cdot \mathbf{b} \rangle = -\frac{1}{2c} \partial_t \mathcal{H}_{R,i}(\mathbf{b}) + \langle \nabla \cdot (\mathbf{a}_p \times \mathbf{e}) \rangle, \tag{36}$$

where the brackets indicate integrating over U_i or smaller. We now see that even if the first term on the right of (36) vanishes, $\langle \mathbf{e} \cdot \mathbf{b} \rangle$ contributes a surface term to (8) that need not vanish. The turbulent EMF is not explicitly resistively limited as in the previous section. since the surface term can dominate. Thus in a steady state for $R_M \gg 1$, an outflow of magnetic helicity may enable fast MFD action. Since the right of (13) is small for large R_M, the magnetic helicity flux has contributions from the small and large-scale field.

Note that the sign of the pseudo-scalar α coefficient changes across the mid-plane of an astrophysical rotator. This means that the magnetic helicity generation in one hemisphere is of opposite sign to that in the opposing hemisphere. One can then imagine that helicity flow could take place across the mid-plane; the loss of helicity from say the top hemisphere into the bottom, acts in congruence with the loss of the opposite sign of helicity from the bottom hemisphere into the top [45] allowing rapid MFD action. Even if there is not strong coupling at the mid-plane between the two hemispheres, an enhanced diffusion at the mid-plane side boundary of each turbulent/convective region could in principle dissipate the helicity require to allow a rapid MFD and a rapid net flux generation.

But there are reasons why a fast MFD in stars and galaxies would more likely involve escape of magnetic helicity out of the external boundaries in addition to merely a redistribution across the mid-plane. First, buoyancy is important for both disks and stars and the Coriolis force naturally produces helical field

structures in rising loops. Second note that the mean Galactic field appears to be quadrupole. This already requires some diffusion at the upper and lower boundaries of the disk [7]. Third, the solar cycle involves a polarity reverse of the mean dipole field and this requires the escape of magnetic fields out through the solar surface. Fourth, note that for the Sun, the most successful dynamo models seem to be interface models [37] where diffusion just below the base of the convection zone is reduced, with no alternative transport mechanism downward. Finally, note that some aspects of solar magnetic helicity loss have been studied [38] and observed [35,36].

Let us explore some implications of depositing relative magnetic helicity to a astrophysical corona. I assume that the rotator is in a steady state over the time scale of interest, so the left sides of (34) and (35) vanish. For the sun, where the mean field flips sign every ~ 11 years, the steady state is relevant for time scales less than this period, but greater than the eddy turnover time ($\sim 5 \times 10^4$ sec). Beyond the ~ 11 year times scales, the mean large and small-scale relative helicity contributions need not separately be steady and the left hand sides need not vanish.

The helicity supply rate, represented by the volume integrals (second terms of (34) and (35)), are then equal to the integrated flux of relative magnetic helicity through the surface of the rotator. Moreover, from (13), we see that the integrated flux of the large-scale relative helicity, $\equiv \mathcal{F}_{R,i}(\overline{\mathbf{B}})$, and the integrated flux of small-scale relative helicity, $\equiv \mathcal{F}_{R,i}(\mathbf{b})$, are equal and opposite. We thus have

$$\mathcal{F}_{R,i}(\overline{\mathbf{B}}) = -2c \int_{U_i} \overline{\mathbf{E}} \cdot \overline{\mathbf{B}} d^3x = 2c \int_{U_i} \mathbf{e} \cdot \mathbf{b} d^3x = -\mathcal{F}_{R,i}(\mathbf{b}). \tag{37}$$

To evaluate this, I use (1) and (2) to find

$$\overline{\mathbf{E}} = -c^{-1}(\alpha \overline{\mathbf{B}} - \beta \nabla \times \overline{\mathbf{B}}), \tag{38}$$

throughout U_i. Thus

$$\mathcal{F}_{R,i}(\overline{\mathbf{B}}) = -\mathcal{F}_{R,i}(\mathbf{b}) = 2 \int_{U_i} (\alpha \overline{\mathbf{B}}^2 - \beta \overline{\mathbf{B}} \cdot \nabla \times \overline{\mathbf{B}}) d^3\mathbf{x}. \tag{39}$$

This shows the relation between the equal and opposite large and small-scale relative helicity deposition rates and the dynamo coefficients.

Now the realizability of a helical magnetic field requires its turbulent energy spectrum, E_k^M, to satisfy [39]

$$E_k^M(\mathbf{b}) \geq \frac{1}{8\pi} k |\mathcal{H}_k(\mathbf{b})| \ , \tag{40}$$

where $\mathcal{H}_k$ is the magnetic helicity at wavenumber k. The same argument also applies to the mean field energy spectrum, so that

$$E_k^M(\overline{\mathbf{B}}) \geq \frac{1}{8\pi} k |\mathcal{H}_k(\overline{\mathbf{B}})|. \tag{41}$$

If I assume that the time and spatial dependences are separable in both E^M and $\mathcal{H}$, then an estimate of power delivered to the corona can be derived. I presume that the change in energy associated with the helicity flow represents an outward deposition rather than an inward deposition. This needs to be specifically calculated for a given rotator and dynamo, but since the source of magnetic energy is the rotator, and buoyancy and reconnection interplay to transport magnetic energy outward, the assumption is reasonably motivated.

For the contribution from the small-scale field, we have

$$\dot{E}^M(\mathbf{b}) = \int \dot{E}^M_k(\mathbf{b})dk \geq \tfrac{1}{8\pi}\int k|\mathcal{F}_{k,R,i}(\mathbf{b})|\,dk \geq \tfrac{k_{\min}}{8\pi}\int |\mathcal{F}_{k,R,i}(\mathbf{b})|\,dk \geq \tfrac{k_{\min}}{8\pi}|\mathcal{F}_{R,i}(\mathbf{b})| = \tfrac{k_{\min}}{8\pi}|\mathcal{F}_{R,i}(\overline{\mathbf{B}})|,$$

where the last equality follows from the first equation in (39). The last quantity is exactly the lower limit on $\dot{E}^M(\overline{\mathbf{B}})$. Thus the sum of the lower limits on the total power delivered from large and small scales is $\frac{k_{\min}}{8\pi}|\mathcal{F}_{R,i}(\overline{\mathbf{B}})| + \frac{k_{\min}}{8\pi}|\mathcal{F}_{R,i}(\mathbf{b})| = 2\frac{k_{\min}}{8\pi}|\mathcal{F}(\overline{\mathbf{B}})|$. Now for a mode to fit in the rotator, $k > k_{\min} = 2\pi/h$, where h is a characteristic scale height of the turbulent layer. Using (39), the total estimated energy delivered to the corona (=the sum of the equal small and large-scale contributions) is then

$$\dot{E}^M \geq 2\frac{k_{\min}}{8\pi}|\mathcal{F}_{R,i}(\mathbf{b})| = 2\frac{k_{\min}}{8\pi}|\mathcal{F}_{R,i}(\overline{\mathbf{B}})| = \frac{\mathrm{V}}{h}\left|\alpha\overline{B}^2 - \beta\overline{\mathbf{B}}\cdot\nabla\times\overline{\mathbf{B}}\right|_{ave}, \quad (42)$$

where V is the volume of the turbulent rotator and *ave* indicates volume averaged. I will assume that the two terms on the right of (42) do not cancel, and use the first term of (42) as representative.

Working in this allowed time range for the Sun ($5\times10^4\mathrm{sec} < t < 11\mathrm{yr}$) I apply (42) to each Solar hemisphere. Using the first term as an order of magnitude estimate gives

$$\dot{E}^M \gtrsim \left(\frac{2\pi R_\odot^2}{3}\right)\alpha\overline{\mathbf{B}}^2 = 10^{28}\left(\frac{R}{7\times10^{10}\mathrm{cm}}\right)^2\left(\frac{\alpha}{40\mathrm{cm/s}}\right)\left(\frac{\overline{B}}{150\mathrm{G}}\right)^2\frac{\mathrm{erg}}{\mathrm{s}}. \quad (43)$$

I have taken $\alpha \sim 40\,\mathrm{cm\ s^{-1}}$ (a low value of α) and have presumed a field of 150G at a depth of 10^4km beneath the solar surface in the convection zone, which is in energy equipartition with turbulent kinetic motions [7].

As this energy is available for reconnection, Alfvén waves, winds, and particle energization, we must compare this limit with the total of downward heat conduction loss, radiative loss, and solar wind energy flux in coronal holes, which cover $\sim 1/2$ the area of the Sun. According to [40], this amounts to an approximately steady activity of $2.5\times10^{28}\mathrm{erg\ s^{-1}}$, about 3 times the predicted value of (43). There is other evidence for deposition of magnetic energy and magnetic + current helicity in the Sun [35,36,41].

Active galactic nuclei (AGN) and the Galactic interstellar medium (ISM) represent other likely sites of mean field dynamos [41]. For the Galaxy, $\dot{E}^M \gtrsim (\pi R^2)\alpha\overline{B}^2 \sim 10^{40}(R/12\mathrm{kpc})^2\times(\alpha/10^5\mathrm{cm/s})(\overline{B}/5\times10^{-6}\mathrm{G})^2\mathrm{erg/s}$ in each hemisphere. This is consistent with coronal energy input rates required by [42] and

[43]. For AGN accretion disks, the deposition rate seems to be consistent with what is required from X-ray observations. Independent of the above, the most successful paradigm for X-ray luminosity in AGN is coronal dissipation of magnetic energy [44].

So boundary terms can potentially alleviate any helicity constraint and allow fast steady dynamo action. But how this specifically happens and the specific boundary physics of a real system will need more study to see what field wave numbers, if any, are preferentially shed.

2.3 Case 3: Time-Dependent Dynamo Action and Dynamical Quenching in a Periodic Box

For the question of actual field amplification and time scales ultimately associated with cycle periods for a real dynamo, a time dependent, non-linear dynamical theory is required. Here the time derivative term of (13) important. (Note also that time dependent dynamo like effects are also important for magnetic field adjustment in Reverse Field Pinches [45].)

An important step forward was the work of [31], based on the Eddy Damped Quasi-Normal Markovian spectral closure scheme. There an approximate set of equations describing the evolution of magnetic and kinetic energy and helicity was derived. References [14,16] performed numerical simulations of the process, and [30] further simplified the equations of [31] by considering a two-scale approach for fully helically forced turbulence. In [30] it was assumed that the large-scale field grows primarily on scale k_1 and the small-scale turbulent field is peaked at $k_2 >> k_1$. The analytic model therein is largely consistent with the forced helical turbulence, periodic box simulations of [14,16]. It should be mentioned however, that while the large-scale field dynamics are consistent with a single scale k_1 at all times, the wavenumber of the small-scale peak seems to migrate from the resistive scale to the forcing scale, in which case the two-scale approach applies but with a migrating k_2 [16,26] unlike that considered in [30]. More work is needed to understand the migration of the small-scale peak. For present purposes we ignore this complication which does not effect the accuracy of the resulting fits of the large-scale field growth.

The basic concept of the successful dynamical quenching model [30] for the α^2 dynamo is that the growth of the large-scale field is the result of a segregation of magnetic helicity. Magnetic helicity of one sign grows on the large scale, while the opposite sign grows on the small scale, up to resistively limited conservation. The growth of the small-scale magnetic helicity also grows current helicity which suppresses α. The two essential equations needed in this analysis are (1) the equation for large-scale magnetic helicity evolution (15) and (2) the equation for total magnetic helicity evolution (13). For more general dynamos with shear, the equations of the dynamical theory are (1) the vector equation for the large-scale magnetic field, and (2) the equation for total magnetic helicity conservation. In general, the paradigm that emerges is that the total magnetic helicity conservation acts as a supplementary dynamical equation that is coupled to the evolution of the large-scale field equation.

Ignoring the boundary terms in (15). and subscript 1 and 2 to indicate mean and fluctuating scales, and M to indicate "magnetic" we then define $H^M \equiv H^0(\overline{\mathbf{B}})/2$. Thus (15) implies

$$\partial_t H_1^M = 2\alpha E_1^M - 2(\lambda+\beta)\overline{\mathbf{J}}\cdot\overline{\mathbf{B}}, \tag{44}$$

where where $E_1^M = \overline{\mathbf{B}}^2/2$. Now we replace spatial derivatives of the mean scale with k_1, and then note that the current helicity term on the right is related to the magnetic helicity $\overline{\mathbf{J}}\cdot\overline{\mathbf{B}} = k_1^2 H_1^M$. We then have

$$\partial_t H_1^M = 2\alpha E_1^M - 2(\lambda+\beta)k_1^2 H_1^M\ . \tag{45}$$

Reference [30] describes the conditions for which this is consistent with the analogous equation that arises from the spectral treatment of [31].

The second equation we need is a re-write of (13). Since the left-most side is the sum of contribution from large and small scales, as is the right-most side, we have in the two-scale approximation

$$\partial_t H_1^M + \partial_t H_2^M = -2\lambda k_1^2 H_1^M - 2\lambda k_2^2 H_2^M. \tag{46}$$

We also need a prescription for α and β. A configuration space approach for α and its pitfalls are described in the next section, but for now, let us extract the dynamical form derived from [31]. For α this is

$$\alpha = -(1/3)\tau(\langle\mathbf{v}\cdot\nabla\times\mathbf{v}\rangle - \langle\mathbf{b}\cdot\nabla\times\mathbf{b}\rangle), \tag{47}$$

where τ is a correlation time of the turbulence at k_2. In externally forced simulations, the first term on the right of (47) (the kinetic helicity) is typically maintained at a fixed value. At early times, the second term on the right is small. Thus the second term on the right (current helicity) can be thought of as a backreaction on the first term. It arises from inclusion of the Navier-Stokes equation. The idea that the second term represents a backreaction was investigated by [46,48] for a stationary system and a correction to α was derived. However, here I will derive the fully dynamical correction following [30]. First re-write (47) as

$$\alpha = \alpha_0 + \tau k_2 H_2^M/3. \tag{48}$$

If we assume that the kinetic helicity is forced maximally, and take $\tau = 2/k_2 v_2$, then $\alpha_0 = 2v_2/3$.

Unfortunately, there is not yet convergence on a rigorous prescription for β in 3-D, but I will consider the two cases discussed earlier, $\beta = \beta_0\alpha/\alpha_0$ and $\beta = \beta_0 \equiv v_2/k_2$=constant as examples. In [26] other prescriptions are considered, including those for which β is quenched but not resistively. In general, the need for the appropriate form of β is an important input, and determines ultimately the form of saturated α in the α^2 dynamo because the two are constrained through $\overline{\mathbf{E}}\cdot\overline{\mathbf{B}}$ as discussed at the end of Sect. 2.1. However, the particular form of β is also less important for illustrating the role of helicity conservation in

the α^2 dynamo than in dynamos with shear as for the former since a range of choices are all relatively successful. I will show that the difference for large k_2/k_1 between choices of β is really quite minimal. More prescriptions for β are considered in [26].

The same type of formalism which leads to the prescription for α, produces a prescription for β involving an approximate *sum* of kinetic + magnetic energies times a correlation time, at least to first order, sort of motivating the case $\beta = \beta_0$. More work is ongoing and indeed for very strong mean fields since β must at least respond to the mean Lorentz force. The case $\beta = \beta_0\alpha/\alpha_0$ is motivated by the fact that at late times one empirical combination of formulae which fit [16] has this relation.

It is important to emphasize though, that as shown in Sect. 2.1, a misleading degeneracy emerges in the steady-state (or near steady state which amounts to the late time evolution) with respect to the choice of β. Again, the reason is that for α^2 dynamos, it is really the $\langle \overline{\mathbf{E}} \cdot \overline{\mathbf{B}} \rangle$ that matters for the growth of the large-scale magnetic helicity, and thus the large-scale field. The main effect of the prescription for β is the saturation value of the field.

Following [30], we need to solve (45) and (46). To do so, I rewrite them in dimensionless form. Define the dimensionless magnetic helicities $h_1 \equiv 2H_1^M k_2/v_2^2$ and $h_2 \equiv 2H_2^M k_2/v_2^2$ and write time in units of $1/k_2 v_2$. I also define $R_M \equiv (v_2/k_1)/\nu_M$. (Note that this definition of R_M is not based on the forcing-scale RMS velocity but on the large scale, k_1^{-1}. We will later employ the previously defined magnetic Reynolds number $R_{M,2} \equiv R_M(k_1/k_2)$.)

Using the above scalings we can replace (45) and (46) with dimensionless equations given by

$$\partial_t h_1 = \frac{4}{3}\left(\frac{k_1}{k_2}\right) h_1(1+h_2) - 2h_1\left[\frac{k_1}{k_2 R_M} + \frac{k_1^2}{k_2^2}(1+q_2 h_2)\right] \tag{49}$$

and

$$\partial_t h_2 = -\frac{4}{3}\left(\frac{k_1}{k_2}\right) h_1(1+h_2) + 2h_1\frac{k_1^2}{k_2^2}(1+q_2 h_2) - \frac{2}{R_M}\frac{h_2 k_2}{k_1},$$

where $q_2 = 0$ in the above equations corresponds to $\beta(t) = \beta_0$ =constant. and $q_2 = 1$ corresponds to $\beta(t) = \alpha(t)\beta_0/\alpha_0$. Solutions of these coupled equations are shown in Figs. 3-8, which are taken from [30]. The key parameters are k_2/k_1, R_M, and q_2. In the figures, I have compared these results to the empirical fits of numerical simulations in [16]. I used $h_1(t=0) = 10^{-3}$, but the sensitivity to $h_1(0)$ is only logarithmic (see (54) below). In Fig. 1, $k_2/k_1 = 5$ was used, following B01, and in Fig. 2 $k_2/k_1 = 20$ was used.

In the figures, the solid lines represent numerical solutions to (49) and (50), whereas the dotted lines represent the formula given in [16], which is an empirical fit to simulation data assuming that α and β are prescribed according to (52) and (53) below. More explicitly, [16] found that the growth of $\overline{\mathbf{B}}$ was well described

by the formula

$$\frac{B_1^2/B_{1,0}^2}{\left(1-B_1^2/B_{1,sat}^2\right)^{1+\frac{\alpha_0 k_1-k_1^2\beta_0}{\nu_M k_1^2}}} = e^{2(\alpha_0 k_1 - k_1^2\beta_0)t}, \tag{50}$$

where $B_{1,0} = B_1(t=0)$. This can be rewritten using the notation above as a dimensionless equation for t in units of $(k_2 v_2)^{-1}$, namely

$$t = \frac{k_2}{2k_1}\frac{\mathrm{Ln}[(h_1/h_0)(1-h_1 k_1^2/k_2^2)^{R_M(k_1/k_2-2/3)-1}]}{2/3-k_1/k_2}. \tag{51}$$

Note that (50) and (51) correspond to α and β quenching of the form

$$\alpha = \frac{\alpha_0}{1+s_B B_1^2/v_2^2} \tag{52}$$

and

$$\beta = \frac{\beta_0}{1+s_B B_1^2/v_2^2}, \tag{53}$$

where $s_B \sim R_M(k_1/k_2)(2/3-k_1/k_2) = R_{M,2}(2/3-k_1/k_2)$, and $R_{M,2} \equiv v_2/k_2\nu_M$. Equations (52) and (53) are derived from those in [16] by re-scaling (55) of [16] with the notation herein. It can also be shown directly that, up to terms of order $1/R_M$, (51) is consistent with that derived by substituting (52) and (53) into (49) and solving for t. Note that in contrast to the suggestion of [16], it is actually the forcing-scale magnetic Reynolds number, $R_{M,2}$, that plays a prominent role in these formulae.

The solutions of (49) and (50) are interesting. Insight can be gained by their sum

$$\partial_t h_1 + \partial_t h_2 = -\frac{2}{R_M}\left(\frac{h_1 k_1}{k_2}+\frac{h_2 k_2}{k_1}\right),$$

which corresponds to (46), the conservation of total magnetic helicity. If we make the astrophysically relevant assumption that $R_M >> 1$, the right hand side of (46) is small for all h_1 and h_2. It follows that $\partial_t(h_1+h_2)=0$ and for $h(t=0)=0$, this implies $h_2 = -h_1$. In this period, we can self-consistently ignore $1/R_M$ in (49). If $q_2 = 1$, this phase ends when $h_2 \to -1$, so that $h_1 \sim 1$. This is manifested in figure 5.

This kinematic phase precedes the asymptotic saturation of the dynamo investigated by other authors, in which all time derivatives vanish exactly. For this to happen, the right hand side of (54) must vanish, which is equivalent to demanding that $h_2 = -(k_1/k_2)^2 h_1$. Since the right hand sides of (49) and (50) are proportional to $1+h_2$ when terms of order $1/R_M$ are neglected, their vanishing requires that $h_2 = -1$, and therefore, that $h_1 = (k_2/k_1)^2$. This is observed in figures 1 and 2. The asymptotic saturation (when the field growth ceases) takes a time of order $t_{sat} \sim R_M k_2/k_1$, which in astrophysics is often huge. Thus,

although in principle it is correct that α is resistively limited (as seen from the solutions in figures 6 & 7) as suggested by [23,25,46,48], this is less important than the fact that for a time $t_{kin} < R_M$ the kinematic value of α applies. The time scale t_{kin} here is given by a few kinematic growth time scales for the α^2 dynamo, more specifically,

$$t_{kin} \sim \mathrm{Ln}[1/h_1(0)](k_2/k_1)/(4/3 - 2k_1/k_2). \tag{54}$$

For $h_1(0) = 0.001$, $k_2/k_1 = 5$, $t_{kin} \sim 37$, as seen in Fig. 5.

Note that t_{kin} is sensitive to k_2/k_1 and independent of R_M. Figure 5 shows that there is significant disagreement in this regime with (52), but this formula was used in [16] only to model the regime $t > R_M$, so the result is not unexpected. We can see from the solution for α itself that indeed the solutions of (49) and (50) do match (52) for $t > R_M$ (figures 6 and 7). Figure 6 shows the difference in the α along with (53) for the two values $R_M = 10^2$ and $R_M = 10^3$. Notice again the disagreement with the formula (52) until $t = R_M$, and agreement afterward. This marks the time at which the resistive term on the right of (49) becomes competitive with the terms involving $(1 + h_2)$. Asymptotic saturation does not occur until $t \sim t_{sat} = R_M k_2/k_1$ as described above.

Finally, note that $q_2 = 0$ corresponds to $\beta = \beta_0$. In general, this leads to a lower value of h_1 in the asymptotic saturation phase because this enforces zero saturation of β, whereas there is still some saturation of α in this limit. (Note that $q_2 = 0$ corresponds to the case of [46] discussed in appendix B of [30].) For large k_2/k_1 the solutions of (49) and (50) are insensitive to $q_2 = 0$ or $q_2 = 1$. This is because the larger k_2/k_1, the smaller the influence of the q_2 terms in (49) and (50). This is highlighted in figure 8 where the result for $q_2 = 0$ is plotted with the [16] fit. This suggests that for large-scale separation, the magnetic energy saturation is insensitive to the form of β quenching. However, in real dynamos, magnetic flux and not just magnetic energy may be needed, so the insensitivity can be misleading because β is needed to remove flux of the opposite sign. From the low k_2/k_1 cases, it is clear that $q_2 = 1$ is a better fit to the simulations of [16].

The physical picture of the quenching process just described is this: helical turbulence is forced at k_2 (= 5 in [16]), and kept approximately constant by forcing. Hence $\alpha_0 = -2\tau H_2^V/3 = const.$ If H_1^M, the magnetic helicity at k_1 (which reaches 1 here as a result of boundary conditions), is initially small — so that $|2k_1^2 H_1^M/3| \ll |\alpha_0|$, (45) (or (49)) shows that it will be exponentially amplified provided that the damping due to $\beta+\lambda$ does not overcome the α effect. Initially, $\alpha = \alpha_0$, acting like a pump that moves magnetic helicity from k_2 to k_1 and driving the dynamo. This kinematic phase lasts until t_{kin} as given by (54). Eventually, the growing H_1^M results in a growing H_2^M of opposite sign, which reduces α through H_2^C. R_M-dependent quenching kicks in at $t = t_{kin}$, but it is not until $t = R_M$ that the asymptotic formulae (52) and (53) are appropriate. Asymptotic saturation, defined by the time at which B_1 approaches its maximum possible value of $(k_2/k_1)^{1/2} v_2$, occurs at $t = t_{sat} = R_M k_2/k_1$. For $t \geq R_M$ the numerical solution of (49) and (50), like the full numerical simulations of [16], is

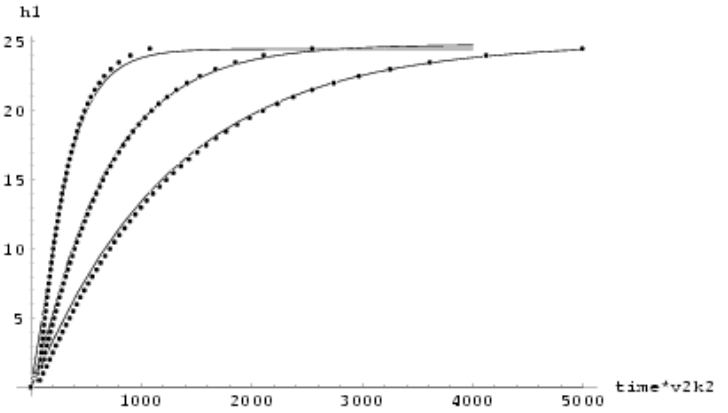


Fig. 3. Solution for $h_1(t)$, $f_h = 1$,$q_2 = 1$. Here $k_2/k_1 = 5$ and the three curves from left to right have $R_M = 100, 250, 500$ respectively. The dots are plotted from the formula used to quasi-empirically fit the simulations in [16]

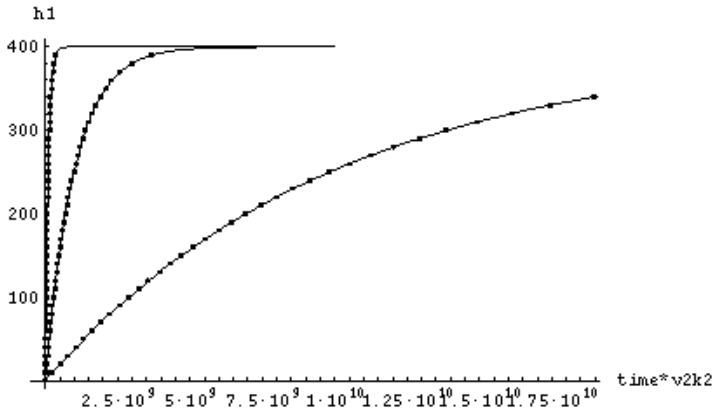


Fig. 4. Solution for $h_1(t)$, $f_h = 1$, $q_2 = 1$. Here $k_2/k_1 = 20$ and the three curves from left to right have $R_M = 10^7, 10^8, 10^9$ respectively. The dots are plotted from the formula for used to quasi-empirically fit simulations of [16]

well fit by the α in (52) with a corresponding β of (53). The two-scale approach is also consistent with B01 in that magnetic helicity jumps from k_2 to k_1 without filling in the intermediate wave numbers.

The emergence of the time scale t_{kin} is interesting because it shows how one can misinterpret the implications of the asymptotic quenching formula (52) and (53). These formulae are appropriate only for $t > R_M$. The large-scale field actually grows kinematically up to a value $B_1 = (k_1/k_2)^{1/2} v_2$ by $t = t_{kin}$ and ultimately up to $B_1 \sim (k_2/k_1)^{1/2} v_2$ by $t = t_{sat}$. For large $R_{M,2}$, these values of B_1 are both much larger than the quantity $v_2/R_{M,2}^{1/2}$, which would have been

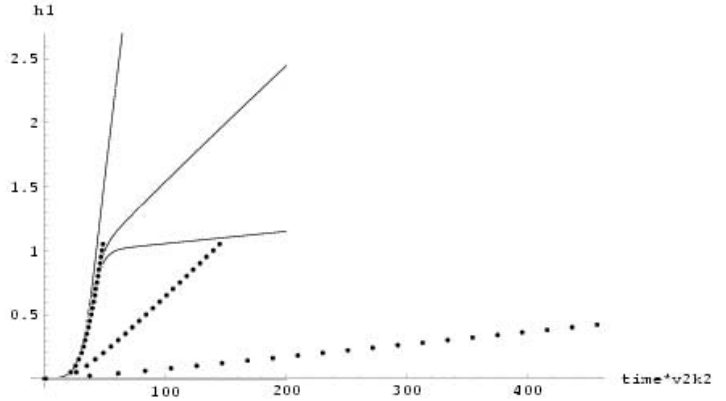


Fig. 5. The early-time solution for $h_1(t)$, $f_h = 1$, $q_2 = 1$. Here for $k_2/k_1 = 5$, and $R_M = 10^2, 10^3, 10^4$ from left to right respectively. Notice the significant departure from the formula of [16] at these early times. For $t < t_{kin}$ there is no dependence on R_M and the growth proceeds kinematically.

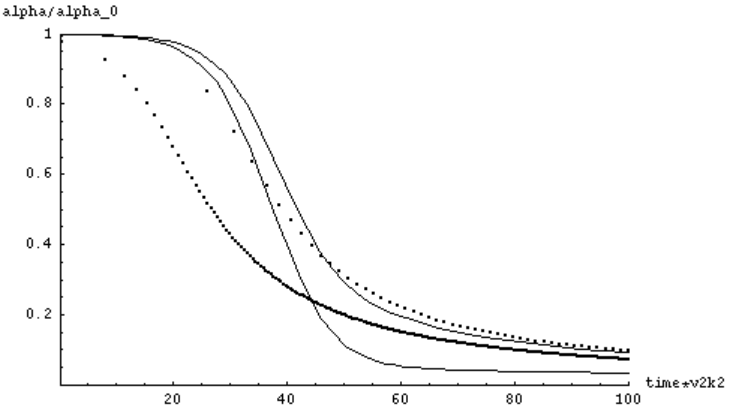


Fig. 6. Solution of $\alpha/\alpha_0(t)$ for $h_1(t)$, $f_h = 1$, $q_2 = 0$. Here $k_2/k_1 = 5$ and the solid lines are the solutions to (49) and (50) for $R_M = 10^2$ (top curve) and $R_M = 10^3$ (bottom curve) respectively. The top and bottom dotted curves are from (52), interpreted from [16]. Notice the long kinematic phase for the solutions, the overshoot, and the convergence with that of (52) at $t = R_M$ for the $R_M = 10^2$ case.

inferred to be the saturation value if one assumed (52) and (53) were valid at all times.

Dynamical quenching or time-dependent approaches recognizing the current helicity as a contributor to α have been discussed elsewhere [8,49,50,51] (see also [45]), but here we have specifically linked the PFL α correction to the helicity conservation in a simple two-scale approach. Other quenching studies for closed systems such as [25] and [48] advocated values of α which are resistively limited and of a form in agreement with (52) but with the assumption of a steady

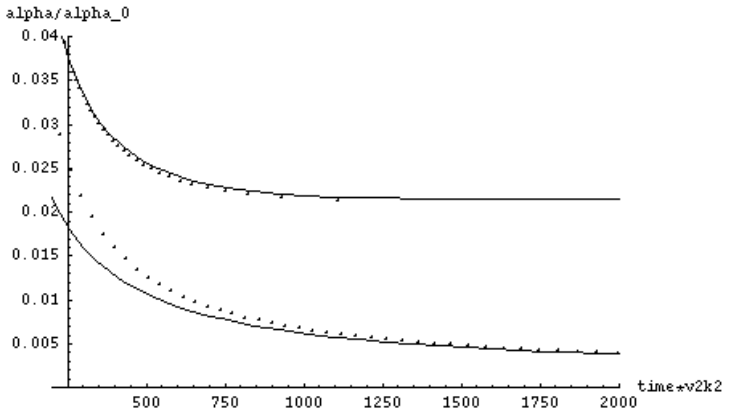


Fig. 7. This is the extension of the previous figure for later times. Notice the convergence of the $R_M = 10^3$ solution to (52) near $t = R_M$.

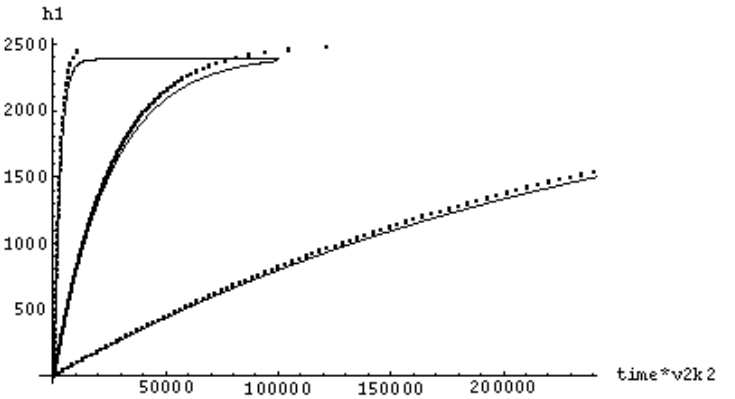


Fig. 8. Solution for $h_1(t)$, $f_h = 1$, $q_2 = 0$. Here $k_2/k_1 = 50$ and the three curves from left to right have $R_M = 10^2, 10^3, 10^4$ respectively. The dotted lines are plotted from the formula used to quasi-empirically fit simulations of [16]. For such large k_2/k_1 the fit to the data is only weakly sensitive to whether $q_2 = 1$ or $q_2 = 0$.

B_1. Assuming (48), and using (45) and (46) in the steady state, their formulae can be easily derived. However, one must also have a prescription for β. If β is proportional to α, then the resistively limited formulae like (24) emerges exactly, which is indeed approximately consistent with (52) and (53) for large R_M. If $\beta(t) = \beta_0$, as in [46], then a formula for α *without* resistively limited quenching (23) emerges. On the other hand, fig. 8 shows that for large k_2/k_1, the dynamo quenching is largely insensitive to β.

Interestingly, if we cavalierly apply these results for the Galaxy (by incorrectly ignoring the shear and assuming an α^2 dynamo that produces force-free large-scale fields), and use $k_2/k_1 = 20$, $B_1(0) = 10^{-9}$, $v_2 = 10$km/s, and $k_2^{-1} \sim 100pc$, we would find the end of the kinematic regime to be at

$t = 140$, or about 1.4×10^9yr. After this stage the field growth would proceed very slowly because of the large R_M, but the saturation field strength at this time is $B_1 \sim v_2/4.5$. Thus quite a large amplification can occur, even with an asymptotically slow dynamo. However this is really an academic exercise since for the Galaxy we need to consider an $\alpha - \Omega$ dynamo, and the boundary terms.

The generalized application of the principle that the magnetic helicity conservations should supplement the mean field dynamo growth equation to account for the backreaction is applied more generally to dynamos with shear in [26].

Finally, note that there is an important puzzle, hidden in the derivations here and those of [46,48] with regard to α quenching that we discuss in the next section.

2.4 Deriving α in Configuration Space: A Puzzle

The two-scale dynamical theory of Sect. 2.3 [30] for α-quenching is appealing because it nicely couples the equations and concepts of magnetic helicity evolution to the current helicity contribution in α, and fits simulation data well. The current helicity contribution was interpreted as a correction to the kinetic helicity contribution of kinematic theory. However, it depends on the current helicity contribution to α as presented in (47) being the total current helicity associated with k_2. To see what we mean by *total* and to show the complication, we consider the derivation of the turbulent EMF in configuration space.

The turbulent EMF can be written in three different ways:

$$\begin{aligned}\langle \mathbf{v} \times \mathbf{b} \rangle &= \langle \mathbf{v}(0) \times \mathbf{b}(0) \rangle + \textstyle\int_0^t \langle \partial_{t'} \mathbf{v}(t') \times \mathbf{b}(t') \rangle dt' + \int_0^t \langle \mathbf{v}(t') \times \partial_{t'} \mathbf{b}(t') \rangle dt' \\ &= \langle \mathbf{v}(t) \times \mathbf{b}(0) \rangle + \textstyle\int_0^t \langle \mathbf{v}(t) \times \partial_{t'} \mathbf{b}(t') \rangle dt' \quad (55) \\ &= \langle \mathbf{v}(0) \times \mathbf{b}(t) \rangle + \textstyle\int_0^t \langle \partial_{t'} \mathbf{v}(t') \times \mathbf{b}(t) \rangle dt'\end{aligned}$$

The three lines in (55) simply correspond to the 3 relevant ways of using the formula $f(t) = f(0) + \int_0^t \partial_{t'} f(t') dt'$, where g is an arbitrary function of time. If I assume that $t \gg 0$, and that widely separated turbulent quantities do not correlate, the first terms on the right of the 2nd and 3rd lines respectively, can be dropped. We then have

$$\langle \mathbf{v} \times \mathbf{b} \rangle = \textstyle\int_0^t \langle \mathbf{v}(t) \times \partial_{t'} \mathbf{b}(t') \rangle dt' = \int_0^t \langle \partial_{t'} \mathbf{v}(t') \times \mathbf{b}(t) \rangle dt'. \quad (56)$$

The second term in (56) as the standard textbook starting point [17,18] for evaluating the turbulent EMF for a kinematic dynamo, but here I have not made any assumptions about the backreaction yet.

To illustrate the point, consider the simple case in which $\nabla \overline{\mathbf{B}} = 0$. The equation for the small-scale field is then

$$\partial_{t'} \mathbf{b} = \overline{\mathbf{B}} \cdot \nabla \mathbf{v} + \mathbf{b} \cdot \nabla \mathbf{v} - \mathbf{v} \cdot \nabla \mathbf{b} - \nabla \times \langle \mathbf{v} \times \mathbf{b} \rangle + \nu_M \nabla^2 \mathbf{b}. \quad (57)$$

The penultimate term goes away when included in (56) and we ignore the last term. The first terms on the right, upon the assumption that the dominant

contributions to correlations are isotropic in $\mathbf{v}$, gives the "textbook" expression for α plus extra terms, that is

$$\langle \mathbf{v} \times \mathbf{b} \rangle \simeq -\frac{\overline{\mathbf{B}}}{3} \int_0^t \langle \mathbf{v}(t) \cdot \nabla \times \mathbf{v}(t') \rangle dt' + Q(v^2 b). \tag{58}$$

The terms symbolized by $Q(v^2 b)$ are typically ignored using some version of the first order smoothing approximation. This is of questionable validity, given that the small-scale field rapidly grows to exceed the mean field. We will come back to the relevance of these terms below.

Now if instead I use the last term of (56) to expand the EMF, I must then invoke the Navier-Stokes equation for the time derivative of the turbulent velocity

$$\partial_t \mathbf{v} = -\mathbf{v} \cdot \nabla \mathbf{v} - \langle \mathbf{v} \cdot \nabla \mathbf{v} \rangle - \nabla p_{eff} + \overline{\mathbf{B}} \cdot \nabla \mathbf{b} + \mathbf{b} \cdot \nabla \mathbf{b} - \langle \mathbf{b} \cdot \nabla \mathbf{b} \rangle + \nu \nabla^2 \mathbf{v} + \mathbf{f}(\mathbf{x}, t), \tag{59}$$

where $\mathbf{f}$ is a forcing function and p_{eff} is the magnetic and thermal pressure. Upon plugging this into (56), the second and sixth terms on the right vanish. If we ignore the viscosity, and assume the dominant contribution to correlations are isotropic, we then have

$$\langle \mathbf{v} \times \mathbf{b} \rangle \simeq \frac{\overline{\mathbf{B}}}{3} \int_0^t \langle \mathbf{b}(t) \cdot \nabla \times \mathbf{b}(t') \rangle dt' + \tilde{Q}(v^2 b, fb) \tag{60}$$

Notice two things about (58) and (60). First they are equal to each other since they were derived from different choices of the expansion of $\langle \mathbf{v} \times \mathbf{b} \rangle$. Second, they do not cleanly include the combination of the total residual helicity, required in (47) when placed into (45) and (46) to derive (49) and (50). The only way that (58) and (60) can have the form of the desired relative helicity is if $Q(v^2 b) = \frac{\overline{\mathbf{B}}}{3} \int_0^t \langle \mathbf{b}(t) \cdot \nabla \times \mathbf{b}(t') \rangle dt'$ and if $\tilde{Q}(v^2 b, fb) = -\frac{\overline{\mathbf{B}}}{3} \int_0^t \langle \mathbf{v}(t) \cdot \nabla \times \mathbf{v}(t') \rangle dt'$. I have been unable to prove that this is the case for the astrophysically relevant weak $\bar{B}$ regime.

There is another approach to calculating the EMF in configuration space that does reveal a similar difference of helicities as that in (47), namely the approach of [52]. But $\mathbf{v}$ and $\mathbf{b}$ enter (47) whereas $\mathbf{v}^{(0)}$ and $\mathbf{b}^{(0)}$, the statistically isotropic parts of $\mathbf{v}$ and $\mathbf{b}$, enter [52]. To see this more explicitly, I write

$$\mathbf{v} = \mathbf{v}^{(0)} + \mathbf{v}^{(A)}, \tag{61}$$

where A indicates an anisotropic contribution, the result of the backreaction from $\overline{\mathbf{B}}$. Similarly,

$$\mathbf{b} = \mathbf{b}^{(0)} + \mathbf{b}^{(A)}. \tag{62}$$

(Even when $\mathbf{b}$ is the result of stirring up an initial seed $\overline{\mathbf{B}}$, there is still a $\mathbf{b}^{(0)}$ which is the statistically isotropic part of $\mathbf{b}$.) We then assume that the statistics of the zeroth order turbulent correlations are those of a homogeneous isotropic, "known" base state. The goal is to express turbulent correlations in terms of the zeroth order quantities. It is sufficient to demonstrate the basic idea invoked

to all orders in $\overline{\mathbf{B}}$ in [52] with that derived to linear order in [55]. Noting that $\langle \mathbf{v} \times \mathbf{b} \rangle^{(0)}$ vanishes, the lowest order contribution to the turbulent EMF is

$$\begin{aligned} \langle \mathbf{v} \times \mathbf{b} \rangle^{(1)} &= \langle \mathbf{v}^{(0)} \times \mathbf{b}^{(1)} \rangle + \langle \mathbf{v}^{(1)} \times \mathbf{b}^{(0)} \rangle \\ &= \int_0^t \langle \mathbf{v}^{(0)}(t) \times \partial_{t'} \mathbf{b}^{(1)}(t') \rangle dt' + \int_0^t \langle \partial_{t'} \mathbf{v}^{(1)}(t') \times \mathbf{b}^{(0)}(t) \rangle dt' \end{aligned} \tag{63}$$

To linear order, using the induction equation for $\mathbf{b}^{(1)}$ and the Navier-Stokes equation for $\mathbf{v}^{(1)}$, it can be shown that by analogy to the derivations of (58) and (60), (combined with a revised first-order smoothing approximation that assumes $|\mathbf{b}^{(1)}/\overline{\mathbf{B}}| < 1$) (63) becomes

$$\langle \mathbf{v} \times \mathbf{b} \rangle^{(1)} = -\frac{\overline{\mathbf{B}}}{3} \left(\int_0^t \langle \mathbf{v}^{(0)}(t) \cdot \nabla \times \mathbf{v}^{(0)}(t') \rangle dt' - \int_0^t \langle \mathbf{b}^{(0)}(t) \cdot \nabla \times \mathbf{b}^{(0)}(t') \rangle dt' \right). \tag{64}$$

FBC showed that in the case of negligible mean field gradients, it is still the zeroth-order kinetic and current helicities which appear most explicitly in α, even to all orders in $\overline{\mathbf{B}}$.

It is clear that the zeroth order helicities are not necessarily equal to those constructed with the full turbulent quantities since

$$\langle \mathbf{v} \cdot \nabla \times \mathbf{v} \rangle = \langle \mathbf{v}^{(0)} \cdot \nabla \times \mathbf{v}^{(0)} \rangle + \langle \mathbf{v}^{(A)} \cdot \nabla \times \mathbf{v}^{(0)} \rangle + \langle \mathbf{v}^{(A)} \cdot \nabla \times \mathbf{v}^{(0)} \rangle \tag{65}$$

and similarly

$$\langle \mathbf{b} \cdot \nabla \times \mathbf{b} \rangle = \langle \mathbf{b}^{(0)} \cdot \nabla \times \mathbf{b}^{(0)} \rangle + \langle \mathbf{b}^{(A)} \cdot \nabla \times \mathbf{b}^{(0)} \rangle + \langle \mathbf{b}^{(A)} \cdot \nabla \times \mathbf{b}^{(0)} \rangle, \tag{66}$$

so the extra terms on the right must be dealt with. One might ask however, if the first terms on the right of (65) and (66) dominate, why can't we simply replace $\langle \mathbf{b}^{(0)} \cdot \nabla \times \mathbf{b}^{(0)} \rangle$ by $\langle \mathbf{b} \cdot \nabla \times \mathbf{b} \rangle$ wherever the former occurs? The reason is that the appropriate helicity which then enters α is $\langle \mathbf{b}^{(0)} \cdot \nabla \times \mathbf{b}^{(0)} \rangle$ to lowest order. Then the H_1^M entering on the left of (45) would be second order. But then the magnetic helicity, H_2^M, entering (46) would also be second order. Thus the current helicity entering α is zeroth order whereas that entering the helicity conservation equation would be second order. There is an ordering mismatch. This is a problem because the success of the model of 2.3 depends on our being able to circumvent this ordering ambiguity and presume that the current helicity entering (47) is exactly k_2^2 times the H_2^M entering (46).

The procedure outlined to derive (64) and the subtlety just described with respect to ordering is basically the "ordering ambiguity" that was discussed in [55]. There it was shown that [46,48] effectively derived the form (63) rather than (47) by linearizing in terms of $\overline{\mathbf{B}}$ but did not identify that they had derived the zeroth order contribution to α. Thus [46,48] were actually using a similar expansion to that of [52]. The subsequent manipulations of [46] and [48] required that they had derived α as a function of the full $\mathbf{v}$ and $\mathbf{b}$, much like our manipulations in Sect. 2.3. The issue also arises subtly in the k space derivation of PFL and is presently unresolved.

3 Conclusions and Open Questions

3.1 Small-Scale Dynamo

For non-helical turbulence, and for $Pr \geq 1$, current simulations indicate that the magnetic field piles up on the resistive scales $k \gg k_f$ when forced externally [11] in a periodic box. The reason for this effect seems to be that the forcing scale inputs shear directly into the small-scale fields, so the power on small scales is the result of cross field structure. But for sufficiently helical turbulence, the spectrum changes: the peak at the resistive scale migrates to the forcing scale (wavenumber k_f) [14]. Can we understand the migration of the small scale field peak as a function of time? Will this picture survive future numerical testing? How do the boundary conditions affect the results?

To explain the change in the small-scale spectrum from the non-helical to helical case, it is possible that what works for the large scale field may also help understand what happens for the small scale field. The kinetic helicity input at k_f also cascades to higher wavenumbers, and so there is a source of helicity at these wavenumbers. Perhaps the change in shape of the small scale spectrum might be modeled by a self-similar set of nested "mean-field" dynamos. The principle is that for each small-scale wavenumber $k_s > k_f$ there is a range of $k_f < k_l < k_s$ for which the inverse cascade field growth driven by the helical turbulence at k_s can overcome the forward cascade of magnetic energy to $k > k_l$. An inverse cascade modeled in this way might account for the overall spectral shape change, but this is presently just speculation.

3.2 Large Scale Periodic Box α^2 Dynamo

The critical value of fractional helicity which determines the migration of the small-scale peak is exactly the same as that for which the kinematic α^2 dynamo has a positive growth rate, and so a large scale field grows at $k_1 < k_f$ in concurrence with the migration of the small-scale peak to k_f (see Sects. 1 and 4). The rate of growth, the saturation level, and the dependence on R_M observed in periodic box simulations of fully non-linear α^2 dynamo seem to be well modeled by a two-scale dynamical quenching model of [30] and Sect. 4. The dynamo α is the difference between the kinetic and current helicities associated with the small-scale field, and the growth of the large scale field is associated with the growth of the large scale magnetic helicity. Because total magnetic helicity is conserved up to a resistive term, the growth of large scale magnetic helicity also means growth of the opposite sign of small-scale magnetic helicity, and thus small-scale current helicity. This eventually suppresses the growth of the large scale magnetic field by reducing α. The time dependent process is non-linear. There are some unresolved issues with this theory however: 1) the theory works well for maximally helical forcing. How does it generalize to arbitrary helical forcing? 2) How is the correction to α to be properly derived in configuration space (see Sect. 5.)? The success of the theory is based on the formula of α from [31] which leaves open the ordering ambiguity discussed in Sect. 5. 3) What is

the appropriate theory for β quenching? The saturation value of the mean field in the α^2 dynamo depends somewhat on the choice of β but the ratio of large to small-scale fields does not. Dynamos with shear and their cycle periods are sensitive to the form of β [26]. 4) Along these lines, how does the dynamical quenching theory based on magnetic helicity conservation apply to the $\alpha - \Omega$ dynamo? This is work in progress [26]. 5) Do dynamos in astrophysics really need to be fast? This is another important reason to study β quenching. But the question should be revisited. Perhaps the kinematic phase (see (54)) can last long enough for substantial amplification, even for asymptotically slow dynamos when no cycle period is required. But one must also consider the observed cycle periods like that of the sun which seem to be fast. Perhaps highly anisotropic turbulence and small cross field structures that can dissipate quickly might even allow fast cycle periods with a thoroughly redefined R_M. See also [28].

3.3 Coronal Activity and Open Boundary Dynamos (OBD)

In addition to these considerations, the role of boundary terms needs to be considered. In a real system there is gravity, rotation, shear, stratification and buoyancy. It may not be enough to just invoke arbitrarily open boundaries to test the OBD but the physics at the boundary itself needs to be studied (e.g. [57]). Winds might be essential. Do mean field dynamos act in symbiosis with jets and winds? We know that the sun is shedding magnetic helicity. Does this enable the mean field dynamo to be fast for all times? In astrophysical rotators, the α effect should depend on height. How does this enter in the non-linear theory?

I showed above that the estimated energy deposition rates to maintain fast mean field dynamo action are consistent with the coronal + wind power from the Sun, Galaxy and Seyfert I AGN all of these sources are natural sites for $\alpha - \Omega$ type dynamos. The helical properties also seem to agree well in the solar case where they can be observed [36]. The steady flow of magnetic energy into coronae thus provides an interesting connection between mean field dynamos and coronal dissipation in a range of sources. A reasonably steady (over time scales long compared to turbulent turnover time scales), active corona with multi-scale helical structures, may provide a self-consistency check for fast dynamo production of magnetic field.

For the α^2 case, figure 9 shows the effect of including an additional loss term in (50), proportional to h_2, on the growth of h_1. The effect is to suppress the backreaction by taking away excess small-scale helicity and allowing the large scale field to grow stronger. Understanding appropriate form of loss term for real $\alpha - \Omega$ dynamos needs more study.

3.4 New Diagram of MFD Operation Is Needed

Related to the boundary issue is the generation of magnetic flux inside of a rotator. For the Galaxy, diffusion of $\overline{\mathbf{B}}$ at the boundary is required to maintain a quadrupole field structure with a net flux inside the disk [7]. The total flux is

conserved, but to obtain a net flux inside the rotator where desired, the return flux must be shed from this region. The surface diffusion of total magnetic field for the Galaxy may be difficult [57] and remains an open question. (One alternative possibility is that the mean field diffuses radially more efficiently at the top of the disk than at the mid-plane.) Note however, that it is $\overline{\mathbf{B}}$ which needs to diffuse, not necessarily the total field or the matter. For the Sun, the solar cycle also requires diffusion of mean field through the boundary at the surface and/or mid-plane.

Since both flux generation inside the rotator, and helicity shedding can both appeal to boundary terms it is possible that the two are related in the simplest $\alpha - \Omega$ dynamo. Note that it was the growth of the small-scale *current* helicity $\langle \mathbf{j} \cdot \mathbf{b} \rangle$ that leads to the suppression of α. Thus one can ask directly if the textbook $\alpha - \Omega$ dynamo has any way of getting rid of this quantity and at the same time generating magnetic flux.

In fact there is something missing when one draws a standard picture of dynamo action either for α^2 or for $\alpha - \Omega$. The latter is shown in figure 10, but the implication holds also shown for the α^2 case. The issue is that the small-scale loop that arises from the kinetic part of the α effect actually induces the *same* sign of the current helicity of the small-scale field as that of the large scale field that it generates. But I have discussed how the growth of the large-scale field should be accompanied by the *opposite* sign of the small-scale helicities. It is intriguing because in the kinematic regime of the α^2 dynamo, the small-scale field does grow first at the smallest scales, and it seems to take nearly until t_{kin} for the peak of the current helicity of opposite sign to that of the large scale to move fully up to the forcing scale (Brandenburg and Maron, personal communication.) In this sense, figure 10 seems limited to the kinematic regime: it would be nice to see a more accurate picture that actually shows the generation of the small-scale current helicity of opposite sign to that of the large scale and how it migrates to the forcing scale graphically.

Finally, note that it is in part the potential for flux generation that distinguishes a mean field dynamo from the kind of dynamo which generates large scale fields as a result of the magneto-rotational instability (MRI). There, the field induces turbulence, which then generates a random component of the field, which is subsequently sheared by the differential rotation. Magnetic energy grows exponentially, even on the largest scale allowed, but there is no real flux generation here and no need for input helicity. Note however that for real accretion discs, helicity is undoubtedly present since all the ingredients are there, density gradient, and turbulence. Thus it is important to understand even in accretion disks, what is the role of helical dynamos. This has not yet been done exhaustively. From this point of view, the mean field dynamo formalism should still apply, with a dynamical theory for quenching, and the MRI may play the role as the source of turbulence. Nevertheless more work is needed.

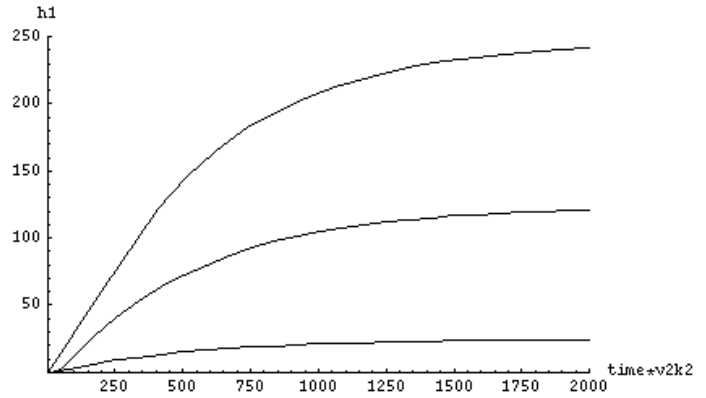


Fig. 9. The effect of adding a loss term proportional to h_2 in (50) on the growth of h_1 for three different values of loss. The loss term is added by introducing multiplicative factors on the last term of (50). From top to bottom, these factors are $10, 5$ and 1(=no loss), respectively. Here $R_M = 200$ and $k_2/k_1 = 5$.

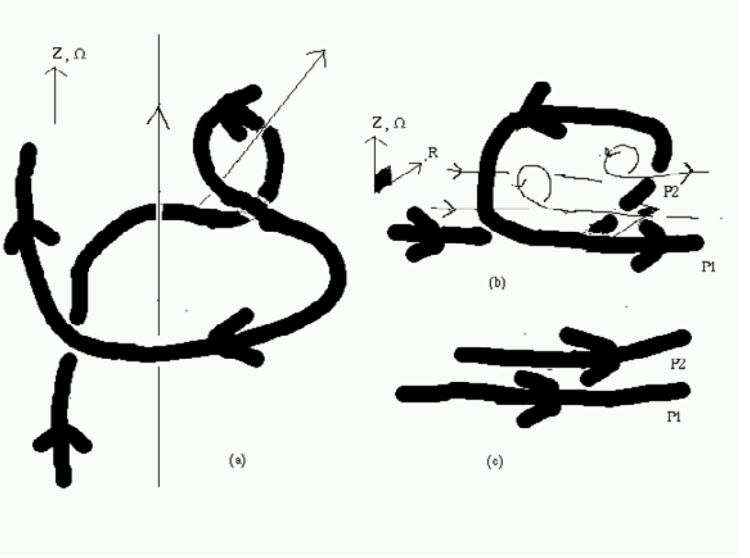


Fig. 10. The kinematic picture and its limitations. This is assumed to be a northern hemisphere so that kinetic helicity is negative. (a) The large scale loop results from shear and the small loop results from the α effect. Note that the current helicity of the smaller loop has the same sign as that of the large scale loop. So the question arises: in a revised picture how can one incorporate the growth of magnetic and/or current helicity of the opposite sign that accompanies growth of large scale helicity? (b) This just shows the effect of multiple small-scale loops as they sum together. (c) the process of reinforcement of the toroidal field is shown.

3.5 What Is the Role of Magnetic Reconnection?

It is generally perceived that reconnection is important for large scale dynamos but the precise way in which this is the case is subtle. In general, the processes of reconnection can serve two roles: it changes the topology of the field, and also dissipates magnetic energy. In the mean-field dynamo formalism, the large scale field is degenerate with respect to small-scale topology. By definition, taking the mean means smoothing out over the small scales. Thus a series of disconnected loops could have the same mean field as an undulating topologically connected field. In the mean field formalism therefore, the role of reconnection is not explicit.

Neither the mean field nor the fluctuating field are the actual field. If one really does want the mean field dynamo to result in a topologically connected actual field, then reconnection would be important for the topology. But it is important to asses the particular application for when this is necessary: the large scale field in the Galaxy is measured mainly by Faraday rotation which provides little information about the actual field topology. The role of reconnection may only be one of ensuring that there is a turbulent cascade which ensures that heat rather than the magnetic field is sink of the kinetic energy.

Another unanswered question relating to dynamos and reconnection is: does reconnection play a role in the conservation of magnetic helicity and in the migration of the peak of the small-scale magnetic energy from the resistive scale to the forcing scale in helically forced turbulence?

Acknowledgments

Thanks to the organizers (E. Falgarone and T. Passot) of the Paris July 2001 workshop on MHD turbulence for an excellent meeting. Thanks to G. Field, A. Brandenburg, and J. Maron for recent collaborations incorporated into the material herein. Thanks also to R. Kulsrud, B. Mattheaus, and A. Pouquet for stimulating discussions. Support from DOE grant DE-FG02-00ER54600 is acknowledged.

References

1. A.A. Ruzmaikin, A.M. Shukurov, D.D. Sokoloff, *Magnetic Fields of Galaxies*, (Kluwer Press, Dodrecht, 1988).
2. R. Kulsrud, R. Cen, J.P. Ostriker, and D. Ryu, ApJ, **480** 481 (1997)
3. E. G. Zweibel & C. Heiles, Nature, **385** 131 (1997).
4. B.D.G. Chandran, S.C. Cowley, & M. Morris, ApJ, **528** 723 (2000).
5. S.H. Lubow, J.C.B. Papaloizou, J.E. Pringle, MNRAS, 267, 235 (1994).
6. A.P. Kazanstev, Sov. Physics. JETP, **26** 1031 (1968).
7. E.N. Parker, *Cosmical Magnetic Fields* (Oxford: Clarendon Press, 1979).
8. Ya. B. Zeldovich, A.A. Ruzmaikin,, and D.D. Sokoloff, *Magnetic Fields in Astrophysics*, (Gordon and Breach, New York, 1983).
9. R.M. Kulsrud & S.W. Anderson, Astrophys. J. **396** 606 (1992).

10. S. Kida, S. Yanase, J. & Mizushima, Physics of Fluids, **3** 457 (1991)
11. J. Maron & S. Cowley, to be submitted to ApJ (2002); http://xxx.lanl.gov/abs/astro-ph/0111008
12. A. Schekochihin, S. Cowley, J. Maron, J., & L. Malyshkin 2002, Phys Rev. E., 65, 6305.
13. R. Beck., A. Brandenburg, D. Moss, A. Shukurov, D. Sokoloff, Galactic Magnetism: Recent Developments and Perspectives, Ann. Rev. Astron. Astrophys., **34**, 155 (1996).
14. J. Maron & E.G. Blackman, ApJL **566**, L41 (2002).
15. E. Vishniac & J. Cho, ApJ, **550** 752 (2000).
16. A. Brandenburg, ApJ, **550** 824 (2001)
17. H.K. Moffatt, H. K. *Magnetic Field Generation in Electrically Conducting Fluids*, (Cambridge University Press, Cambridge, 1978).
18. F. Krause & K.-H. Rädler *Mean-field magnetohydrodynamics and dynamo theory*, (Pergamon Press, New York, 1980).
19. M. Steenbeck, F. Krause, & K.-H. Rädler, Z. Naturforsch. **21a**, 369 (1966).
20. S.A. Balbus & J. Hawley, Rev. Mod. Phys. **70**, 1 (1998).
21. T.G. Cowling, *Magnetohydrodynamics*, (Wiley Interscience, New York, 1957).
22. J.H. Piddington, *Cosmical Electrodynamics*, (Krieger Press, Malbar, 1981).
23. S.I. Vainshtein & F. Cattaneo, ApJ **393** 165 (1992).
24. L.L. Kitchatinov, V.V. Pipin, G. Rüdiger, G., & M. Kuker, Astron. Nachr., **315**, 157 (1994).
25. F. Cattaneo, & D.W. Hughes, Phys. Rev. E. **54**, 4532 (1996).
26. E.G. Blackman & A. Brandenburg, submitted to ApJ (2002); http://xxx.lanl.gov/abs/astro-ph/0204497.
27. S. Vainshtein, Phys. Rev. Lett., **80**, 4879 (1998).
28. A. Brandenburg, W. Dobler, & K. Subramanian, Astron. Nachr, **323** 99 (2002).
29. A. Brandenburg, A. Bigazzi, & K. Subramanian, MNRAS **325** 685 (2001)
30. G.B. Field & E.G. Blackman, ApJ **572** 685 (2002).
31. A. Pouquet, U. Frisch, & J. Léorat, J. Fluid Mech. **77**, 321 (1976).
32. G.B. Field in *Magnetospheric Phenomena in Astrophysics*, R. Epstein & W. Feldman, eds. AIP Conference Proceedings 144 (American Institute of Physics, Melville NY, 1986), p324.
33. E.G. Blackman, & G.B. Field, Astrophys. J. **534** 984 (2000a).
34. M.A. Berger & G.B. Field, J. Fluid Mech. **147** 133 (1984).
35. J.W. Bieber & D.M. Rust, ApJ, **453** 911 (1995)
36. D.M. Rust & A. Kumar, Astrophys. J. **464**, L199 (1994).
37. J.A. Markiel & J.H. Thomas, Astrophys. J. **523** 827 (1999)
38. A. Ruzmaikin, in *Magnetic Helicity in Space and Laboratory Plasmas*, edited by A. Pevtsov, R. Canfield, & X. Brown, (Amer. Geophys. Union, Washington, 1999), p111; M.A. Berger & A. Ruzmaikin, Journ. Geophys. Res. **105**, 110481
39. U. Frisch, A. Pouquet, J. Léorat & A. Mazure, J. Fluid Mech. **68**, 769 (1975).
40. G.L. Withbroe, & R.W. Noyes, Ann. Rev. Astron. Astrophys. **15**, 363 (1977);
41. E.G. Blackman, & G.B. Field, Mon. Not. R. Astron. Soc. **318**, 724 (2000a).
42. B.D. Savage, in *The Physics of the Interstellar Medium and Intergalactic Medium*, A. Ferrara, C.F. McKee, C. Heiles, & P.R. Shapiro eds. ASP conf ser vol 60. (Astronomical Society of the Pacific, San Francisco, 1995), p233.
43. R.J. Reynolds, L.M. Haffner, S.L. Tufte, Astrophys. J. **525**, L21 (1999).

44. M. DeVries & J. Kuijpers, Astron. & Astrophys. **266**, 77 (1992); F. Haardt & L. Maraschi, Astrophys. J. **413**, 507 (1993); G.B. Field & R.D. Rogers., Astrophys. J. **403**, 94 (1993); T. DiMatteo E.G. Blackman & A.C. Fabian, Mon. Not. R. Astron. Soc. **291** L23 (1997); A. Merloni & A.C. Fabian, Mon. Not. R. Astron. Soc. **332** 165 (2000).
45. H. Ji, Phys. Rev. Lett. **83** 3198 (1999); H. Ji in *Magnetic Helicity in Space and Laboratory Plasmas*, edited by A. Pevtsov, R. Canfield, & X. Brown, (Amer. Geophys. Union, Washington, 1999), p167; H. Ji, & Prager, S. C. in press Magnetohydrodynamics, astro-ph/0110352 (2001)
46. A.V. Gruzinov & P.H. Diamond P.H., Phys. Rev. Lett., **72** 1651 (1994); A.V. Gruzinov & P.H. Diamond, Physics of Plasmas, **2** 1941 (1995);
47. A.V. Gruzinov & P.H. Diamond, Physics of Plasmas, **3** 1853 (1996).
48. A. Bhattacharjee & Y. Yuan, Astrophys. J. **449** 739 (1995).
49. N. Kleeorin, D. Moss, D., I. Rogachevskii, D. Sokoloff, A&A, **361** L5 (2000)
50. N. Kleeorin, I. Rogachevskii, A. Ruzmaikin, A&A, **297** L59 (1995)
51. N. Kleeorin & I. Rogachevskii, Phys. Rev. E., **59** 6724 (1999)
52. G.B. Field, E.G. Blackman, & H. Chou, Astrophys. J. **513**, 638 (1999).
53. Ya.-B. Zeldovich, Sov Phys. JETP, **4** 460 (1957).
54. E.G. Blackman and G.B. Field, Physics of Plasmas, **8** 2407 (2001)
55. E.G. Blackman & G.B. Field, Astrophys. J. **521** 597 (1999).
56. W.N. Brandt, T. Boller, A.C. Fabian,& M. Ruszkowski, Mon. Not. R. Astron. Soc. **303**, L53 (1999).
57. R.R. Rafikov& R.M. Kulsrud, Mon. Not. R. Astron. Soc. **314**, 839 (2000).
58. F. Cattaneo, Astrophys. J. **434**, 200 (1994).
59. A. Brandenburg, & W.Dobler A&A, **369** 329 (2001).
60. R. Arlt & A. Brandenburg A&A 380, 359 (2001)